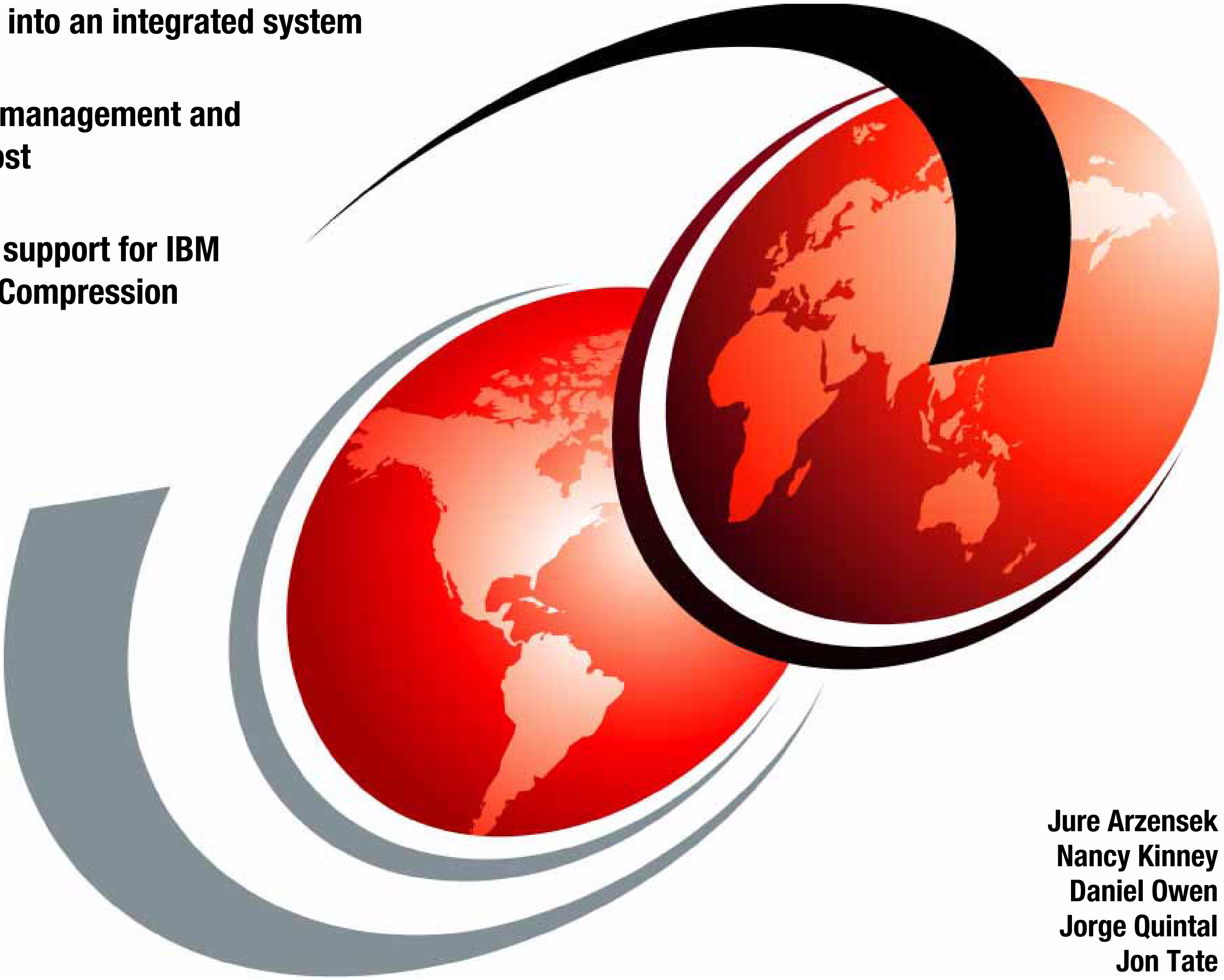
IBM

Implementing the IBM Storwize V7000 Unified Disk System

Consolidates storage and file serving workloads into an integrated system

Simplifies management and reduces cost

Integrates support for IBM Real-time Compression



Jure Arzensek
Nancy Kinney
Daniel Owen
Jorge Quintal
Jon Tate

Redbooks

ibm.com/redbooks

International Technical Support Organization

Implementing the IBM Storwize V7000 Unified Disk System

April 2014

Note: Before using this information and the product it supports, read the information in “Notices” on page xi.

Second Edition (April 2014)

This edition applies to Version 7.1.0.5 of the IBM Storwize V7000 code and Version v1.4.2.0-27-2273 of the IBM Storwize V7000 File Module code.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

Active Cloud Engine®
AFS™
AIX®
BladeCenter®
DS4000®
Easy Tier®
FlashCopy®
Global Technology Services®
GPFS™
IBM®
PartnerWorld®
POWER®
PureFlex®
Real-time Compression™
Redbooks®
Redbooks (logo) ®
Storwize®
System Storage®
System x®
Tivoli®
XIV®
z/OS®

The following terms are trademarks of other companies:

Intel, Intel Xeon, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

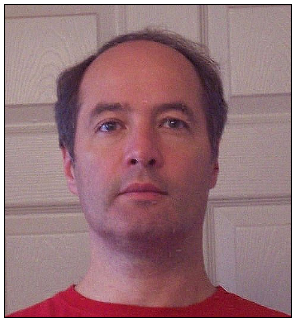
Preface

This IBM® Redbooks® publication introduces the IBM Storwize® V7000 Unified Disk System, a virtualized storage system that consolidates block and file workloads into a single storage system. Advantages include simplicity of management, reduced cost, highly scalable capacity, performance, and high availability. It also offers improved efficiency and flexibility through built-in solid-state drive optimization, thin provisioning, IBM Real-time Compression™, and nondisruptive migration of data from existing storage. The system can virtualize and reuse existing disk systems, which offers a greater potential return on investment.

We suggest that you familiarize yourself with the following Redbooks publications to get the most from this book:

- *Implementing the IBM Storwize V7000 V6.3*, SG24-7938
- *Implementing the IBM System Storage SAN Volume Controller V6.3*, SG24-7933
- *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859
- *SONAS Implementation and Best Practices Guide*, SG24-7962
- *SONAS Concepts, Architecture, and Planning Guide*, SG24-7963

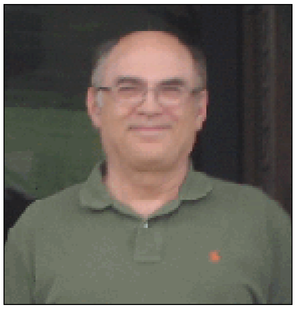
Authors



Jure Arzensek is an Advisory IT Specialist for IBM Slovenia. He works as a Product Field Engineer for the Central and Eastern Europe, Middle East, and Africa Level 2 team, supporting IBM PureFlex® and IBM BladeCenter® products. He has been with IBM since 1995 and has worked in various technical support and technical education roles in Europe, the Middle East and Africa and Central Europe, Russia, and Turkey. Jure holds a degree in Computer Science from the University of Ljubljana. His other areas of expertise include IBM System x® servers, storage area network (SAN), IBM SAN Volume Controller, IBM Storwize V7000, IBM System Storage® DS3000, IBM DS4000®, and IBM DS5000 products, and network operating systems for the Intel platform. He has co-authored 13 other Redbooks publications.

Nancy Kinney is an Infrastructure Architect. Previously, she worked as an IBM US Remote Technical Support Engineer for IBM Global Technology Services® in the Austin IBM AIX® system and IBM System Storage center, where she acquired end-to-end experience in troubleshooting Fibre Channel (FC) technologies. She holds an IBM Midrange Specialist Certification, as well as NetApp/N-Series NCDA Certification. She has a wide range of experience working with FC storage technologies from multiple vendors, as well as working with multipathing drivers for operating systems that attach to storage and networking technologies.

Daniel Owen is the Performance Architect for the IBM Storwize V7000 Unified Disk System. Before joining the Systems Technology Group's System Storage team, he worked within the IBM POWER® Systems group, developing technology to improve the performance of applications on IBM POWER and IBM AIX systems. Daniel is a Chartered Engineer who has more than a decade of experience working on the performance of computer systems.



Jorge Quintal is a Storage Systems Management Consultant who provides development support for IBM Real-time Compression. He joined IBM through the acquisition of Sequent Computer Systems in 1999. He has since worked for Storage Lab Services as one of the original members working with SAN File System, on the IBM SAN Volume Controller, network-attached storage (NAS), and as lead for N-Series services development and implementations. Jorge also worked for an extended period as an IBM XIV® Technical Advisor.

Jon Tate is a Project Manager for IBM System Storage SAN Solutions at the International Technical Support Organization (ITSO), San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2 support for IBM storage products. Jon has 27 years of experience in storage software and management, services, and support, and is both an IBM Certified IT Specialist and an IBM SAN Certified Specialist. He is also the UK Chairman of the Storage Networking Industry Association.

This book was produced by a team of specialists working at Brocade Communications Systems, San Jose; IBM Manchester Labs, UK; and the IBM International Technical Support Organization, San Jose Center.

Previous authors:

Andreas Baer
Nitzan Iron
Tom Jahn
Paul Jenkin
Jorge Quintal
Bosmat Tuv-El

We extend our thanks to the following people for their contributions to this project, including the development and Product Field Engineering teams in the IBM UK Hursley Park Lab:

Robin Findlay
Carlos Fuente
Geoff Lane
Andrew Martin
Cameron McAllister
Paul Merrison
Steve Randle
Matt Smith
Barry Whyte

Muhammad Zubair
Chris Canto
Peter Eccles
IBM Hursley

Duane Bolland
Jackson Shea
IBM Beaverton

Norm Bogard
IBM Orlando

Chris Saul
IBM San Jose

Sangam Racherla
IBM ITSO

Achim Christ
Nils Haustein
Michael Jahn
Thomas Luther
Alexander Saupp
IBM Germany

Special thanks to the Brocade staff for their unparalleled support of this residency in terms of equipment and support in many areas:

Silviano Gaona
Brian Steffler
Marcus Thordal
Jim Baldyga
Brocade Communications Systems

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form:

  ibm.com/redbooks

- Send your comments by email:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Introduction

The IBM Storwize V7000 Unified Disk System integrates the serving of storage and file-related services, such as file sharing and file transfer capabilities, in one system. The Storwize V7000 Unified system can provide storage system virtualization as well, using the mature virtualization capabilities of the IBM SAN Volume Controller. It is an integrated storage server, storage virtualization, and file server appliance.

1.1 A short history lesson

IT infrastructure concepts and products advance and change over time, adjusting to changing requirements and tasks. For instance, in many computing environments, a centralized approach with a strictly specified design and infrastructure components has been superseded by a client/server approach, with a decentralized and less proprietary, more interoperable infrastructure. This infrastructure is easier to adapt to new concepts and tasks.

As always, IT architects have to work with the concepts, technology, and designs that are available at a particular time. For instance, servers in client/server environments were able to use only internal storage previously. This changed with external Redundant Array of Independent Disks (RAID) and RAID storage systems. If we look at storage as a service, this previously internal-only service was delegated to a specialized device, which enabled new infrastructure concepts.

The serving of storage from a central element in the infrastructure to many storage clients and the implementation of high availability by mirroring data to two independent storage servers is more prevalent in the industry. These specialized storage servers are, in essence, storage server appliances. What was simply called a server because it housed all elements in one device and provided a service to clients is now a client, specifically a storage client. IT architects need to use the service that is provided by the storage servers to build services on top of those.

Servers were becoming used for more specialized tasks. Later, the hardware and software was adapted to support this specialization, to a certain degree.

One of these tasks was organizing files in file systems and making the file system space and the files accessible by using file-sharing clients. These adapted servers are known as *file servers*. With external storage, these devices are storage clients in one aspect and file servers in another. They *use* a lower-level service (storage) and *provide* a higher-level service (file serving).

To enhance the functionality, availability, disaster tolerance, and other aspects of a file-serving subinfrastructure, new types of devices were developed and introduced. These were specialized, single-purpose file servers (a *file server appliance*). They provide only the functions that are implemented in the product of the particular vendor, without the explicit ability to use them for different tasks (as with multipurpose servers).

This specialization, which can have many other advantages, led to appliances that were called *network-attached storage* (*NAS*). Although, strictly speaking, this is serving files, not providing storage.

This gave us a modular, layered infrastructure in which each logical layer corresponds with a class of device that provides services that are related only to that layer. A storage server serves storage to a storage client. A file server serves files to a file client. These devices are connected to each other by using different forms of external networks and using specialized protocols to provide their services.

When file server appliances started to be used to generate storage volumes from files in their file systems, they made these volumes accessible to storage clients by using Internet Small Computer System Interface (iSCSI) and Fibre Channel (FC) technology. This meant that these devices provided services that belonged to two different functional layers, and they acted as both file servers and storage servers. To distinguish them, they were called *unified storage*. The IBM approach to this type of storage is to take a storage server that can provide storage virtualization and integrate it with a file server product in one device.

This integrated product is the IBM Storwize V7000 Unified storage system, which is shown in Figure 1-1.

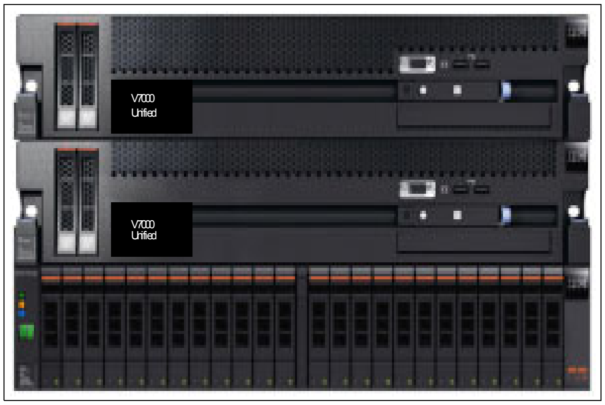


Figure 1-1 IBM Storwize V7000 Unified system

Internally, this device is still built with a layered approach: A storage server serves storage to external storage clients and to the internal file server, and the file server provides file services to external file clients. This is a truly integrated storage server, storage virtualization, and file server appliance.

1.2 About the rest of this book

Initially, we provide the basic terminology and concepts of storage and file services and how they relate to each other. Building on that, we introduce file sharing and file transfer methods, in general. We then briefly explain the architecture of the Storwize V7000 and the specifics of the implementation of the file-related functionality and the access control options.

Storage-related functions, such as virtualization and copy services, are covered more briefly because these topics are already covered in available Redbooks publications. These functions are the same as in the IBM Storwize V7000 virtualized, software-defined storage system. We recommend having a copy of the following book available to help you understand the Storwize V7000 information: *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

Information about the file server-related part of the product, such as the IBM General Parallel File System (GPFS™) and differences in Scale Out Network Attached Storage (SONAS) are included also. For more information about SONAS, see the following publications:

- *SONAS Implementation and Best Practices Guide*, SG24-7962
- *SONAS Concepts, Architecture, and Planning Guide*, SG24-7963

For GPFS, we recommend reading *Implementing the IBM General Parallel File System (GPFS) in a Cross Platform Environment*, SG24-7844.

The theoretical part of the book is followed by the implementation chapters, which describe the user interfaces, planning for implementation, and the actual implementation of the Storwize V7000 Unified system. This includes antivirus implementation, performance and monitoring overview, backup and recovery, and troubleshooting and maintenance.

This book helps you understand file services and how they are implemented in the Storwize V7000 Unified system and how to successfully implement, run, and maintain the system.

1.3 Latest release highlights

Although the Storwize V7000 Unified system received various small updates in various areas, there are several new enhancements and features worthy of note.

1.3.1 USB Access Recovery

If a system is relocated or the network is changed, access to the management interface can get lost unless you plan carefully. This change allows for a quick and easy way to recover the file module console and reset the management Ips. This is covered in detail in the implementation and troubleshooting chapter.

1.3.2 Call-home installation enhancement

This is an enhancement that allows our clients to use call home setup, which enables better tracking for quick access to machines installed in the field or when sending an IBM service representative into the field. After the System Setup wizard completes the user is prompted to complete the Support Services wizard.

1.3.3 SSH tunnel from the IFS Management Node to the Storwize V7000 Node

This enables the Host Software Group third-party plug-ins for V7000 to work on the V7000 Unified system as Chapter 9, “GUI and CLI setup and use” on page 83 explains.

2

Terminology and file serving concepts

This chapter describes terminology that is used for storage and file services and for client/server file sharing and file transfer protocols.

The IBM Storwize V7000 Unified Disk System provides file and block storage. The terminology in this chapter pertains mostly to file storage. For detailed block storage information, consult the IBM Redbooks publication titled *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

2.1 Terminology for storage and file services

The Storwize V7000 Unified system gives you the ability to manage both file and block level data within a single system. With an integrated product such as this system, it is essential to use clear, consistent terminology.

The main focus when designing infrastructures is the service that the products provide. (For example: Is it the kind of service that the client asks for? Are the requirements met?) Even though IBM products perform input/output (I/O) operations, it is services such as the mapping of logical volumes and the sharing of files that are important to keep in mind in this context.

2.1.1 Terminology for random access mass storage

The lowest level of service that is provided to permanently store and retrieve data for use in upper layers in computer systems is random access mass storage, hereafter referred to as *storage*. For instance, storage is provided by hard disk drives (HDDs), solid-state drives (SSDs), and (by using HDDs and SSDs) by Redundant Array of Independent Disks (RAID) controllers and the controllers of external storage systems and storage servers. Often, the term *block storage* is used, although in many cases it is not necessary to include the word *block*.

The *block* prefix originates in the disk data organization architecture. For instance, the fixed-block architecture (FBA or FB) uses records of data that are written in "blocks" of a specific size. Logical block address (LBA), an FBA that is common today, uses a fixed sector size of 512 bytes. Referring to storage with the *block* prefix leaves out other architectures, such as count key data (CKD), that are used in IBM z/OS® systems, for example. In cases where it is necessary to differentiate it from CKD, it makes sense to refer to the storage as *fixed block storage* (FB storage), without omitting the word *fixed*.

LBA storage, as used by open systems, provides the storage to upper layers of the system as a sequence of blocks. This is called a *volume*. The *block* prefix is not necessary for the term *volume*, either, to differentiate it from upper layer facilities and protocols. The term is not needed because these protocols do not provide storage or volumes in any form. Instead, they *use* storage in the form of volumes and provide upper layer services.

Storage that is provided by physical random access mass storage devices, such as HDDs or SSDs, is referred to as *physical volume* (PV) storage. Examples are storage that is used in RAID and in LVM software components of an operating system.

When using physical volumes to start with, there are volumes that are called *logical volumes* (LVs) being presented in different places. The term *logical volume* is overused. To be understood, it depends on either the context that it is being used in or on a further definition or clarification. An LV can be a logical volume that is provided by a computer system's internal RAID adapter, an external storage system, or an LVM. All of these LVs have in common that they also present a sequence of blocks that can be used in the same way a PV would be. For upper layers that use these volumes, the source is apparent. This means that they do not know whether they are using a PV or an LV. For instance, an LV, like a PV, might be used by a file system (FS) or by a database management system (DBMS). The diagram in Figure 2-1 on page 7 shows a storage server that is providing logical volumes to a storage client.

A computer system that accesses a logical volume that is mapped to it is a *storage client*. A logical volume that is served to a storage client by a storage server (storage system) is often referred to as a LUN. This use of *LUN* is a misnomer, because LUN is short for logical unit number, an addressing scheme that is used to *identify* a logical unit (LU). This addressing scheme is used for Small Computer System Interface (SCSI) based systems, by using the variations of SCSI, Internet SCSI (iSCSI), and Fibre Channel (FC). A logical volume is only one type of SCSI *logical unit* (LU), identified by a LUN. However, LUs are *not* limited to logical volumes. They can be other devices that use the SCSI addressing scheme as well, such as tape devices.

To summarize: With *storage servers,* we provide *storage* service when we map logical volumes to storage clients.

Figure 2-1 depicts a storage server that is providing logical volumes to a storage client.

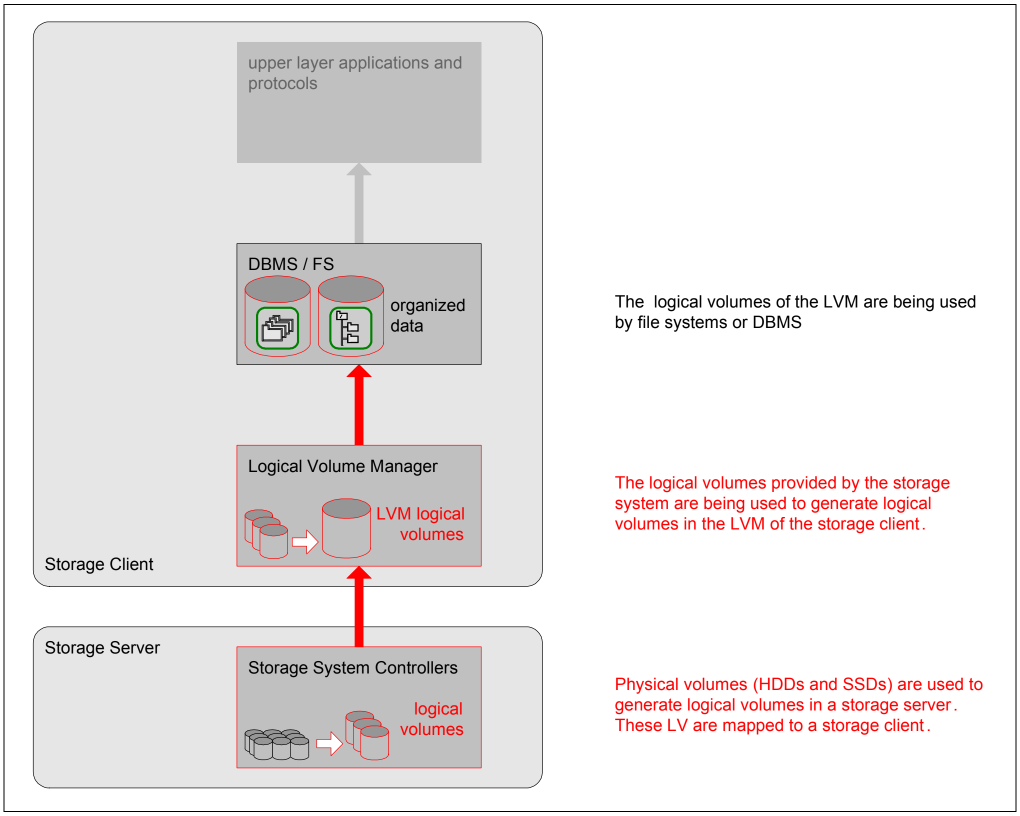

upper layer applications and protocols
DBMS / FS
organized data
The logical volumes of the LVM are being used by file systems or DBMS
Logical Volume Manager
LVM logical volumes
The logical volumes provided by the storage system are being used to generate logical volumes in the LVM of the storage client.
Storage Client
Storage Server
Storage System Controllers
logical volumes
Physical volumes (HDDs and SSDs) are used to generate logical volumes in a storage server. These LV are mapped to a storage client.

Figure 2-1 Storage server that is providing logical volumes to a storage client

2.1.2 Terminology for file systems and file sharing, file access, and file transfer

Operating systems of storage clients apply a structure that organizes the storage space of logical volumes to store and retrieve data. Usually, a Logical Volume Manager (LVM) takes the logical volumes from storage servers to build LVM logical volumes. On top of the LVM logical volumes are the facilities to structure them to enable writing, accessing, and reading data, such as file system (FS) or database management system (DBMS).

Sometimes, there is a bit of confusion about *file systems* and *file serving and file sharing*. This is partly rooted in the naming of the networking protocols that are used to share and access file resources.

A *file system* is used to control how information is stored and retrieved. Without a file system, information that is placed in a storage area is one large body of information, with no way to tell where one piece of information stops and the next begins.

By separating the information into pieces and giving each piece a name, the information is easily distinguished and identified. Taking its name from the way that paper-based information systems are named, each piece of information is called a *file*. The structure rules and the logic that are used to manage the groups of information and their names is called a *file system*.

There are many different kinds of file systems. Each one has a different structure and logic. Each one also has different properties of speed, flexibility, security, size, and other characteristics. Some file systems are designed to be used for specific applications. For example, the ISO 9660 file system is designed specifically for optical disks.

File systems can be used on many different kinds of storage devices. Each storage device uses a different kind of media. The most common storage device in use today is a hard disk drive with media that is a disc coated with a magnetic film. The film has ones and zeros "written" on it that send electrical pulses to a magnetic read/write head. Other media that are used are magnetic tape, optical disc, and flash memory. In some cases, the computer's main memory (RAM) is used to create a temporary file system for short-term use.

File systems are used to implement types of data to store, retrieve, and update a set of files. *File system* refers to either the abstract data structures that are used to define files or the actual software or firmware components that implement the abstract ideas.

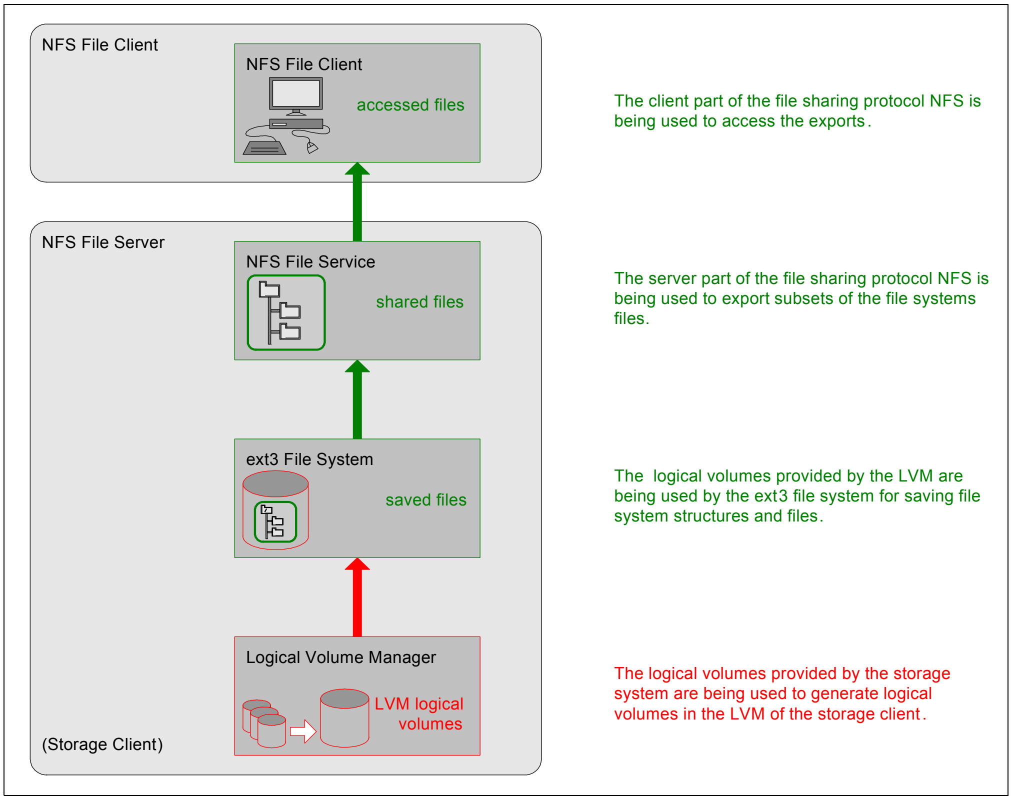
Some file systems are used on local data storage devices; others provide file access through a network protocol, for example, Network File System (NFS), Server Message Block (SMB), or Plan 9 Filesystem Protocol clients (known as Plan 9 or 9P). Some file systems are *virtual*, in that the "files" supplied are computed on request (for example, proxies) or are merely a mapping into a different file system that is used as a backing store. The file system manages access to both the content of files and the metadata about those files. It is responsible for arranging storage space. Reliability, efficiency, and tuning regarding the physical storage medium are important design considerations.

File sharing is a way of making data that is already organized in a file system *accessible* to users of other (network-connected) computer systems. These file sharing protocols, also called *file access protocols*, use client/server infrastructures. We do not describe Peer-to-Peer (P2P) file sharing protocols in this book. The *file server* part of the protocol makes the files accessible for the *file client* part of the protocol. It is common and technically correct to say that files are *shared*, not only because the server shares access to the files with the client but also because these files might be accessed by multiple clients (shared).

Some common file sharing protocols allow users to access files on another computer system's file system in a similar way that they access data in their local file system. These are the different versions of the NFS protocol and the SMB protocol. The SMB protocol often is referred to as *Common Internet File System* (CIFS), which is merely a *specific dialect* (one version) of the SMB protocol. See 2.2.2, "The Server Message Block protocol" on page 13, for more information about the SMB protocol. The newest versions of the NFS and SMB file sharing protocols are NFS 4.1 and SMB 2.1.

The terms that are used for NFS and SMB servers and clients are *NFS file server*, *NFS file client*, *SMB file server*, and *SMB file client*. Sometimes, the software instances are called *NFS file service* and *SMB file service* to distinguish them from the computer system's hardware that they run on. In short, the terms NFS server and NFS client, as well as SMB server and SMB client, are used.

The diagram in Figure 2-2 on page 9 shows an example of a file server that is using logical volumes, the ext3 file system, and the NFS file sharing protocol to make files accessible to an NFS client.



The client part of the file sharing protocol NFS is being used to access the exports.

The server part of the file sharing protocol NFS is being used to export subsets of the file systems files.

The logical volumes provided by the LVM are being used by the ext3 file system for saving file system structures and files.

The logical volumes provided by the storage system are being used to generate logical volumes in the LVM of the storage client.

Figure 2-2   File server housing a file system and sharing files, file client accessing files

Other methods of making files accessible and transferring files are the File Transfer Protocol (FTP), the Hypertext Transfer Protocol (HTTP) and its variations, and the Secure Shell (SSH) based protocols: Secure Copy Protocol (SCP) and Secure FTP (SFTP). The terms that are used for FTP and HTTP servers and clients are *FTP file server* (for short: *FTP server*), *FTP file client* (for short: *FTP client*), *HTTP server*, and *HTTP client*.

Specialized file servers are commonly called *network-attached storage* (NAS). Although NAS products are network-attached, they are file servers that use internal or external storage (in which case, they are storage clients). Although the term *NAS* is widely used, it might help to think of it as a file server appliance or simply as a special-purpose file server.

The diagram in Figure 2-3 on page 10 shows the different layers from storage to the access of files and how they relate to each other. The example that is shown is using a storage server to provision logical volumes to a storage client (the file server). The file server uses these logical volumes to organize data in the form of files in file systems. The file server then provides access to the content of these file systems to file clients by using file access protocols.

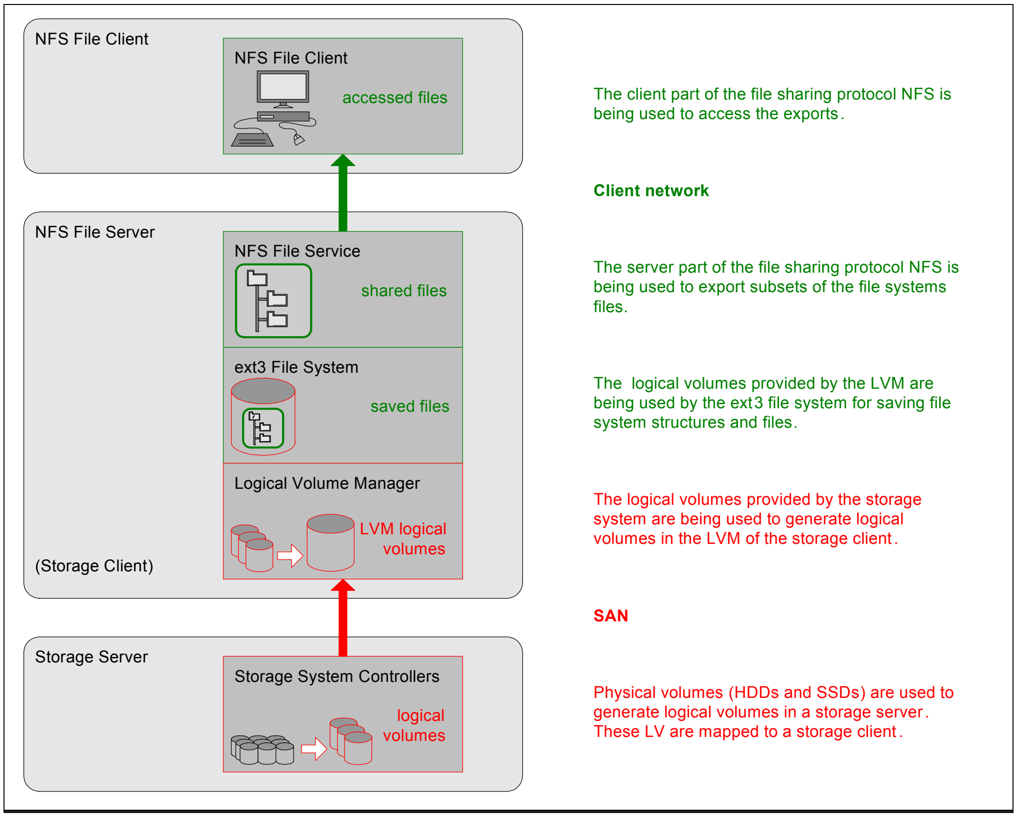

NFS File Client
NFS File Client
accessed files
The client part of the file sharing protocol NFS is being used to access the exports.
Client network
NFS File Server
NFS File Service
shared files
The server part of the file sharing protocol NFS is being used to export subsets of the file systems files.
ext3 File System
saved files
The logical volumes provided by the LVM are being used by the ext3 file system for saving file system structures and files.
Logical Volume Manager
LVM logical volumes
The logical volumes provided by the storage system are being used to generate logical volumes in the LVM of the storage client.
(Storage Client)
SAN
Storage Server
Storage System Controllers
logical volumes
Physical volumes (HDDs and SSDs) are used to generate logical volumes in a storage server. These LV are mapped to a storage client.

Figure 2-3 Storage server providing LV to file server, file server sharing files, file client accessing files

2.2 File serving with file sharing and file transfer protocols

One important aspect of distributed computing is being able to access files on a remote system. To the user, remotely located files should have a transparent image. That is, the user should not need to be concerned whether the file is on a remote or local system. This also means that only a single set of commands should be defined to control local and remote files. The file sharing protocols NFS and SMB use the concept of integrating access to shared parts of remote file systems into the local file system structures. The following list includes some of the design issues for file access protocols:

- **Access:** A file must appear to a user as a local file, regardless of whether it is on a remote or local machine. The path name to both local and remote files need to be identical.
- **Concurrency:** Access to a file must be synchronized so that a modification by one process does not cause an undesired effect to other processes that depend on that file.
- **Failure and recovery:** Both the client and the server must provide a recovery mechanism so that if one or both fails, the other carries out a set of expected behaviors.
- **Mobility:** If a file is moved from one system to another, the client must still be able to access that file without any alteration.
- **Scalability:** A wide range of network sizes and workloads must be supported.

Other ways of serving and transferring files are, for example, FTP, HTTP, SCP, and SFTP. They are designed with different goals in mind. For instance, HTTP is used mainly to serve content to World Wide Web (WWW) clients, whereas FTP is used to transfer files to and from remote locations without the goal of making the remote files available to processes of the local machine.

2.2.1 The Network File System protocol

The NFS is a distributed file system protocol developed by Sun Microsystems in 1984. It is designed to enable client computer users to access files over a network from remote systems as though they were accessing files in local storage.

Overview of the Network File System protocol

The portions of a file system tree that are made accessible are called *exports* on the server side and *mounts* on the client side. So basically, local physical file systems on an NFS server are made accessible to NFS clients. NFS, because of its roots, is found mostly on UNIX-like and Linux-like systems and is included in the base operating system or distributions. NFS is the standard for these systems to share and access files over a network. There are NFS implementations available for many other operating systems also.

NFS initially used only the stateless User Datagram Protocol (UDP) as the transport layer, but implementations that use the stateful Transmission Control Protocol (TCP) started to appear, too. NFS is based on Remote Procedure Call (RPC), which is an interprocess communication protocol that is used to start routines in another address space, such as in a remote computer system. NFS is described and defined in Request for Comments (RFC) documents and can be implemented at no charge.

> **Note:** RFC documents are sometimes called *Internet standards*. This is not accurate, although some RFCs become standard. An RFC might, for example, propose and describe Internet-related protocols and methods. Even though not officially a standard, it allows implementations of protocols to adhere to a specific RFC so that implementations based on a certain RFC can interact and be compatible with each other. The name *RFC* can be misleading. When an RFC is published, it is not changed. If there is a need for a change, a new RFC is published with a new RFC number.

Network File System Version 2

The first version of NFS that was available outside of Sun Microsystems was NFS Version 2 (NFSv2). It was published in 1989 in RFC 1094. NFSv2 needs a port mapper (`rpc.portmap`, `portmap`, `rpcbind`), which assigns the ports on which the NFS services listen. These ports are temporary and change when the NFS server restarts. More configuration is necessary for the ports that are used to become permanent, which makes NFS usable through firewalls that allow traffic to only specified ports.

NFSv2 has some limitations related to use cases, scalability, and performance. It supports files only up to 2 GB because of the 32-bit file size. The size of any single data transfer cannot exceed 8 KB, which hinders performance because of the high number of NFS requests. Another performance-limiting factor is that NFSv2 works only in a synchronous way. Data must be written to the file system by the NFS server before the write is acknowledged to the client (called *stable writes*). This limits the scalability of NFSv2.

NFSv2 does not support Kerberos authentication. To grant NFS clients access to exports, the access is granted to the computer system that the NFS client is running on. This means that any user on that system is able to access the exports. The limits for users are only file and directory permissions.

Note: Kerberos is an Internet protocol that is designed to add security to networked servers. It uses secret-key strong cryptography to deliver user authentication for networked applications. Kerberos was developed at the Massachusetts Institute of Technology (MIT) and is available at no charge. It is also included in commercial products.

Network File System Version 3

NFS Version 3 (NFSv3) was published in 1995 in RFC 1813. It still uses the port mapper. Version 3 removed some limitations and introduced some changes. The file offset is now 64 bit, which supports files larger than 2 GB. The maximum transfer size limit of 8 KB is gone; the client and server can agree upon the transfer size.

NFSv3 introduced an asynchronous method of operation (so called *unstable writes*). With unstable writes, the server does not need to acknowledge the write to the file system to the client immediately, so it can delay the write. The server must acknowledge the write only when it receives a commit request. This speeds up client writes and enables the server to efficiently write the data, mostly independent from the client's write operations. The NFSv3 client can detect uncommitted data in an error situation and can recover from that.

When NFSv3 was introduced, support for TCP as a transport layer increased. Some vendors had already added support for NFSv2 with TCP as a transport. Sun Microsystems added support for TCP as a transport at the same time that it added support for Version 3. Using TCP as a transport made using NFS over a wide area network (WAN) more feasible.

NFSv3 still grants access to the computer system, does not authenticate the user, and does not support Kerberos. This limits the use of NFSv3 to trusted networks.

Sun Microsystems handed over the maintenance of NFS to the Internet Engineering Task Force (IETF) before NFS Version 4 was defined and published.

Network File System Version 4

NFS Version 4 (NFSv4) was published in 2000 in RFC 3010. In 2003, it was revised and published in RFC 3530. It was also influenced by Andrew File System (IBM AFS™) and CIFS, which included performance improvements, mandated strong security, and introduced a *stateful* protocol. The NFSv4 server does not rely on a port mapper anymore. NFSv4 requires TCP as the transport layer; it listens on the well-known port TCP 2049. It is a stateful protocol, maintaining the state of objects on the server. This way, the server knows about the intentions of clients and some problems with stateless operation can be avoided.

NFSv4 improves performance and functionality with, for instance, file system semantics for the Microsoft Windows operating systems. It supports Windows access control lists (Windows ACLs), but it does not support the Portable Operating System Interface (POSIX) that is based on UNIX ACL, in addition to UNIX permissions. The security model in NFSv4 builds upon Kerberos, Low Infrastructure Public Key Mechanism (LIPKEY), and Simple Public Key Mechanism Version 3 (SPKM-3). The method that is used is agreed upon by the NFS client and NFS server. Other security mechanisms are negotiated, such as which encryption algorithm is being used.

NFS Version 4.1 (NFSv4.1) was published in 2010 in RFC 5661. It offers new features, such as the option to use clustered installations with parallel processing.

Note: There are a broad range of ACL formats, which differ in syntax and semantics. The ACL format that is defined by Network File System Version 4 (NFSv4) is called *NFSv4 ACL*. GPFS supports the NFSv4 ACL format, so this implementation is sometimes referred to as *GPFS NFSv4 ACL*. However, the implementation of NFSv4 ACLs in GPFS does not imply that GPFS or the Storwize V7000 Unified system supports NFSv4. The Storwize V7000 Unified system supports NFS version 2 (NFSv2) and NFS version 3 (NFSv3).

The Storwize V7000 Unified system stores all user files in GPFS format. Access protection in the Storwize V7000 Unified system is implemented in GPFS by using NFSv4 ACLs. It is enforced for all of the protocols that are supported by the Storwize V7000 Unified system: CIFS, NFS, FTP, HTTPS, and SCP.

2.2.2 The Server Message Block protocol

The Server Message Block (SMB) protocol is an application-layer network protocol for sharing file and printing resources in a distributed computing environment. Common Internet File System (CIFS) is considered the modern dialect of SMB. It enables the access to shared remote resources as though they were part of the local system. Most use of SMB involves computers that are running the Microsoft Windows operating system.

The SMB protocol was developed by IBM and further developed by Microsoft, Intel, IBM, 3Com, and others. An early mention of SMB is in the *IBM Personal Computer Seminar Proceedings* document from October 1984. In 1987, the SMB protocol was officially defined in a Microsoft and Intel document called *Microsoft Networks/OpenNET-FILE SHARING PROTOCOL*. Thereafter, it was developed by Microsoft and others. It has been used mainly with client computers that are running the IBM OS/2 OS versions and the Microsoft Windows OS, where the SMB protocol with the functionality to act as an SMB server and SMB client is built-in. SMB server and SMB client implementations for other platforms became available later and are in widespread use today.

Overview of the original Server Message Block protocol

Because SMB protocol has been developed further over the years, this results in many variants of the protocol, called *dialects*. It retains compatibility with earlier versions, with the ability to negotiate the dialect to be used for a session. The dialects are defined by a standard command set and are identified by a standard string, such as PC NETWORK PROGRAM 1.0 (the first dialect of the SMB protocol), MICROSOFT NETWORKS 3.0, DOS LANMAN 2.1, or NTLM 0.12 (the SMB dialect NTLAN Manager, which is designated as CIFS).

The name of the SMB protocol refers to the packets of data that are sent between the client and server (the Server Message Blocks). Each SMB contains a request from an SMB client or a response from an SMB server (client/server, request-response protocol).

The SMB protocol is used to provide access to resources and to access resources, called *shares*. Shares can be subsets of a file system and its contents, printers, serial ports, and some other resources. The term *share* is also used because the resources might be accessed by multiple clients (shared between them), with the protocol that is providing the locking mechanisms. The SMB protocol, as a presentation and application layer protocol, is implemented on top of various transport layer protocols, with NetBIOS over TCP/IP being the most common. In general, it is independent from the transport protocol, but it is connection-oriented. With changes, it can be implemented on top of a stateless protocol, such as UDP, also.

As stated before, the SMB protocol is compatible with the highest level of functionality of earlier versions, for both the server and the client. When an SMB client connects to an SMB server, they identify which dialect of the SMB protocol that they both understand and negotiate which one to use. The goal is to agree upon the dialect with the highest level of functionality that they both support.

There are two levels of access control. The first level is the *share level*, in which a client needs to provide the password for the share. The second level is the *user level*, in which the client authenticates with the server by using a user name and password. When the client is authenticated, the server sends a user ID (UID) to the client, which uses this UID in all later SMBs. Thereafter, the access to shares (which are not protected at the share level) is possible. For file serving, the SMB client can now open, read, write, and close files by sending the appropriate SMBs to the SMB server.

SMB permits locking files and records. To enhance performance, a mechanism called *opportunistic locking* (*oplock*) was introduced. It enables SMB protocol clients to cache data. This mechanism can lead to data loss during connection failures or server failures. To prevent data loss in case of failures, an option to disable opportunistic locking is provided in the SMB protocol implementation. Disabling oplocks also removes the performance benefits. Figure 2-4 shows how an SMB server manages oplocks:

1. SMB client 1 requests oplock for file 1 from SMB server.
2. SMB server grants oplock for file 1 to SMB client 1 and starts caching data for file 1.
3. SMB client 2 requests access to the file 1.
4. SMB server breaks oplock for file 1, SMB client 1 writes cached data to the SMB server.
5. SMB server grants access to file 1 to SMB client 2.

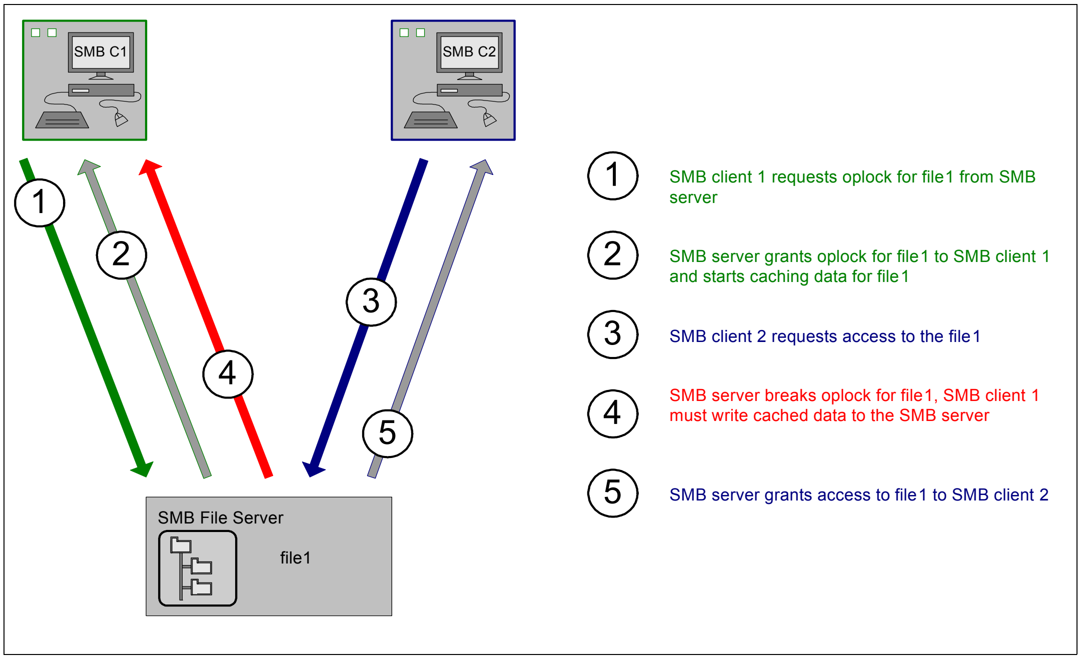

SMB C1
SMB C2
1
2
3
4
5
SMB File Server
file1
1 SMB client 1 requests oplock for file1 from SMB server
2 SMB server grants oplock for file1 to SMB client 1 and starts caching data for file1
3 SMB client 2 requests access to the file1
4 SMB server breaks oplock for file1, SMB client 1 must write cached data to the SMB server
5 SMB server grants access to file1 to SMB client 2

Figure 2-4 SMB server managing oplocks

The SMB protocol enables you to set attributes for files, directories, and extended attributes, too. It also supports ACLs.

What CIFS is and what is it not: CIFS is *not synonymous* with the SMB protocol. It is a *dialect* of the SMB protocol.

Microsoft started to call its versions of the SMB protocol the *Microsoft SMB protocol*. The first dialect of the Microsoft SMB protocol was the NT LAN Manager dialect. This dialect of the SMB protocol was based on the implementation of the SMB protocol in the Microsoft NT4 and Windows 2000 (NT5) OS. Microsoft proposed this specific dialect of SMB to the Internet Engineering Task Force (IETF) to become a standard with the name *CIFS*. During this time, the terms SMB and CIFS started to be used interchangeably, driven by Microsoft, because many expected that CIFS would become a standard and would be the successor (rather than just a dialect) of the SMB protocol. But CIFS did not become a standard and has not been published as an RFC document, either. In 2002, the Storage Network Industry Association (SNIA) CIFS Work Group published a document for CIFS with the title *Common Internet File System (CIFS) Technical Reference Revision: 1.0*, which does not constitute either a standard or specification, nor does it claim to be.

There is no real specification (documented in an RFC, for instance) that would enable developers to use to implement the SMB protocol (to specify, or spec) and be interoperable with various other implementations (that would adhere to the specification also). For instance, to be interoperable with the SMB protocol implemented in Microsoft products, developers must adjust to the changes that Microsoft might make.

After CIFS (the NT LAN Manager dialect of the SMB protocol), the Microsoft SMB protocol was developed further, and the term *SMB protocol* is being used by Microsoft and others. The extended version of CIFS is the Server Message Block (SMB) Version 1.0 Protocol (SMB 1). The most recent development, the complete redesign of the SMB protocol, is officially called the *Server Message Block (SMB) Version 2 Protocol*.

For SMB 1, the authentication support is enhanced. It is now compatible with the Generic Security Services application programming interface (GSS API). It supports Kerberos authentication through the GSS API.

It is now possible to query for older file versions, in case that is supported by the file system. Other enhancements make the protocol more efficient, such as the implementation of SMB server-side only operations (without the need to transfer the file to the SMB client and back to the SMB server). For example, for an SMB client-initiated copy operation from one directory to another, it makes no sense to read and write the file (transfer the data back and forth over the network, wasting network resources) because it still is done in CIFS. Concerning the use of TCP as the transport protocol, in addition to NetBIOS over TCP, SMB 1 supports running directly on TCP (Direct TCP), without the need for NetBIOS.

Quotas can be used by the SMB 1 client to limit file space that is used if the SMB server supports quotas.

Kerberos and GSS API: The Kerberos API is not standardized, but Kerberos 5 includes a GSS API implementation.

GSS API is an application programming interface (API) to enable software vendors to implement security-related services with a common API instead of supporting each other directly. GSS API is an IETF standard. With GSS API being implemented in the SMB 1 protocol, Kerberos can be used for authentication of SMB 1 clients.

SMB Version 2 protocol

All versions of the SMB protocol, including CIFS and SMB 1, are evolutionary changes and enhancements to the original SMB protocol. The Server Message Block (SMB) Version 2 Protocol (SMB 2), introduced in 2006 by Microsoft, is a newly developed file sharing protocol. It features a different set of commands, but is based on the concepts of the original SMB protocol. SMB 2.1 was introduced with Microsoft Windows 7 and Windows Server 2008 R2.

SMB 2 reduces complexity and increases efficiency, so it improves performance (especially for high-latency networks). Also, it is more scalable and introduces other enhancements, such as connection-error handling, improved message signing, and support for symbolic links. The SMB 2 protocol supports only TCP as the transport protocol by using either Direct TCP or NetBIOS over TCP.

SMB 3.0 (previously SMB 2.2) is the latest version and was introduced with Windows 8 and Windows Server 2012. To be compatible with older SMB versions (including the original SMB protocol versions), the older SMB protocol is used to negotiate the SMB version that is to be used by the SMB client and the SMB server. Although the protocol is proprietary, its specification was published to allow other systems to interoperate with Microsoft operating systems that use the new protocol.

SMB 3.0 also includes several changes that add functions and improve performance over SMB 2.

Note: At the time of publishing the Second Edition of this book, the Storwize V7000 Unified Disk System supports SMB 2.

2.2.3 File Transfer Protocol

The File Transfer Protocol (FTP) is an open systems interconnection (OSI) application layer protocol that is designed to transfer files between computer systems. It was initially written by Abhay Bhushan and was developed further and defined in RFC documents. FTP is a client/server protocol with the *FTP server* providing access to files (and file space) and *FTP clients* browsing the content that is made available, plus downloading files from and uploading files to the FTP server.

Historically, FTP was command-line based, but graphical user interface (GUI) implementations across many OSs became available and are in widespread use. FTP clients might mimic the look and behavior of file managers or file browsers in operating systems' GUIs, but FTP is not designed to integrate transparently into the representation of file system trees as file access protocols such as NFS and SMB do. There is no share to connect to and no export to mount. Instead, an FTP client connects to an FTP server and must either authenticate or, if permitted, connect as *anonymous* (no authentication needed). Then, the client can browse the FTP repository and transfer files. When files have been downloaded to the FTP clients, users of the client's system can use the files from within the file system (as opposed to the file access protocols NFS and SMB, which are used to work with files that are still in the file system of the remote system).

FTP uses TCP with the server by using two well-known ports, the *control* (or command) port TCP 21 for commands and authentication, and the *data* port TCP 20 for data transfers.

There are two modes of operation with FTP, *active mode* and *passive mode*. With active mode, the FTP client uses a random unprivileged port (above TCP 1023) to connect to the FTP server's control port (TCP 21, where the FTP server is listening for incoming FTP client requests). Using this control connection, the FTP client informs the FTP server about its *own* listening port for the transfer of data. The client-side listening port that is known to the FTP server is not a well-known port. Instead, it is the random port that the FTP client used to connect to the FTP server, plus 1 (the FTP client tells the FTP server only the port; it does *not initiate* the data transfer connection). Then, the server connects with its data port (TCP 20) to this connection-specific port of the client. This data connection is *initiated by the server* and the client side is *listening*, a behavior that is normally expected from a server. This mode of operation was called *active mode* in retrospect. It is the FTP server that *actively* tries to establish the data transfer connection to a client.

When servers or daemons “listen” to ports for connection attempts by clients from outside a firewall-secured network, they need these ports opened or forwarded by firewalls or they would wait indefinitely. Therefore, firewalls are usually configured to allow incoming traffic to servers on specific, well-known ports. The behavior of the FTP protocol in active mode, where the FTP client is listening to a random and temporary non-privileged port, leads to implications with client-side firewalls. Connection attempts from outside, if not explicitly allowed, are usually blocked.

To overcome this issue, the *passive mode* was introduced to FTP. In this mode, the FTP client uses two non-privileged ports above port TCP 1023, usually consecutive ones, to initiate two connections to the FTP server. It uses the first port to establish the control connection to the FTP server control port, TCP 21. It tells the FTP server to use passive mode (the FTP server listens on a port for the data transfer connection). The server-side data port is a non-privileged port above port TCP 1023 on the server side, randomly chosen by the FTP server. The FTP server sends the FTP client the information about which port the server uses to listen. With this knowledge, the FTP client can establish the data transfer connection from its second port to the FTP server listening port. Therefore, only the client is initiating connections, and the client-side firewall is likely to let the connections pass.

The server-side firewall must be configured to allow the incoming connection attempts to the non-privileged ports to pass, which is the drawback to that method from the server-side point of view. The issue is that the firewall must be configured to allow connection attempts to the ports above 1023. One way to reduce that problem is with FTP servers that support limiting of listening ports to a range of ports. Only these ports need to be configured by the firewall to allow incoming connection attempts.

2.2.4 Hypertext Transfer Protocol

The Hypertext Transfer Protocol (HTTP) is an OSI application layer client/server networking protocol. It is an IETF standard, and HTTP Version 1.1 is the most recent. HTTP is used to transfer files between HTTP servers and HTTP clients. The HTTP server software makes the content available to the HTTP clients, using port TCP 80 by default (other TCP ports can be used, but they must be specified by the HTTP client). The HTTP client software can have functionality to interpret these files and display the result as layout. Such HTTP client software is commonly called a *web browser*. HTTP is a core technology for the WWW.

To encrypt the requests by a client and the actual content while in transit, the Secure Sockets Layer (SSL) and its successor, Transport Layer Security (TLS) can be used. This is called *HTTP Secure* (*HTTPS*) or *HTTP over SSL/TLS*. HTTPS uses port TCP 443 by default. HTTPS is specified in RFC documents.

2.2.5 Secure Copy Protocol

The Secure Copy Protocol (SCP) is a client/server protocol that is used by SCP clients to copy files to and retrieve files from SCP servers. It is also possible to initiate a transfer of files between two remote systems. It uses the Secure Shell infrastructure for encrypted transfer. Usually, the SSH daemon provides the SCP function. It might act as both the SCP client and the SCP server. The port that is used is TCP 22. There is no official standard for SCP.

To connect to a Storwize V7000 Unified system environment using SCP, a prerequisite is that an SCP client is installed, available, and functioning properly. Windows clients do not have an SCP client that is installed by default, so it must be installed before this protocol can be used.

2.2.6 Secure Shell File Transfer Protocol

Secure Shell File Transfer Protocol (SSH FTP), also sometimes referred to as *Secure FTP* (SFTP), is a file transfer protocol that was developed in the context of SSH2 and is the standard file transfer protocol to use with SSH2. It runs over a secured connection that SSH provides. It is not equal to simple FTP or to FTP over SSH. It is a newly developed protocol. It provides functionality similar to FTP as compared to the SSH-based SCP protocol.

3

Architecture and functions

In this chapter, we provide an overview of the architecture of the IBM Storwize V7000 Unified Disk System and its functions.

3.1 High-level overview of Storwize V7000 Unified system

To serve logical volumes and files, the hardware and software to provide these services are integrated into one product. Viewed from its clients, one part of the Storwize V7000 Unified Disk System is a storage server and the other part is a file server; therefore, it is called *Unified*.

The Storwize V7000 Unified system is a single, integrated storage infrastructure with unified central management that simultaneously supports Fibre Channel, IP Storage Area Networks (iSCSI), and network-attached storage (NAS) data formats and is centrally managed.

3.1.1 Storwize V7000 Unified storage subsystem: the Storwize V7000

The V7000 Unified system uses internal storage to generate and provide logical volumes to storage clients, so it is a storage system. It can manage the virtualization of external storage systems also.

The storage subsystem consists of the hardware and software of the Storwize V7000 storage system (Storwize V7000). Initially, it runs the SAN Volume Controller and Storwize V7000 Version 7.1 (at the time of writing).

The storage subsystem is used for these functions:

- Provision of logical volumes to external storage clients
- Provision of logical volumes to the internal storage clients, the file modules

3.1.2 Storwize V7000 Unified system file server subsystem: The file modules

In addition to providing logical volumes, the Storwize V7000 Unified Disk System provides access to file system space and to files in those file systems. It uses file sharing protocol, file access protocols, and file transfer or file copy protocols, so it acts as a file server.

The file server subsystem of V7000 Unified system consists of two IBM Storwize V7000 file modules. These file modules perform the functions of the IBM Storwize V7000 Unified system software, initially running Version 1.4.2 (at the time of writing).

The file modules of the V7000 Unified system are internal storage clients of the system. They use the logical volumes that are provided by the system to save files and share files to file clients. The base operating system (OS) of the file modules is RedHat 6.1. They use a distributed file system, the IBM General Parallel File System (GPFS), to store and retrieve files. To make the content of the GPFS accessible to file clients, the file modules use the following file sharing and file access protocols:

- File Transfer Protocol (FTP)
- Hypertext Transfer Protocol Secure (HTTPS)
- Network File System (NFS)
- Secure Copy Protocol (SCP)
- Secure FTP (SFTP)
- Server Message Block (SMB)

Figure 3-1 on page 21 shows a high-level system diagram of the V7000 Unified system with a virtualized storage system, a storage client, and a file client.

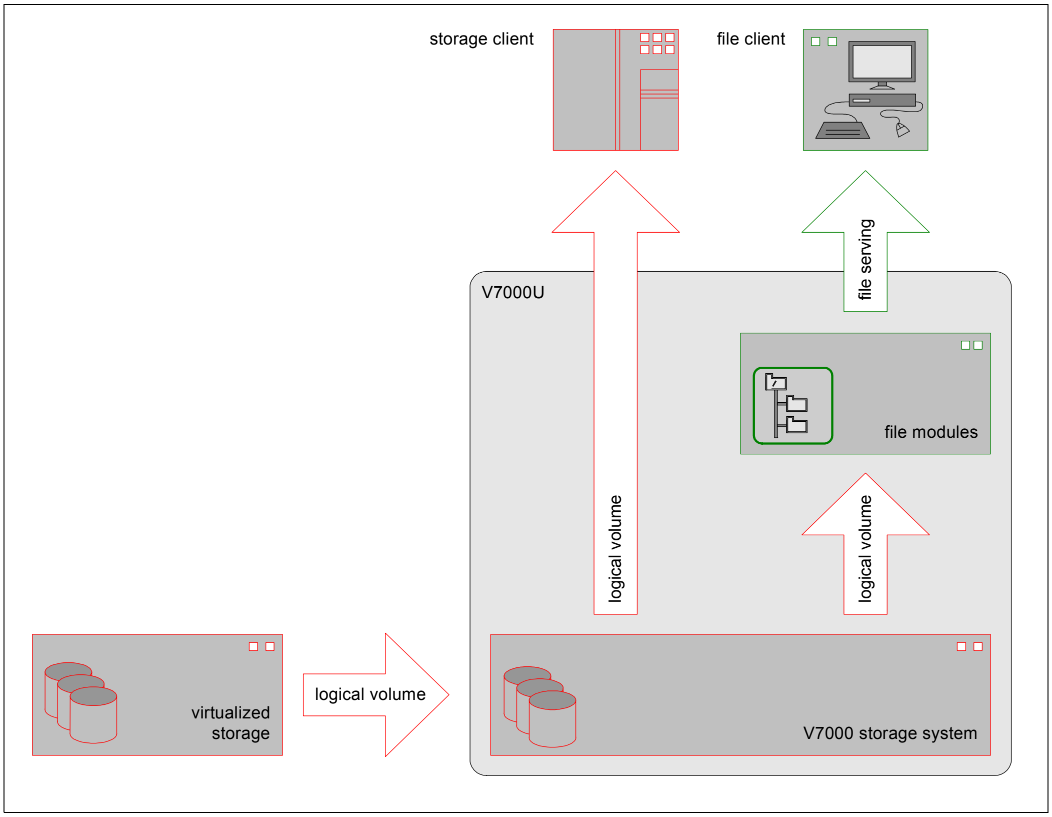


Figure 3-1 Storwize V7000 Unified system, high-level system diagram

3.2 Storwize V7000 Unified system configuration

The Storwize V7000 Unified Disk System consists of a single IBM Storwize V7000 control enclosure, 0 - 9 Storwize V7000 expansion enclosures (storage server subsystem), two file modules (file server subsystem), and the connecting cables. You can add up to three control enclosures for more I/O groups. Each additional control enclosure, together with the associated expansion enclosures (up to nine per control enclosure), provides a new volume I/O group. The file modules remain directly connected to the original control enclosure, which presents I/O group 0. That is the default configuration. Plan to have only block volumes in the new I/O groups. File volumes that are created for you when a new file system is created must continue to be in I/O group 0.

> **Note:** Adding more file modules to the system is not supported.

3.2.1 Storwize V7000 Unified storage subsystem configuration

The Storwize V7000 control enclosure houses two node canisters, the redundant power supplies (764 W with battery backup) and the two types of internal drive bays:

- Enclosure for 12 3.5-in. drives
- Enclosure for 24 2.5-in. drives

The expansion enclosures contain two Switched Bunch of Disks (SBOD) canisters of either type (which can be mixed freely) and two 500 W power supply units (PSUs). The basic Storwize V7000 Unified system supports up to 120 3.5-in. drives, 240 2.5-in. drives or a mixture of both. The V7000 Unified system supports the same dual-port 6 Gb serial-attached SCSI (SAS) hard disk drives (HDDs) and solid-state drives (SSDs) as the Storwize V7000:

- 146 GB 2.5-inch 15k RPM SAS HDD
- 300 GB 2.5-inch 15k RPM SAS HDD
- 1 TB 2.5-inch 7.2k RPM NL HAS HDD
- 200 GB 2.5-inch SSD (E-MLC)
- 400 GB 2.5-inch SSD (E-MLC)
- 800 GB 2.5-inch SSD
- 1.2 TB 6 Gb SAS 2.5-inch SFF HDD
- 300 GB 6 Gb SAS 10k RPM 2.5-inch SFF HDD
- 600 GB 6 Gb SAS 10k RPM 2.5-inch SFF HDD
- 900 GB 6 Gb SAS 10k RPM 2.5-inch SFF HDD
- 2 TB 3.5-inch 7.2k RPM HDD
- 3 TB 3.5-inch 7.2k RPM NL SAS HDD
- 4 TB 6 Gb NL SAS 3.5-inch 7.2k RPM LFF HDD

The Storwize V7000 subsystem of the V7000 Unified system is used to create virtual volumes and provides them as logical volumes to storage clients. The protocols used are FC, Fiber Channel over Ethernet (FCoE), and Internet Small Computer System Interface (iSCSI).

The following interfaces are on the two node canisters:

- Eight 2, 4, 8 Gb FC ports, which are fitted with shortwave transceivers, SFP+:
  - Four ports for external connectivity to FC storage clients
  - Four ports for internal connectivity to the file modules
- Each canister has two USB ports, for a total of four. The USB ports are used for installation and maintenance tasks.
- Four 1 GbE ports for external connectivity to iSCSI storage clients and for management (at least one 1 GbE port of each Storwize V7000 controller canister must be connected to the client network).
- Four 10 GbE ports for connectivity to iSCSI and FCoE storage clients. This is optional. It requires a Host Interface Module (HIM) in each Storwize V7000 controller canister.
- Four 4x 6 Gb SAS connectors for up to five expansion enclosures in the first SAS chain and for up to four expansion enclosures in the second SAS chain.

3.2.2 Storwize V7000 Unified system file server subsystem configuration

The file server subsystem of the V7000 Unified system consists of two file modules. They are IBM System x servers x3650 M3.

The following details are for one file module:

- Form factor: 2U
- Processor: Single four core Intel Xeon C3539 2.13 GHz, 8 G L3 cache (or similar)
- Cache: 72 GB
- Storage: Two 600 GB 10 K SAS drives, RAID 1
- Power supply units: Two (redundant), 675 W

The following interfaces are on one file module:

- Four 1 GbE ports
  - Two ports for external connectivity to file clients and file level remote copy
  - Two ports for the management network between the file modules for Unified system clustering
- Two ports 10 GbE for external connectivity to file clients and file level remote copy
- Two ports 8 Gb FC, one port is internally connected to each Storwize V7000 node canister

The internal and external interfaces of the V7000 Unified system, including the optional 10 GbE interfaces on the Storwize V7000 subsystem, are shown in Figure 3-2.

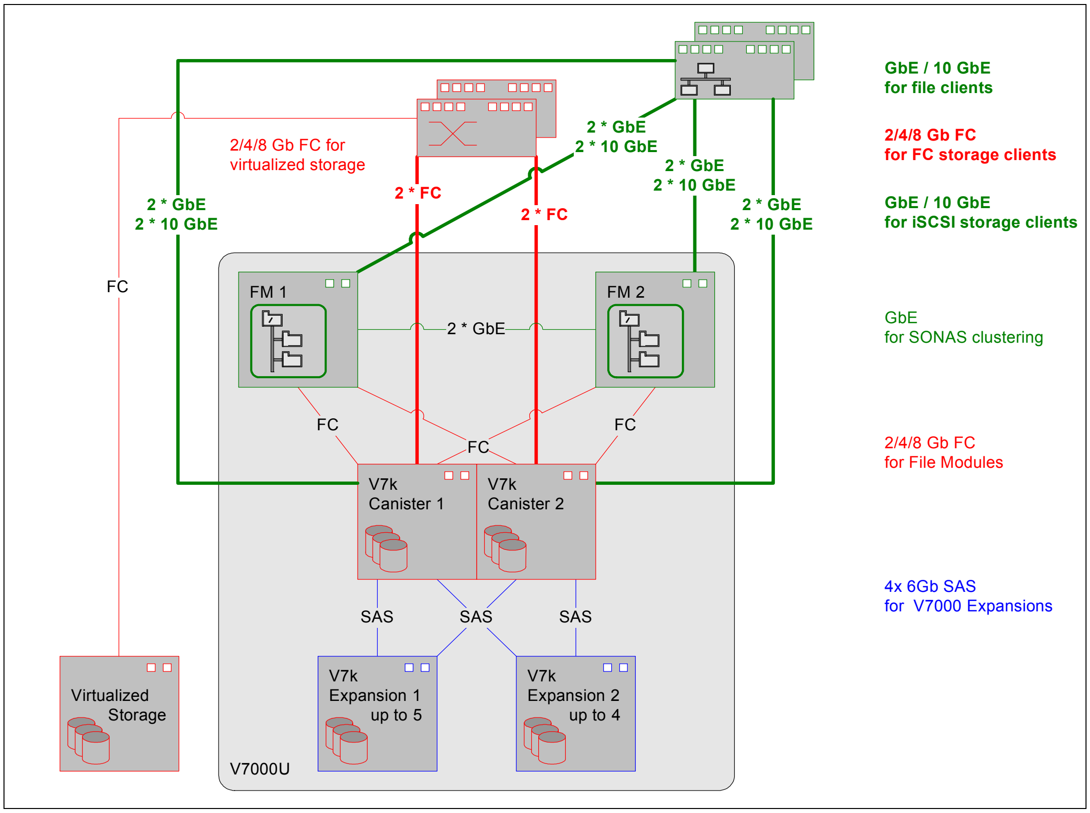


Figure 3-2 Storwize V7000 Unified system, internal and external interfaces

3.3 Storwize V7000 Unified system storage functions

The V7000 Unified system provides the storage service by mapping logical volumes to storage clients.

The storage functions implemented in and supported by the V7000 Unified system are the same as in the Storwize V7000 7.1 release. The Storwize V7000 uses RAID to protect against drive failures for the internal SAS storage. The other storage functions of the Storwize V7000 can be used on both internal and external virtualized storage systems.

The protocol used to obtain access to external storage systems is the Fibre Channel Protocol (FCP). The protocols that are used by storage clients to access the logical volumes that are mapped to them are FC, FCoE, and iSCSI. The externally connected FC ports on the Storwize V7000 controllers work as initiator and target.

For storage system virtualization, the Storwize V7000 supports up to 128 storage pools (MDisk groups). The extent sizes for these pools can be configured to be 16 MB through 8 GB. The system can manage 2^22 extents. For example, with a 16 MB extent size, the system can manage up to 16 MB X 4,194,304 = 64 TB. The maximum number of volumes for use by the file modules and for external storage clients is 8192, or 2048 per I/O Group. Volumes might be in striped, sequential, or image mode (for online volume migration). Volume mirroring can be used to protect against failures of MDisk groups (for instance, two external virtualized storage systems). Thin provisioning can be used with 64 KB or 256 KB grain size. Compressed volumes have a default grain size of 64 KB. The system can support up to 800 compressed volumes or 200 volumes per I/O Group. IBM Easy Tier® hybrid storage pools with two tiers can be used (SSD and HDD). The automatic extent level migration works based on access history in the previous rolling 24-hour period. Easy Tier is also supported by compressed volumes, and extent level migration is based on reads.

For more information about Storwize V7000 block storage, see the IBM Redbooks publication titled *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

3.4 Storwize V7000 Unified system file serving functions

The Storwize V7000 Unified system provides file sharing and file transfer services, broadly called *file serving*. The protocols that are implemented in the V7000 Unified system are described in Chapter 2, “Terminology and file serving concepts” on page 5. Here, we list specifics of the software architecture, operational characteristics, and components of the V7000 Unified system software.

The file-related functionality of the system is provided by the V7000 Unified system software, which runs on the file modules.

- The file modules provide the means for file serving capabilities (NFS, SMB, HTTPS, FTP, SCP, and SFTP).
- The Storwize V7000 back-end provides the connection to the storage.

Both file modules of the V7000 Unified system work together as a cluster, working in parallel to provide the functions of the Storwize V7000 software.

The V7000 Unified system software provides multiple elements and integrated components that work together in a coordinated manner to provide the file-related services to clients. Figure 3-3 on page 25 shows an overview of the software components that run on the V7000 Unified system file modules.

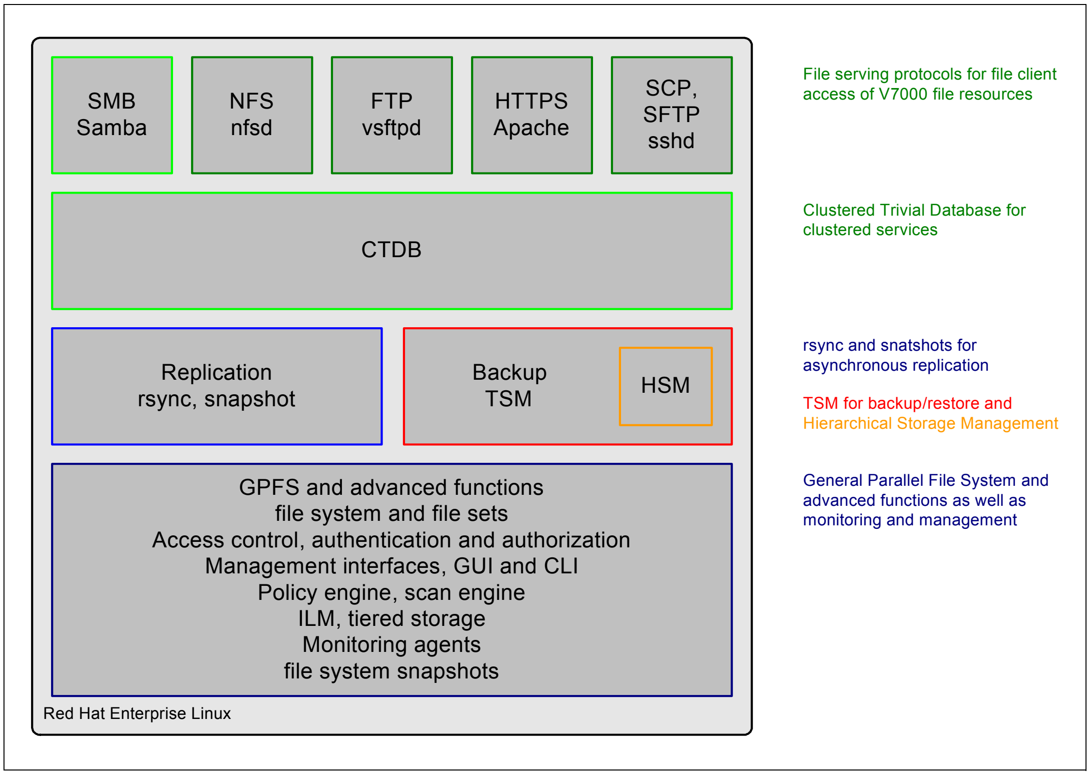

SMB
Samba
NFS
nfsd
FTP
vsftpd
HTTPS
Apache
SCP,
SFTP
sshd
CTDB
Replication
rsync, snapshot
Backup
TSM
HSM
GPFS and advanced functions
file system and file sets
Access control, authentication and authorization
Management interfaces, GUI and CLI
Policy engine, scan engine
ILM, tiered storage
Monitoring agents
file system snapshots
Red Hat Enterprise Linux
File serving protocols for file client access of V7000 file resources
Clustered Trivial Database for clustered services
rsync and snatshots for asynchronous replication
TSM for backup/restore and
Hierarchical Storage Management
General Parallel File System and advanced functions as well as monitoring and management

Figure 3-3 Overview of Storwize V7000 Unified system file modules software

The software that runs on the Storwize V7000 Unified system file modules provides integrated support for policy-based, automated placement and subsequent tiering and migration of data. You can provision storage pools and store file data either according to its importance to the organization or according to performance requirements. For example, you can define multiple storage pools with various drive types and performance profiles. You can create a higher performance storage pool with fast drives and define a less expensive (and lower performance) storage pool with higher capacity Nearline drives. Sophisticated policies are built into the V7000 Unified system, which can migrate data between pools based on many characteristics, such as capacity threshold limits and age of the data for use in an information lifecycle management (ILM) strategy. Policies are also used for compression by placing known compressed and uncompressed files into separate file system pools.

The V7000 Unified system software supports remote replication, point-in-time copying (file system-level snapshots), and automated storage tiering, all managed as a single instance within a global namespace. Asynchronous replication is specifically designed to cope with connections that provide low bandwidth, high latency, and low reliability. The asynchronous scheduled process picks up the updates on the source V7000 Unified system and writes them to the target V7000 Unified system by using snapshots and the *rsync* tool. The *rsync* tool is a standard Linux utility and is included in all popular Linux distributions, which includes the file modules.

3.4.1 Storwize V7000 Unified system file sharing and file transfer protocols

The network file sharing protocols and file transfer protocols that are supported by the Storwize V7000 Unified system are NFS, SMB, FTP, SCP, SFTP, and HTTPS. The V7000 Unified system uses GPFS to organize the data and save it to the storage of the Storwize V7000 part of the system. The cluster manager is used to provide cross-node and cross-protocol locking services for the file serving functions in NFS, SMB, FTP, SCP, SFTP, and HTTPS. The SMB file sharing function maps semantics and access control to the Portable Operating System Interface (POSIX)-based GPFS with a native NFSv4 access control list (ACL).

3.4.2 Storwize V7000 Unified system NFS protocol support

You can use the NFS client/server communication standard to view, store, and update files on a remote computer. By using NFS, a client can mount all or a portion of an exported file system from the server to access data. You must configure the NFS server to enable NFS file sharing in the Storwize V7000 Unified system. It provides several benefits:

- Supports NFSv2 and NFSv3 with NFSv4 partial implementation
- Supports normal NFS data access functions with NFS consistency guarantees
- Supports authorization and ACLs
- Supports client machine authorization through NFS host lists.
- Supports enforcement of access control lists (ACLs)
- Supports reading and writing of the standard NFSv3 and POSIX bits
- Supports the NFSv3 advisory locking mechanism
- Provides Semi-transparent node failover (the application must support network retry)

The V7000 Unified system software file system implements NFSv4 ACLs for security, regardless of the actual network storage protocol used. This method provides the strength of the NFSv4 ACLs even to clients that access the V7000 Unified system by the NFSv2, NFSv3, CIFS, FTP, and HTTPS protocols.

NFS protocol limitations and considerations

The flexibility of the NFS protocol export options allow for potentially unsafe configurations. Adhering to good practice guidelines reduces the potential for data corruption. The following considerations and limitations apply to V7000 Unified Disk System Version 1.4, although these might change with a future release.

- NFSv4 is not supported.
- NFS Kerberos functionality, for example SecureNFS, is not supported.
- Do not mount the same NFS export on one client from both Storwize V7000 Unified system file modules, because data corruption might occur.
- Do not mount the same export twice on the same client.
- Do not export both a directory and any of its subdirectories from a server if both are part of the same file system.
- Do not export the same file system, or the same file, through multiple exports to the same set of clients.
- A client must never access the same file through two different server:export paths. The client cannot distinguish that the two objects are the same, so write ordering is not possible and client-side caching is affected.

- In the V7000 Unified system, each export is assigned a new file system ID even if the exports are from the same file system. This process can lead to data corruption, which is why it is not good practice.
- Although use of nested mounts on the same file system is strongly discouraged, it is possible to create nested mounts by using the V7000 Unified system. If nested mounts are configured on the system, it is your responsibility to *exercise extreme caution* to avoid any possibility of corruption.
- POSIX ACLs for NFSv2 and NFSv3 are not supported on the Storwize V7000 Unified system.
- Mount the Storwize V7000 Unified system NFS exports only by using an IP address. Do not mount a V7000 Unified system NFS export by using a DNS Resource Record entry name. If you mount a V7000 Unified system NFS export by using a host name, ensure that the name is unique and remains unique, because this restriction prevents data corruption and unavailability.
- When an NFS client detects an NFS server change, such as an NFS server reboot or a new NFS server that is assuming NFS server responsibility from the previous NFS server, while writing data asynchronously, the NFS client is responsible for detecting whether it is necessary to retransmit data and for retransmitting all uncommitted cached data to the NFS server if retransmission is required.
- Storwize V7000 Unified system failover is predicated on this expected client behavior. For example, when an NFS client is writing data asynchronously to one of V7000 Unified system file modules, if the other file module assumes the NFS server role, the NFS client must detect the server change and retransmit all uncommitted cached data to the other file module to ensure that all of the data is safely written to stable storage.
- The Storwize V7000 Unified system uses the group IDs (GIDs) supplied by the NFS client to grant or deny access to file system objects, as defined in RFC 5531. When a user is defined in more than 16 groups, to get the access control that you want, you must appropriately define the groups that are transmitted from the client and appropriately define mode bits or ACLs on the V7000 Unified system.
- Files that are created on an NFSv3 mount on a Linux client are visible only through CIFS clients mounted on the same server node. CIFS clients that are mounted on different server nodes cannot view these files.

3.4.3 Storwize V7000 Unified system SMB and CIFS protocol support

The SMB and CIFS protocol functions of the Storwize V7000 Unified system are provided by an implementation of Samba. It is clustered by using the clustered trivial database (CTDB).

Samba is a software package that is available under the GNU General Public License (GPL) to provide file sharing functions to SMB file clients (such as Microsoft Windows operating systems). Samba provides the following functions:

- Name resolution
- Access control through authentication and authorization
- Integration with a Windows Server domain as a Primary Domain Controller (PDC) or as a domain member
- Option to be part of an Active Directory (AD) domain
- Service announcement for browsing of resources
- File sharing and print queue sharing

SMB and CIFS protocol access in the V7000 Unified system has been tested from SMB file clients that were running Microsoft Windows (2000, XP, Vista 32-bit, Vista 64-bit, 2008 Server), Linux with SMB client, Mac OS X 10.5, and Windows 7.

GPFS is a POSIX-compliant UNIX style file system. For SMB file clients, the V7000 Unified system maps UNIX ACLs to Windows access control semantics. A multitude of file access concurrency and cross-platform mapping functions are performed by the Storwise V7000 Unified system software, especially in the cluster manager. The V7000 Unified system implementation for SMB file access includes the following characteristics:

- File access by using SMB is supported only for file systems that are on internal storage of the V7000 Unified system storage part, not for external virtualized storage systems.
- SMB protocol version support:
  - The SMB 1 protocol is supported.
  - SMB 2.0 and later is not fully supported. See the note in “SMB protocol limitations” on page 29.
- SMB data access and transfer capabilities are supported by normal locking semantics.
- Consistent locking across platforms by supporting mandatory locking mechanisms and strict locking.
- User authentication provided through Microsoft Active Directory (AD) or through Lightweight Directory Access Protocol (LDAP).
- Consistent central ACL enforcement across all platforms.
- ACLs are enforced on files and directories and can be modified by using Windows tools.
- Semi-transparent failover if the SMB or CIFS implementation supports the network retry.
- Supports the `win32` share modes for opening and creating files.
- File lookup is not case-sensitive.
- Support for DOS attributes on files and directories.
- Archive bit, ReadOnly bit, system bit, and other semantics do not require POSIX attributes.
- MS-DOS and16-bit Windows short file names.
- Supports generation of 8.3 file names (eight-character file names with a three-character file extension.
- Notification of changes to file semantics to all clients that are in sessions with the file.
- Opportunistic locks and leases are supported for enabling client-side caching.
- Offline or de-staged file support (by the Storwize V7000 Unified system hierarchical storage management [HSM] function through IBM Tivoli® Storage Manager):
  - Offline files are displayed with the IBM hour glass symbol in Windows Explorer.
  - Recall to disk is apparent to the application. No additional operation is needed.
  - Windows Explorer can display file properties without the need to recall offline files.
- Storwize V7000 Unified system snapshots are integrated into the Volume Shadow Copy Service (VSS) interface.
  - Allows users with the appropriate authority to recall older file versions from the Storwize V7000 Unified system snapshots
  - Supports file version history for file versions that are created by Storwize V7000 Unified system snapshots

- The standard CIFS time stamps are available:
  - Created time stamp:
    - The time when the file was created in the current directory.
    - When the file is copied to a new directory, a new value is set.
  - Modified time stamp:
    - The time when the file was last modified.
    - When the file is copied elsewhere, it keeps the value in the new directory.
  - Accessed time stamp:
    - The time when the file was last accessed.
    - This value is set by the application, but not all applications modify it.

SMB protocol limitations

Consider the following SMB protocol limitations when you are configuring and managing the Storwize V7000 Unified system:

- Alternate data streams are not supported. One example is an NTFS alternate DataStream from a Mac OS X operating system.
- Server-side file encryption is not supported.
- Level 2 opportunistic locks (*oplocks*) are currently not supported. Therefore, Level 2 oplock requests are not granted.
- Symbolic links cannot be stored or changed and are not reported as symbolic links. But symbolic links created thought NFS are respected if they point to a target under the same exported directory.
- SMB signing for attached clients is not supported.
- SSL secured communication to Active Directory is not supported.
- Storwize V7000 Unified system acting as a distributed file system (DFS) root is not supported.
- Windows Internet Naming Service (WINS) is not supported.
- Retrieving quota information by using **`NT_TRANSACT_QUERY_QUOTA`** is not supported.
- Setting quota information by using **`NT_TRANSACT_SET_QUOTA`** is not supported.
- Managing the Storwize V7000 Unified system by using the Microsoft Management Console Computer Management Snap-in is not supported, with the following exceptions:
  - Listing shares and exports
  - Changing share or export permissions
- Users must be granted permissions to traverse all of the parent folders on an export to enable access to a CIFS export.

SMB-specific and CIFS1-specific limitations

- CIFS extensions for UNIX are not supported.
- You cannot create a shadow copy of a shared folder by using a remote procedure call (RPC) from a shadow copy client.
- Backup utilities, such as Microsoft Volume Shadow Copy Service, cannot create a shadow copy of a shared folder by using an RPC.

SMB2-specific limitations

- SMB 2.1 is not supported.
- The Storwize V7000 Unified system does not grant durable or persistent file handles.

Storwize V7000 Unified system FTP and SFTP support

The V7000 Unified system provides FTP access from FTP clients by using a vsftpd server. The following characteristics apply:

- Supports file transfer to and from any standard FTP client.
- Supports user authentication through Active Directory and LDAP.
- Supports enforcement of ACLs and retrieval of POSIX attributes. ACLs cannot be modified by using FTP because there is no support for the **`chmod`** command.
- Supports FTP resume for clients that support the network trying again if there is a node failover.
- Characters for file names and directory names are UTF 8-encoded.
- You cannot use the FTP protocol with the PuTTY utility to access the Storwize V7000 Unified system Service IP address, because PuTTY attempts to list files. That is not permitted by the V7000 Unified system Service IP for FTP service. However, PuTTY SCP is supported.
- When you are using FileZilla to view a directory listing on a Storwize V7000 Unified system, all file time stamps have a constant time offset. The time offset is caused by FileZilla automatically converting the time stamps from UTC to the local time zone. This conversion can be customized by adding the V7000 Unified system to the site manager and adjusting the server time offset in the Advanced tab.
- When opening the FTP session, specify one of the IP addresses that are defined on the Public Networks page of the GUI. When you are prompted for the Name, specify the user ID in this format: *domain\user*. When prompted for the password, enter the password.

Storwize V7000 Unified system HTTPS support

The V7000 Unified system supports simple read-only file transfer of files through the HTTPS protocol from any HTTP client by using the Apache HTTP server. All transfers are using HTTPS to provide access control. The following features are supported through HTTPS:

- Supports read-only file transfer of appropriately formatted files.
- Supports user authentication through Active Directory and LDAP.
- Supports enforcement of ACLs. ACLs cannot be viewed or modified with this protocol.
- On node failover during a file transfer, the transfer is canceled and must be tried again on the other file module. Partial retrieve is supported, which minimizes duplicate transfers in a failover situation.
- Characters for file names and directory names are UTF 8-encoded.

The Storwize V7000 Unified system software uses HTTP aliases as the vehicle to emulate the share or export concept. For example, share XYZ is accessible by `http://server.domain/XYZ`. The system redirects all HTTP access requests to HTTPS.

The web-based Distributed Authoring and Versioning (WebDAV) and the Representational State Transfer (REST) API are currently not supported in the V7000 Unified system. They are known requirements.

Storwize V7000 Unified system SCP and SFTP support

The Storwize V7000 supports the transfer of files between an SCP client and Storwize V7000 Unified system by using the `sshd` daemon. All of the default options that are implemented in this protocol are supported. Also, SFTP is available by `sshd` to transfer files in a manner similar to using FTP.

Storwize V7000 Unified system locking characteristics

POSIX byte range locks that are set by NFS clients are stored in IBM GPFS and Windows clients that are accessing the Storwize V7000 Unified system by using the SMB protocol to accept these POSIX locks. The mapping of SMB protocol locks to POSIX locks is updated dynamically on each locking change.

Unless the application specifically knows how to handle byte range locks on a file or are designed for multiple concurrent writes, concurrent writes to a single file are not desirable in any operating system.

To maintain data integrity, locks are used to guarantee that only one process can write to a file (or to a byte range in a file) at a time. Although file systems traditionally locked the entire file, newer ones, such as GPFS, support the ability for a range of bytes within a file to be locked. Byte-range locking is supported for both the SMB protocol and the NFS protocol, but this requires the application to know how to use this capability.

If another process attempts to write to a file (or a section of one) that is already locked, it receives an error and waits until the lock is released.

The Storwize V7000 Unified system supports the standard DOS and NT file system (deny mode) locking requests. These requests allow only one process to write to an entire file at a particular time, as well as byte-range locking. In addition, the V7000 Unified system supports the Windows locking known as *opportunistic locking* or *oplock*.

SMB protocol byte range locks set by Windows SMB file clients are stored both in the V7000 Unified system cluster-wide database, and by mapping them to POSIX byte range locks in GPFS. This mapping ensures that NFS file clients see relevant SMB protocol locks as POSIX advisory locks, and NFS file clients accept these locks.

3.4.4 Storwize V7000 Unified system cluster manager

The *cluster manager* is a core Storwize V7000 Unified system component, as shown in Figure 3-4 on page 32. The cluster manager coordinates and orchestrates V7000 Unified system functions and advanced functions. It runs on only one of the file modules and can fail over to the second file module if an issue occurs, such as a system hang or other failure.

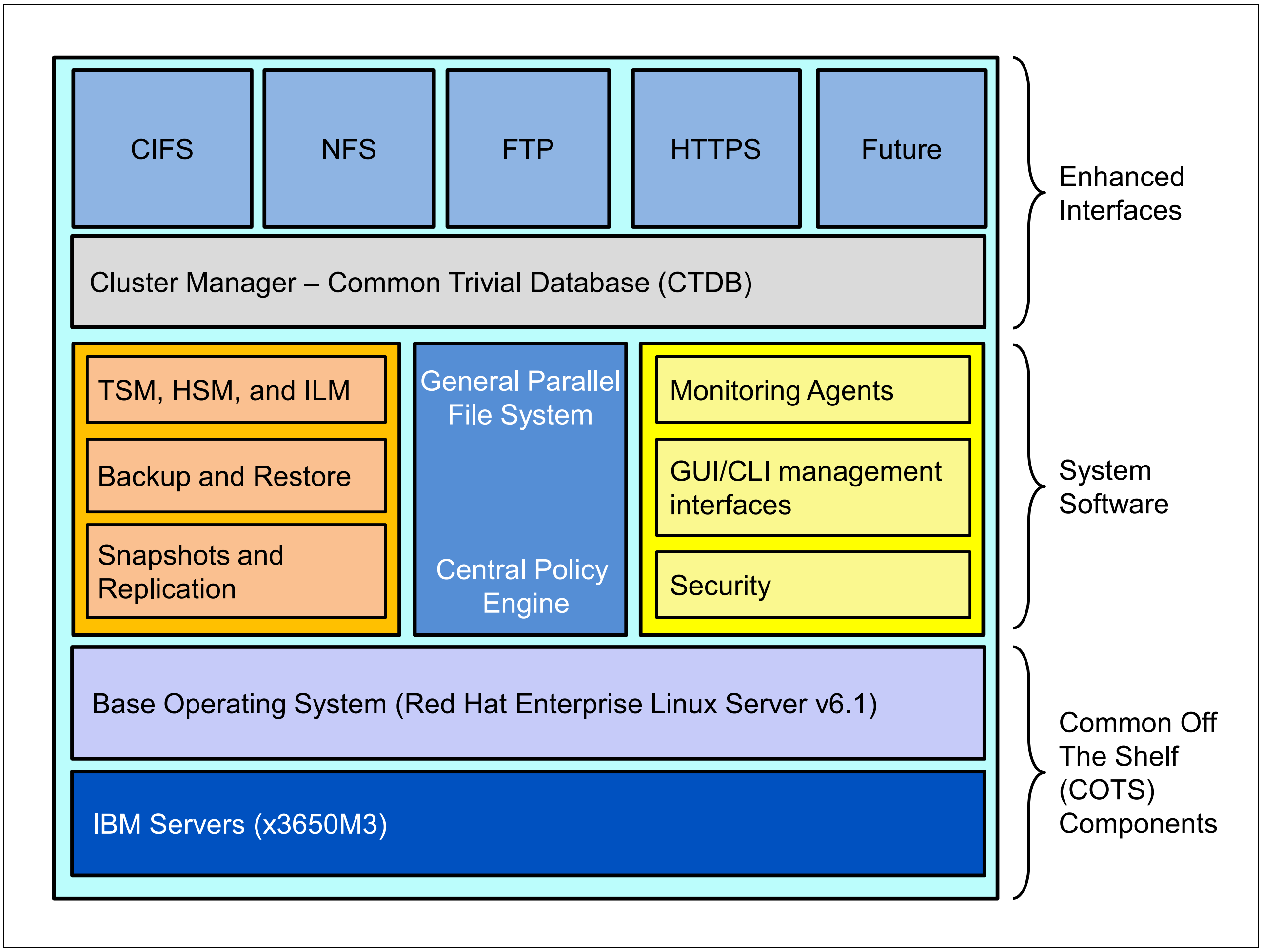

CIFS
NFS
FTP
HTTPS
Future
Cluster Manager – Common Trivial Database (CTDB)
Enhanced Interfaces
TSM, HSM, and ILM
Backup and Restore
Snapshots and Replication
General Parallel File System
Central Policy Engine
Monitoring Agents
GUI/CLI management interfaces
Security
System Software
Base Operating System (Red Hat Enterprise Linux Server v6.1)
IBM Servers (x3650M3)
Common Off The Shelf (COTS) Components

Figure 3-4 Storwize V7000 Unified system Cluster Manager

The Storwize V7000 Unified system cluster manager provides the clustered implementation and management of the file modules, including tracking and distributing record updates across both file modules in the cluster. It controls the public IP addresses that are used to publish the file services, and moves them as necessary between the file modules. By monitoring scripts, the cluster manager monitors and determines the health state of the file modules. If a file module has a problem, such as a hardware or software failure, the cluster manager dynamically migrates the affected public IP addresses and in-flight workloads to the other file module. It uses the *tickle-ACK* method with the affected clients so that they re-establish the TCP connection to the other file module. With this method, acknowledgement packets get exchanged, which allows for the remaining file module to send an appropriate reset packet so that clients know that the connection to the original file module is to be reset. Otherwise, clients would time out, possibly after a long time.

The Storwize V7000 Unified system software works in an active-active, high-available, and workload-sharing manner with the clustering functionality provided by the cluster manager. If a file module fails, the V7000 Unified system software automatically fails over the workload to the remaining file module. From a workload allocation standpoint, the V7000 Unified system uses the Domain Name System (DNS) to perform round-robin access to spread the workload as equally as possible on an IP address basis across the file modules.

The V7000 Unified system allocates a single network client to one file module. The V7000 Unified system software does not rotate a single client's workload across file modules. This process is unsupported by DNS or the SMB protocol. It would decrease performance because caching and read-ahead are done in the file module. It is for this reason that any one individual client is assigned to one file module during a session.

One of the primary functions of the cluster manager is to support concurrent access from concurrent users, spread across multiple various network protocols and platforms to many files. The V7000 Unified system software also supports, with the appropriate authority, concurrent read and write access to the same file, including byte-range locking. *Byte-range locking* means that two users can access the same file concurrently, and each user can lock and update a subset of the file.

All file accesses from the users to GPFS logically traverse the cluster manager. This logically implies that the cluster manager handles metadata and locking but does not handle data transfer. In other words, the cluster manager is not in-band for data transfer.

The clustered trivial database (CTDB) functionality provides important capabilities for the cluster manager to provide a global namespace to all users from any file access protocol, in which both file modules appear as a single file server. The CTDB also assures that all of the V7000 Unified system SMB components on both file modules can talk to each other in a high-performance manner and update each other about locking and other information.

3.4.5 Storwize V7000 Unified system product limits

Consider the following limits when you are configuring and managing the Storwize V7000 Unified system:

- Application installation that is not for the V7000 Unified system is not supported on V7000 Unified system file modules.
- The number of shares and exports that can be created per service (CIFS, NFS, FTP, SCP, and HTTPS) is limited to 1000 per protocol.
- If you are naming a share by using the command-line interface (CLI), you can use up to 80 characters. However, the graphical user interface (GUI) limits you to 72 characters.
- The "global" share name is reserved and cannot be used.
- Restricting ports by VLAN, service, or other criteria, is not possible.
- VLAN 1 is not supported for V7000 Unified system client traffic.

  This restriction is intended to prevent security exposure and reduce the probability of network configuration errors. VLAN 1 is used within the industry as the default or native VLAN. Many vendors use VLAN ID value 1 for management traffic by default. Configuring VLAN 1 as available within the network can be a security exposure because VLAN 1 might span large parts of the switched network by default. Common practice in the industry strongly discourages the use of VLAN 1 for user client traffic. Setting VLAN 1 for user client traffic can require explicit steps that differ by vendor and can be prone to configuration error.
- The READ_NAMED, WRITE_NAMED, and SYNCHRONIZE ACLs have no effect on the V7000 Unified system.
- Task scheduling has the following limitations:
  - Only schedules in equal space increments of 3 hours are supported.
  - Space increments other than 3-hour increments are not supported.

- For the following reasons, it is strongly suggested that only the UTF-8 character set is selected when you connect to V7000 Unified system CLI through Secure Shell (SSH), to the V7000 Unified system GUI through browser, or to V7000 Unified system shares and exports through NFS and FTP:
  - By selecting UTF-8 encoding in the SSH client for the connection to the V7000 Unified system CLI.
  - By selecting UTF-8 as locale for the connection to the V7000 Unified system in the FTP client.
  - All Storwize V7000 Unified system internal scripts and tools currently use `LANG=en_US.UTF8` and handle file names and directory names as though they contain only UTF-8 characters. Users can create files and directories by using different locales. For example, by using an external Linux client that is set to `LANG=ISO-8859-1`, `LANG=is_IS` or `LANG=de_DE`, or DBCS locales, such as `LANG=euc_JP`. The V7000 Unified system kernel NFS daemon treats file and directory names as a stream of bytes. So by using NFS mounts, you can theoretically copy those files and directories into the V7000 Unified system. The system kernel NFS daemon is not aware of locales, so it can copy files or directories with non-UTF-8 characters into the system.
  - UTF-8 uses the most-significant bit to encode characters that are not in the ASCII character set, which includes only characters with hexadecimal values 0x01 - 0x7f and decimal values 1 - 127. The UTF-8 encoding enforces that, if 1 byte in a file or directory name is greater than hexadecimal 0x7f, a second, maybe a third, and maybe a forth, byte must follow to complete a valid character. Therefore, files and directories that are created in a non-UTF-8 locale that have such a byte greater than 0x7f in their names are invalid when interpreted as UTF-8.
  - The CLI command input, the GUI, and some output, such as messages and log entries, currently require UTF-8 format only.
  - Multibyte Character Set (MBCS) support is limited to file and directory names. MBCS characters in object names (for example, user names) are not supported.
  - Current limitations:
    - Non-UTF-8 characters in file and directory names are not displayed correctly in the CLI, the GUI, messages, or log entries.
    - Non-UTF-8 file and directory names can be read only from clients that have the same language setting. For example, if an NFS client defined as ISO-8859-1 is used to create a file, a CIFS client or a different NFS client that is using UTF-8 cannot see or access that file.
    - Non-UTF-8 file and directory names cannot be backed up or restored.
    - Non-UTF-8 file and directory names cannot be specified when you are using the Storwize V7000 Unified system CLI, because the CLI interprets characters only as UTF-8. Attempting to restore a file name that contains a non-UTF-8 character does not restore the file with that file name because the byte representation is different.
    - Non-UTF-8 file and directory names might cause problems in other Storwize V7000 Unified system areas, including asynchronous replication, backup, and file access methods such as FTP, HTTPS, SCP, SMB, and NFS.
    - Non-UTF-8 file and directory names can be represented differently in different locales. Some locales will not be able to represent the byte combination at all, so they might treat the file names as invalid and will not process them correctly, if at all.
    - Object names that use multibyte non-UTF-8 characters can be limited to as few as 25% of the maximum number of characters allowed for the names of the same object that are composed of only 1-byte UTF-8 characters.

  - A directory that contains non-UTF-8 characters in its path cannot be the root of a share or export nor of a file set.
  - The following characters cannot be backed up or restored:
    `chr(24)`
    `chr(25)`
    `newline`
    / (slash)
- Wild cards and quotation marks (“) are *not* supported for backup. If you require that those characters be backed up, contact your IBM support representative.

Notes:

- Windows Service for UNIX (SFU) and Subsystem for UNIX based Applications (SUA) NFS does not handle non-ASCII UTF-8 file names correctly.
- Microsoft Internet Explorer (IE) does not correctly display non-ASCII characters in FTP file names and directory names that use UTF-8 for file name encoding. Such files or directories cannot be accessed. For example, IE does not correctly parse FTP directory listings that contain space characters within user or group names.
- The Storwize V7000 Unified system is tested for up to 6 VLAN-based subnets.
- For further configuration limits and restrictions, consult the Storwize V7000 Unified Support web page:

  http://www.ibm.com/storage/support/

  Also, locate the V1.4 Configuration Limits and Restrictions for IBM Storwize V7000 Unified web page:

  http://www-01.ibm.com/support/docview.wss?uid=ssg1S1004227

4

Access control for file serving clients

This chapter describes control of access to resources for file serving clients. *Access control* is a broad term about controlling who (user) or what (system) is granted access to which resources and can have many criteria. Two key concepts that are used to control access are authentication and authorization:

- *Authentication* provides and verifies credentials to ensure the identity of a user.
- *Authorization* grants a user access to a specific service or to specific resources, usually after successful authentication.

4.1 Authentication and authorization overview

The objective of authentication is to verify the claimed identity of users and components. Authentication methods include unique user IDs, keys, and digital certificates.

As the first access control process, authentication provides a way of identifying a user, usually by having the user provide credentials before access is granted. This can be done by entering a valid user name and valid password. Typically, the process of authentication is based on each user having a unique set of criteria for gaining access. The authentication server compares the entered authentication credentials to user credentials that are stored in a database. If the credentials match, the user is deemed to be identified and verified. If the credentials do not match, authentication fails.

After a successful authentication, the user can access services or resources based upon the associated authorization. Authorization might be based on a user ID (UID) and a matching UID in access control lists (ACLs) or other means of mapping a specific user to a specific resource or service.

4.1.1 UNIX authentication and authorization

UNIX authentication is system-based. Granting access to resources within the system or for shared resources that are being accessed from the system, the authorization is based on the UID and group identifier (GID). Users enter a user name and another credential, such as a password or a private Secure Shell (SSH) key), to log on to a UNIX system. The system looks up the user's UID in local files or an external directory service, such as a Lightweight Directory Access Protocol (LDAP) directory, and then verifies the received credential. The information for credential verification might be stored locally (for instance, hashes of passwords are stored in `/etc/shadow`, and public SSH keys are stored either in `.ssh/authorized_keys`) or in the external directory service, such as an LDAP directory.

When a user has successfully logged on to a UNIX system, that person is trusted (authenticated) on the system and on other systems that trust the particular system that the user just logged on to. For example, for file sharing through the NFS file sharing protocol, a Network File System (NFS) file server administrator creates an NFS export and grants access to the user's system. For NFS file access, the UNIX NFS client that is running on the user's system sends the user's UID with each file access request. The NFS service that is running on the NFS file server considers this UID as authenticated, assuming that the system correctly authenticated the user's UID. The UID is not used for another specific authentication by the NFS server, but for the *authorization* of the user's access to file resources. For example, if a user's system is authenticated but the UID of the user does not match the UID for resources on the NFS server system, the user still has no access to the resources of the NFS server because the user is not *authorized* to access them.

This means that to be able to access remote NFS resources, the user's system must be authenticated by the remote server, and the user must be authenticated by the local system and authorized to access its resources (by a successful logon to a system). Also, the UID that the NFS client of the user's system provides must match a valid UID on the NFS server's system for specific resources.

4.1.2 Windows authentication and authorization

Microsoft Windows authentication and authorization is session-based. A user logs on by using a user name and password on the Windows system. The system looks up the user's security identifier (SID) in the local Windows registry or on the Windows domain controller, such as an Active Directory server, and then verifies the received credential. The information for credential verification can be stored locally (in the Windows registry) or on the external Windows domain controller (the Active Directory server).

A user who is successfully logged on to a Windows system is trusted on this system. However, the user still must authenticate and be authorized to use other services provided by other network-connected systems before getting access. For instance, to access shares through the Server Message Block (SMB) protocol, an SMB file server administrator creates an SMB share and customizes the ACL (for authorization) of this share to grant the user access to the share.

For access to SMB shares, the SMB client that is running on the user's system sends an authentication request to an authentication server (for instance, AD). The authentication server checks whether the requesting user is allowed to use the service, and then it returns a session credential (which is encrypted with the key of the SMB server) to the SMB client. The SMB client sends the session credential to the SMB server. The SMB server decrypts the session credential and verifies its content. When the session credential is verified, the SMB server knows that the user was authenticated and authorized by the authentication server.

4.1.3 UNIX and Windows authentication and authorization

To provide heterogeneous file sharing for UNIX and Windows, the IBM Storwize V7000 Unified system must support the authentication methods for UNIX and Windows. The V7000 Unified system uses Windows authentication for incoming SMB connection requests and UNIX authentication for incoming NFS, Hypertext Transfer Protocol (HTTP), Secure Copy Protocol (SCP), and Secure File Transfer Protocol (SFTP) requests.

4.2 Methods used for access control

Depending on the dominating operating system environment, the size of the infrastructure, and other variables, different methods are used to control access.

4.2.1 Kerberos

Kerberos is a network authentication protocol for client/server applications. It uses symmetric key cryptography. The user password, in clear text format, is never sent over the network. The Kerberos server grants a ticket to the client for a short span of time. This ticket is used by the client of a service while it is communicating with the server to get access to the service, for instance, to access to SMB file server shares. Windows Active Directory authentication is based on Kerberos. MIT Kerberos is a free implementation of Kerberos protocol that is provided by the Massachusetts Institute of Technology.

For more information about Kerberos, see the MIT website:

http://web.mit.edu/kerberos/#what_is

4.2.2 User names and user IDs

UNIX system and UNIX based appliances, such as the Storwize V7000 Unified system, use user names and UIDs to represent users. The user name is typically a human-readable sequence of alphanumeric characters, and the UID is a positive integer value. When a user logs on to a UNIX system, the operating system looks up the UID and then uses it for further representation of the user.

User names, UIDs, and the mapping of user names to UIDs are stored locally in the `/etc/passwd` file or on an external directory service, such as AD, LDAP, or Network Information Service (NIS).

4.2.3 Group names and group identifiers in UNIX

UNIX systems use groups to maintain sets of users that have the same permissions to access certain system resources. A UNIX system also maintains group names and group identifiers (GIDs), which is similar to using user names and UIDs. A UNIX user can be a member of one or more groups, but one group is the primary or default group. UNIX groups are not nested. They contain users only, not other groups.

Group names, GIDs, the mapping of group names to GIDs, and the membership records of users in groups are stored either locally in the `/etc/group` file or in an external directory service, such as AD, LDAP, or NIS. The primary group of a user is stored in the `/etc/passwd` file or in an external directory service.

4.2.4 Resource names and security identifiers in Windows

Microsoft Windows refers to all operating system entities as *resources*, including users, groups, computers, and other resources. Each resource is represented by a security identifier (SID). Windows groups can be nested. For instance, one group can include one or more users and one or more groups. Resource names and SIDs are stored locally in the Windows registry or in an external directory service, such as Active Directory or LDAP directory.

4.2.5 UID, GID, and SID mapping in the Storwize V7000 Unified system

The Storwize V7000 Unified system stores all user data in the GPFS file system, which uses UIDs and GIDs for authorization. For SMB share access, the Storwize V7000 Unified system needs to map SIDs to UIDs and GIDs to enforce access control. NFS clients send the UID and GID of a user that requests access to a file. The Storwize V7000 Unified system uses the Linux default access control mechanism by comparing the received UID and GID with the UIDs and GIDs stored in an IBM General Parallel File System (GPFS).

The UIDs and GIDs that the NFS clients use must match the UIDs and GIDs that are stored in the GPFS. There is a requirement to allow the remapping of external UIDs and GIDs used by the NFS client to different UIDs and GIDs that are stored in the GPFS.

For HTTP, SFTP, and SCP access, the Storwize V7000 Unified system requires users to authenticate with a user name. The system needs to map the user name to one UID and one or more GIDs for GPFS access control.

When an SMB client that is using Microsoft Windows connect to the Storwize V7000 Unified system, it first contacts the Active Directory to check the user name and password combination. The UID and GID pair that is created is then stored in the `idmap` database in the

Storwize V7000 Unified system. The first time that a user logs in, the ID mapping is created. After that, it is picked up from the database directly.

For NFS access from UNIX clients, the UID is provided by the UNIX client. In there is mixed access from Windows and UNIX, Active Directory with Services for UNIX (SFU) can be used.

4.2.6 Directory services in general

Storing user and group information in local files works well for small organizations that operate only a few servers. Whenever a user is added or deleted, the group membership is changed or a password is updated. This information must be updated on all servers. Storing this information in local files does not scale for large organizations that have many users who need selected access to many servers and services.

Directory services enable you to store and maintain user and group information centrally on an external server. Servers look up this information in the directory server rather than storing the information in local files.

4.2.7 Windows NT 4.0 domain controller and Samba primary domain controller

A *domain* is a Windows NT concept in which a user might be granted access to several computer resources with the use of user credentials. A domain controller (DC) is a server that responds to authentication requests and controls access to various computer resources. Windows 2000 and later versions introduced Active Directory, which largely eliminated the concept of primary and backup domain controllers. Primary domain controllers (PDCs) are still used by customers. The Samba software can be configured as the primary domain controller and the client can run Samba on Linux. The Samba4 project has the goal to run Samba as an Active Directory server.

4.2.8 Lightweight Directory Access Protocol

The Lightweight Directory Access Protocol (LDAP) is a directory service access protocol that uses TCP/IP. LDAP was developed as a lightweight alternative to the traditional Directory Access Protocol (DAP). The *lightweight* in the name refers to the fact that LDAP is not as network intensive as DAP.

An LDAP directory is usually structured hierarchically as a tree of nodes. Each node represents an entry within the LDAP database. A single LDAP entry consists of multiple key/value pairs that are called attributes, and is uniquely identified by a distinguished name.

4.2.9 Microsoft Active Directory

Active Directory (AD) is a Microsoft created technology that was introduced in Windows 2000 and provides the following network services:

- Directory service, which is based on LDAP
- Authentication service, which is based on Kerberos
- Domain Name System (DNS) function

4.2.10 Services for UNIX and Identity Management for UNIX

Services for UNIX (SFU) is a Microsoft Windows component for Windows Server 2003 with AD. Identity Management for UNIX is used instead in Windows Server 2008 with AD, which provides interoperability between Microsoft Windows and UNIX environments. The Storwize V7000 Unified system uses it primarily for UID/GID/SID mapping.

4.2.11 Network Information Service

Network Information Service (NIS) is a directory service protocol for centrally storing configuration data of a computer network. NIS protocols and commands were originally defined by Sun Microsystems. The service is now widely implemented. Originally called Yellow Pages or YP, some of the binary names still start with yp. The original NIS design was seen to have inherent limitations, specifically in the areas of scalability and security. Therefore, modern and secure directory systems, primarily LDAP, are used as an alternative. The NIS information is stored in so called *NIS maps*, typically providing the following information:

- Password-related data similar to data stored in */etc/passwd*
- Group related data similar to data stored in */etc/group*
- Network configuration such as netgroups

4.2.12 Access control list in general

Generally, an access control list (ACL) is a list of permissions that is attached to a resource. An ACL describes which identities are allowed to access the respective resource (for instance read, write, execute). ACLs are the built-in access control mechanism of UNIX and Windows systems. Storwize V7000 Unified system uses a Linux built-in ACL mechanism for access control to files that are stored on GPFS.

4.2.13 GPFS NFSv4 ACLs

There is a broad range of ACL formats that differ in syntax and semantics. The ACL format that is defined by NFSv4 is also called *NFSv4 ACL*. GPFS ACLs implement the NFSv4 style ACL format, which is sometimes referred to *GPFS NFSv4 ACL*. The Storwize V7000 Unified system stores all user files in GPFS. The GPFS NFSv4 ACLs are used for access control of files that are stored on the Storwize V7000 Unified system.

Note: The implementation of NFSv4 ACLs in GFPS does not imply that GPFS or the Storwize V7000 Unified system supports NFSv4. NFSv4 support of the Storwize V7000 Unified system is planned for a future release.

4.2.14 POSIX bits

The POSIX bits of a file are a way to specify access permissions to files. UNIX file systems allow you to specify the owner and the group of a file. You can use the POSIX bits to configure access control for the owner, the group, and for all other users to read, write to, or execute the file. POSIX bits are less flexible than ACLs.

The change of the POSIX bits of a GPFS file system triggers a modification of its GPFS NFSv4 ACL. Because the Storwize V7000 Unified system uses GPFS NFSv4 ACLs for access control, the Storwize V7000 Unified system administrators and IBM service personnel should never change the POSIX bits of files that are stored in GPFS.

4.2.15 ACL mapping

GPFS NFSv4 ACLs and Windows ACLs are not compatible. For instance, Windows supports unlimited nested groups that are not fully supported by GPFS NFSv4 ACLs. The Storwize V7000 Unified system maps Windows ACLs on a best fit basis to GPFS NFSv4 ACLs, which results in some limitations. It is a known limitation that in this aspect the current Storwize V7000 Unified system is not fully compatible with Windows SMB file sharing.

4.3 Access control with Storwize V7000 Unified system

The authentication configuration of the Storwize V7000 Unified system consists of two elements, the configuration of a directory service and the refinement of the ID mapping. In the Storwize V7000 Unified system implementation, it is essential to define an initial owner when you are creating a share. Only this owner has initial access from a file client and can start to define directory structures and associated ACLs for all other designated users of this share. It cannot be displayed and listed afterward and cannot be changed if there is any data that is stored on the share.

4.3.1 Authentication methods supported

Storwize V7000 Unified system supports the following authentication methods:

- Active Directory
- Active Directory with Microsoft Windows Services for UNIX (SFU), and Identity Management for UNIX
- Samba Primary Domain Controller (PDC)
- LDAP
- LDAP with MIT Kerberos
- NIS
- Local authentication service that is configured internally on Storwize V7000 Unified system

Storwize V7000 Unified system uses the following other authentication elements within the methods supported:

- Netgroups: It is a group of systems that are used to restrict access for mounting NFS exports on a set of systems and deny mounting on the rest of the systems. The Storwize V7000 Unified system supports the netgroup being stored in NIS.
- Kerberos: The Storwize V7000 Unified system supports Kerberos with Active Directory (mandatory) and LDAP (optional).
- Secure Sockets Layer/Transport Level Security (SSL/TLS): These protocols are primarily used to increase the confidentiality and integrity of data that is sent over the network. These protocols are based on public-key cryptography and use Digital Certificate based on X.509 for identification.

With the Storwize V7000 Unified system, you can configure only one authentication method at one time, for instance AD. External authentication server needs to be configured separately and the Storwize V7000 Unified GUI or CLI does not provide any means to configure or manage the external authentication server. This is true even for the Kerberos server.

The Storwize V7000 Unified system provides server-side authentication configuration for various protocols, which include NFS, SMB, FTP, SCP, SFTP, and HTTP. For NFSv3, only the protocol configuration is performed. Kerberos has a few special steps to be performed on the V7000 for NFSv3, though. Because authentication happens on the NFSv3 client side, it needs to be configured on the client side mainly.

It is required that the V7000 Unified system is synchronized in time with the authentication servers. Authentication does not work if time is not synchronized. Authentication configuration does not ensure synchronization. Therefore, you need to ensure this manually.

4.3.2 Active Directory authentication

To use Active Directory (AD), the V7000 Unified system must be configured for and joined to the AD domain. This automatically creates the required computer account in the AD. The *public clustername* that is specified during installation is used as the computer account name. For Kerberos to work, it is important to always use this name for file sharing protocol access.

Authentication is provided for all supported file access protocols except NFS. Active Directory with Services for UNIX (SFU) must be configured for access through NFS for Windows Server 2003 with AD. On Windows Server 2008 with AD, the Identity Management for UNIX must be enabled in AD.

4.3.3 Active Directory with Services for UNIX or Identity Management for UNIX

Active Directory with SFU or with Identity Management for UNIX is the correct choice for clients with the following conditions:

- You use Windows Server 2003 or Windows Server 2008 with Active Directory to store user information and user passwords.
- You plan to use NFS.
- You plan to use asynchronous replication.

The primary Windows group that is assigned to an Active Directory user must have a group ID (GID) assigned. Otherwise, the user is denied access to the system. Each Active Directory user must have a valid UID and GID assigned to be able to mount and access exports.

The primary UNIX group setting in Active Directory is not used by the Storwize V7000 Unified system. The V7000 Unified system always uses the primary Windows group as the primary group for the user. This results in new files and directories that are created by a user through the SMB protocol being owned by their primary Windows group, not by the primary UNIX group. For this reason, it is best that the UNIX primary group is the same as the Windows primary group that is defined for the user.

It is difficult to add SFU at a later time, when data is already stored on the Storwize V7000 Unified system with AD. This is because the UIDs and GIDs used internally by GPFS must match the UIDs and GIDs stored in SFU. If conflicting UIDs and GIDs are stored in SFU, this is not possible. We therefore recommend to configure the Storwize V7000 Unified system with Active Directory and SFU at the time of installation, before the system is put into production.

The ACLs that are copied from the source Storwize V7000 Unified system to the target Storwize V7000 Unified system include UIDs and GIDs of the source V7000 Unified system. They are inconsistent with the UIDs and GIDs of the target V7000 Unified system.

To enable all NFS users of more than one V7000 Unified system to access all systems with the same identity, the authentication schema in Active Directory must be changed to UID or GID. Existing user entries must be edited (mapping from SID to UID or GID). You must add UNIX user information for new users while you are creating the user account in AD.

4.3.4 Samba primary domain controller authentication

NT4 PDC (primary domain controller) is the earlier domain controller concept that was used by Microsoft Windows NT and Windows 2000. This authentication method is no longer supported by Microsoft.

The open source Samba community developed the Samba PDC which supports this authentication type. Samba PDC emulates an NT4 PDC on Linux systems.

4.3.5 LDAP authentication

LDAP can be used in environments where Windows and UNIX clients are used. The Storwize V7000 Unified system supports LDAP with Kerberos for SMB protocol access. LDAP is not supported for secured NFS, FTP, HTTP, or SCP protocol access.

4.3.6 Network Information Service

Network Information Service (NIS) is used in UNIX-based environments for centralized user and service management. NIS keeps user, domain, and netgroup information. A *netgroup* is used to group client machine IP and host name, which can be specified while you are creating NFS exports. NIS is also used for user authentication for services, including SSH, FTP, HTTP, and more. The Storwize V7000 Unified system uses NIS for netgroup support and ID mapping. NIS default domain is used to resolve the netgroup information. You cannot change NIC default domain, but it is supported to get netgroup information from other domains. The NIS client configuration needs server and domain details of the NIS server.

Three different modes of NIS configuration are supported:

- NIS for netgroup and Active Directory or Samba PDC for authentication and Active Directory increment ID mapping.

  This is used for mixed environments with Windows and UNIX users. In this mode, the V7000 Unified system supports both the SMB and the NFS protocol and netgroups.

- NIS with ID mapping as an extension to Active Directory or Samba PDC and netgroup support.
- Plain NIS without any authentication, only for netgroup support (only NFS).

4.4 Access control limitations and considerations

Consider the following limitations when you are configuring and managing the Storwize V7000 Unified system.

4.4.1 Authentication limitations

For Active Directory with the SFU UID, GID, or SID mappings extension, consider these factors:

- Enabling SFU for a trusted domain requires a two-way trust between the principal and the trusted domain.
- To access the Storwize V7000 Unified system, users and groups must have a valid UID or GID assigned to them in AD. The allowed range is 1 through 4294967295. It is advisable to keep the lower range greater than 1024 to avoid conflict with the CLI users. Starting the command with a range of less than 1024 generates a warning message that asks for confirmation. Use the **`--force`** option to override it.
- For user access, the primary group on the Storwize V7000 Unified system is the Microsoft Windows primary group, not the UNIX primary group that is listed in the UNIX attribute tab in the user's properties. Therefore, the user's primary Microsoft Windows group must be assigned a valid GID.

For Active Directory with the NIS mappings extension:

- Because spaces are not allowed in UNIX names, implement the following conventions for mapping Active Directory users and groups to NIS:
  - Convert all uppercase characters to lowercase characters.
  - Replace every space character with the underscore character. For example, an Active Directory user named *CAPITAL Name* has the corresponding name, *capital_name*, on NIS.
- If Active Directory is already configured on the Storwize V7000 Unified system, you can use only the **`--idMapConfig`** option of the **`cfgad`** CLI command to change the high value of the range. The high value of the range can be changed only to a higher value. You cannot change the high value of the range to a lower value. You cannot change the low value of the range, and you cannot change the range size.

  For example, in the Storwize V7000 Unified system, if you used the **`cfgad`** CLI command with the **`--idMapConfig`** option to configure Active Directory and specify the value for the **`--idMapConfig`** option as 3000 - 10000:2000, you can use only the **`cfgad`** CLI command with the **`--idMapConfig`** option to increase the value of 10000 for the high value of the range. You cannot decrease the value of 10000 for the high value of the range. You cannot change the value of 3000 for the low value of the range, and you cannot change the value 2000 for the range size.
- You might need to make one of these changes:
  - Change from NIS ID mappings to Active Directory ID mappings
  - Change the ID mapping parameters of an already existing Active Directory configuration by using the **`--idMapConfig`** option of the **`cfgad`** CLI command
  - Change the low value of the range, decrease the high value of the range, or change the range size.

To make any of those changes, you must perform the following steps in this sequence:

a. Submit the **cleanupauth** Storwize V7000 Unified CLI command, and do not specify the **--idmapDelete** option.

b. Submit the **cleanupauth** Storwize V7000 Unified CLI command, and do specify the **--idmapDelete** option.

c. Submit the **cfgad** Storwize V7000 Unified CLI command with the options and values that you want for the new Active Directory configuration.

If you do not perform the preceding steps in sequence, results are unpredictable and can include loss access to data.

- UIDs and GIDs less than 1024 are denied access for the FTP, SCP, and HTTPS protocols for all of the supported authentication schemes, other than Active Directory with SFU.
- Authentication configuration commands stop and restart the SMB, NFS, FTP, SCP, and HTTPS services. This action is disruptive. Connected clients lose their connections, and file operations are interrupted. File services resume a few seconds after an authentication configuration command completes.

4.4.2 Authorization limitations

When you are managing authorization, the following Storwize V7000 Unified system implementation details apply:

- When a child file or child directory is created, the ACL that the file is initially assigned depends on the ACL type, the file system settings, and the ACL of the parent directory. Depending on these variables, the results in GPFS might be slightly different than in Microsoft Windows. For example, if the parent directory is set to have two ACLs, with full access for owner and for everyone, the Windows default is to create two ACLs for the child: Allow full access for owner and allow full access for everyone. The GPFS default creates six ACLs: Allow and deny ACLs for owner, group, and everyone.
- The special permissions Write Data/Create File and Create Folder/Append Data cannot be set separately for files. If either of these permissions is set, both are set. Enabling one always enables the other, and disabling one always disables the other. For directories, they can be set separately on condition that these access control entries (ACEs) are not inherited by files. You can configure two separate ACEs:

  – The ACE that is inherited by files has both special permissions enabled or both disabled

  – Another ACE that is inherited by directories where one of the preceding special permissions is enabled and the other disabled.

  In this case, the “Apply to” field of the Permission Entry panel can contain these values:

  – This folder only
  – This folder and subfolders
  – Subfolders only

  If you attempt to specify these values:

  – This folder, subfolders, and files
  – This folder and files
  – Files only

  The following security message is displayed:

  ```
  Unable to save permission changes on folder. The parameter is incorrect.
  ```

- The BypassTraversalCheck privilege that can be used on Windows servers is not supported on the Storwize V7000 Unified system. To read the content of a subdirectory, a user must have READ permission in the ACL of this subdirectory and must have traversal permission (SEARCH in Windows, execute in POSIX) for all of the parent directories. You can set the traverse permission in the Everyone group ACE at the share root and have this privilege inherited by all subdirectories.
- ACLs can be managed through NAS protocols by an authorized user.
  - The default ACL on a file system root directory (700 root root) prevents users from accessing the file system.
  - ACL inheritance stops at file set junction points. New file sets always have the default ACL: 700 root root.
- For security reasons, creating an export does not allow the setting of an owner for existing directories. The owner can be changed only if the directory is empty. When a directory contains files or directories or linked file sets, you cannot change the owner when you are creating an export.
- If the Owner option is omitted when you are creating an export for a nonexistent directory, the directory is created, and it inherits the ACL if inheritance is configured. If ACL inheritance is not configured, a new export for a nonexistent directory is assigned the default ACL, which is 700 root root.
- If the directory is empty, the owner can be changed by deleting and re-creating the export with the owner option.
- Using POSIX commands, such as **`chmod`**, overwrites any previous ACLs and creates an ACL with entries only for owner, group, and everyone.
- When you create an export, you can create multiple levels of directories if they do not exist. However, if multiple directory levels are created, the owner is set only for the leaf directory. All other directories are created as owner root and can be accessed only if an appropriate ACL inheritance was configured.
- The Storwize V7000 Unified system automatically creates Owner, Group, and Everyone special entries in the ACL to support interoperability with NFS. This is unlike Microsoft Windows, which does not use these special ACL entries. An inherited ACL might look different from the parent ACL because the owner and group entries changed. Other ACL entries are not affected by this special behavior.
- Only the root user can change ownership to a different user.

5

Storage virtualization

This chapter is a general description of virtualization concepts and an explanation about storage virtualization in the IBM Storwize V7000 Unified Disk System.

5.1 User requirements that storage virtualization can meet

In a non-virtualized storage environment, every system is an “island” that must be managed separately. The storage virtualization helps to overcome this obstacle and improves the management process of different storage systems.

You can see the importance of addressing the complexity of managing storage networks by applying the total cost of ownership (TCO) metric to storage networks. Industry analyses show that storage acquisition costs are only about 20% of the TCO. Most of the remaining costs are related to managing the storage system.

When considering virtualization, it is important to recognize that there are many ways to achieve it. Therefore, the best option might vary, depending on requirements. Because of that, the path is usually different for each specific environment.

However, these are several general benefits that can address any concerns:

- Easier administration:
  - Simplified management, because the integrated approach means fewer hardware and software layers to manage separately, so fewer resources are needed.
  - Less management required equates to lower cost.
  - Ideally, a single level of control for use of advanced functions, such as copy services.
  - Decoupling use of advanced functions from, for example, the hardware and technology that is used as the storage back-end.
  - Decoupling use of multiple, virtualized operating system environments from the actual server and hardware that are used.
- Improved flexibility:
  - Shared buffer resources mean that flexible assignment of resources to multiple levels is possible, as required.
  - As needs change over time, there is the potential for automation of this flexible resource assignment.
- Improved resource use:
  - Shared buffers plus their flexible assignment can lead to less resources and fewer resource buffers being required.
  - Delayed acquisitions are an option with this consolidated approach.
  - Cost savings result from the consolidated block and file virtualization approach.

5.2 Storage virtualization terminology

Although *storage virtualization* is a term that is used extensively throughout the storage industry, it can be applied to a wide range of technologies and underlying capabilities. Technically, most storage devices can be claimed to be virtualized in one form or another.

Therefore, we need to define the concept of storage virtualization as used in this book:

- Storage virtualization is a technology that makes one set of resources look and feel like another set of resources, preferably with more desirable characteristics.

- It is a logical representation of resources that is not constrained by physical limitations:
  - It hides part of the complexity.
  - It integrates new functions with existing services.
  - It can be nested or applied to multiple layers of a system.

It is also important to understand that virtualization can be implemented at various layers within the input/output (I/O) stack. We must clearly distinguish between virtualization at the disk layer and virtualization at the file system layer. The focus of this book is virtualization at the disk layer, which is referred to as *block-level virtualization* or *block aggregation layer.*

Virtualization devices can be inside or outside of the data path. When this is the case, they are called in-band and symmetrical or out-of-band and asymmetrical virtualization:

- Symmetrical: In-band appliance

  The device is a storage area network (SAN) appliance n the data path, and all I/O operations flow through the device. This kind of implementation is also referred to as *symmetric virtualization* or *in-band virtualization*.

  The device is both target and initiator. It is the target of I/O requests from the host perspective and the initiator of I/O requests from the storage perspective. The redirection is performed by issuing new I/O requests to the storage. The IBM SAN Volume Controller and IBM Storwize V7000 storage system use symmetrical virtualization.

- Asymmetrical: Out-of-band or controller-based

  The device is usually a storage controller that provides an internal switch for external storage attachment. In this approach, the storage controller intercepts and redirects I/O requests to the external storage as it does for internal storage. The actual I/O requests are redirected. This kind of implementation is also referred to as *asymmetric virtualization* or *out-of-band virtualization*.

Figure 5-1 on page 52 shows variations of these two virtualization approaches.

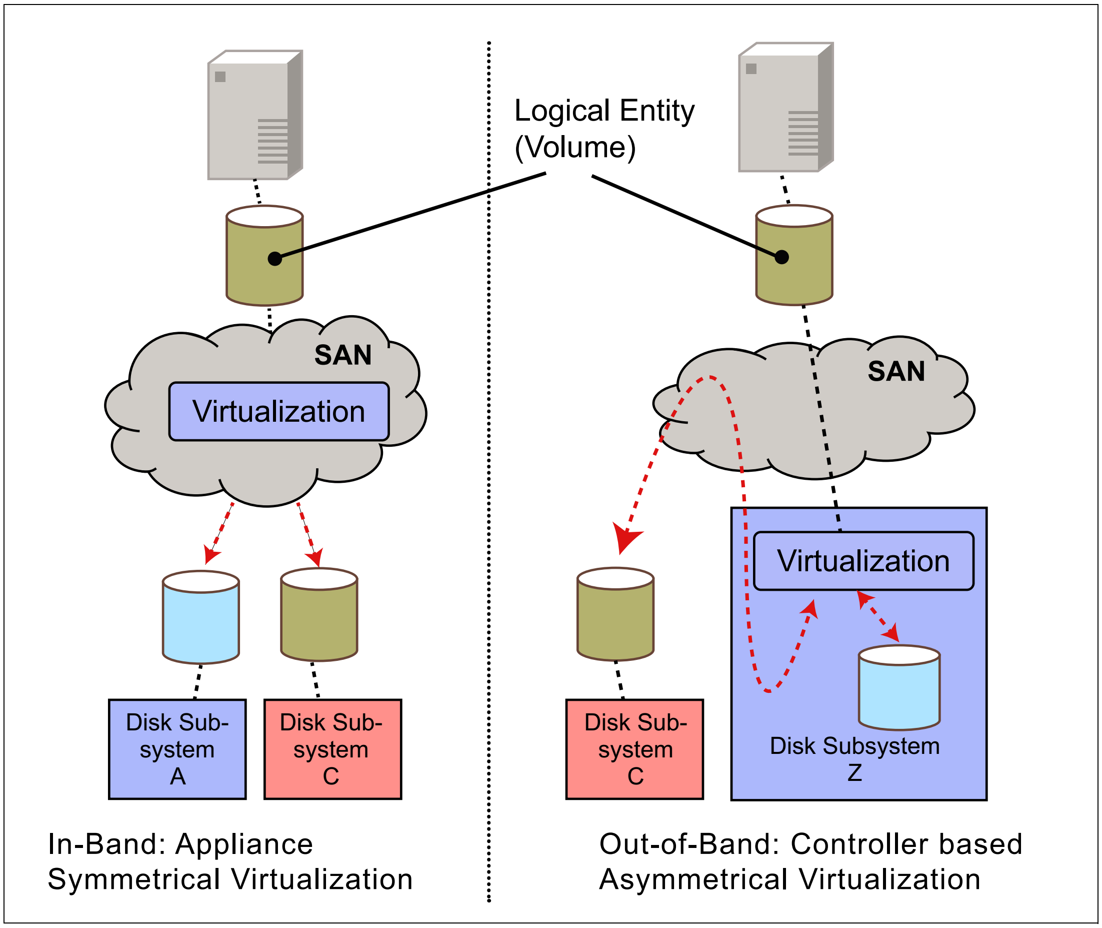


Figure 5-1 Overview of block-level virtualization architectures

In terms of the storage virtualization concept, the focus in this chapter is on block-level storage virtualization in a symmetrical or in-band solution.

Figure 5-2 on page 53 shows in-band storage virtualization on the storage network layer.

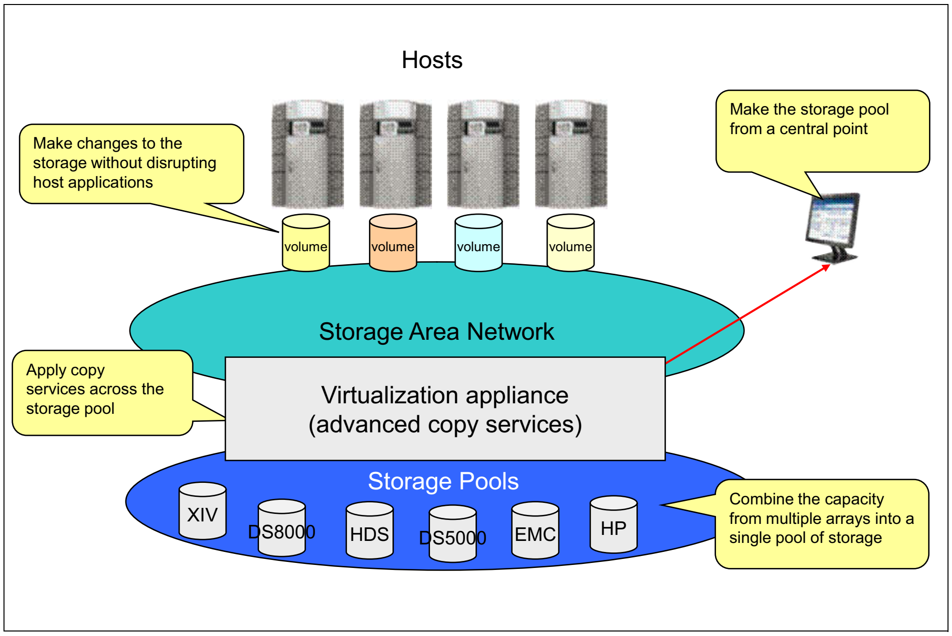

Hosts
Make changes to the storage without disrupting host applications
Make the storage pool from a central point
volume
volume
volume
volume
Storage Area Network
Apply copy services across the storage pool
Virtualization appliance (advanced copy services)
Storage Pools
XIV
DS8000
HDS
DS5000
EMC
HP
Combine the capacity from multiple arrays into a single pool of storage

Figure 5-2 In-band storage virtualization on the storage network layer

The IBM Storwize V7000 Unified Disk System inherits its fixed block-level storage virtualization features entirely from the IBM SAN Volume Controller and Storwize V7000 product family. These features can be used independently from the file system or as enhancements that are built into the Storwize V7000 Unified system, based on the software stack that runs on the two file modules.

5.2.1 Realizing the benefits of Storwize V7000 Unified storage virtualization

The Storwize V7000 Unified system can manage external storage arrays and its own internal storage.

Managing external storage in the V7000 Unified system reduces the number of separate environments that must be managed down to a single environment and provides a single interface for storage management. Moreover, advanced functions, such as mirroring and IBM FlashCopy®, are provided in this system, so there is no need to purchase them for each new disk subsystem.

Migrating data from external storage to the Storwize V7000 Unified system can be done easily because of the virtualization engine that this system offers. This process connects the external storage array to the existing logical unit number (LUN) on the V7000 Unified system and copies the data with the data migration procedure.

In addition, free space does not need to be maintained and managed within each storage subsystem, which further increases capacity use.

5.2.2 Using internal physical disk drives in the Storwize V7000 Unified system

The Storwize V7000 Unified system recognizes internal physical disk drives as *drives* and supports Redundant Array of Independent Disks (RAID) arrays.

The RAID array can consist of different numbers of physical disk drives, depending on the RAID level. However, in general, there can be up to 16 drives in one RAID array. This RAID array with the managed disk (MDisk) is then added to a *pool* layer, which manages performance and capacity. Multiple MDisks can belong to one pool. The logical storage entities, called the *volumes*, are created in these pools. By default, volumes in the pool are striped across all of the MDisks in a pool. The stripe size is known as an *extent*.

These volumes are then mapped to external hosts to provide storage capacity to them. They are seen by the hosts as a Small Computer System Interface (SCSI) disk. This is a volume with underlying special properties and capabilities, such as two independent copies with volume mirroring or, in the case of a thin-provisioned volume, a much smaller real capacity.

These are the logical steps that are required on the V7000 system to set up the volumes and make them accessible to a host:

1. Select the physical drives.
2. Create the RAID array (when that is complete, this array is an MDisk).
3. Add this MDisk to a pool.
4. Create the volumes in this pool.
5. Map the volumes to the external host.

> **Note:** In the Storwize V7000 Unified system, these steps are required for external hosts that are attached through an Internet Small Computer System Interface (iSCSI), Fibre Channel over Ethernet (FCoE), or a Fibre Channel (FC) SAN to the Storwize V7000 component. It works the same for the volumes that are used by the file modules to host the data for file systems and to enable host access through file protocols. Both file modules are directly attached to the Storwize V7000 component as hosts, but they are not displayed as hosts in the standard host panels in the GUI. The volumes that are used for file systems are visible in several places in the GUI windows.

Figure 5-3 shows virtualization layers that use internal disks.

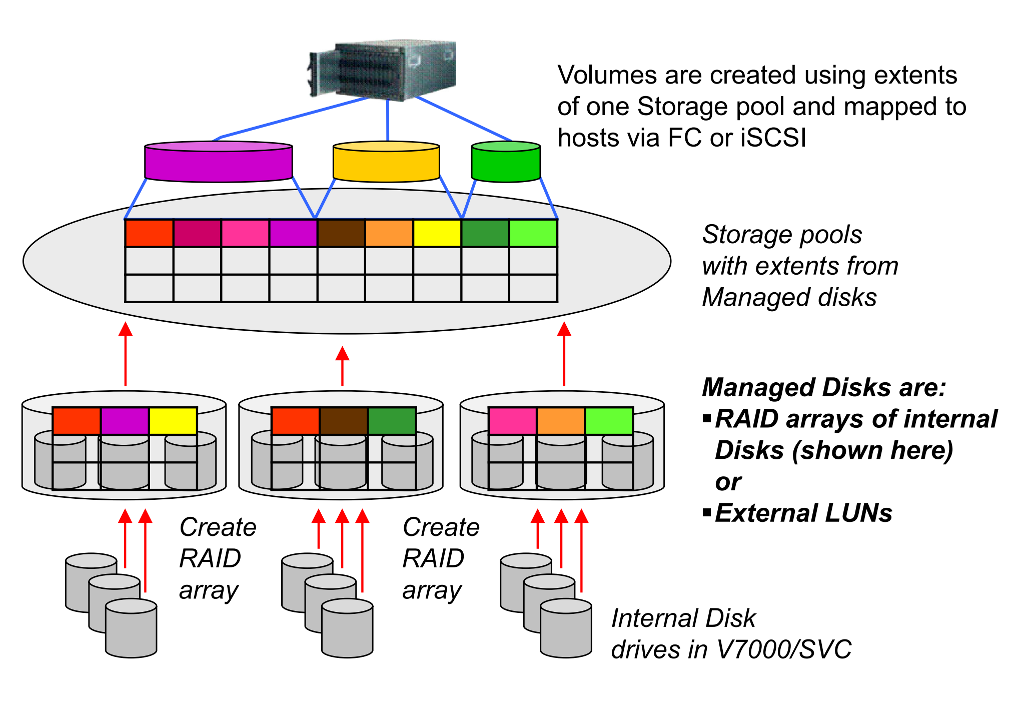

Volumes are created using extents of one Storage pool and mapped to hosts via FC or iSCSI
Storage pools with extents from Managed disks
Managed Disks are:
▪RAID arrays of internal Disks (shown here) or
▪External LUNs
Create RAID array
Create RAID array
Internal Disk drives in V7000/SVC

Figure 5-3 Virtualization layers using internal disks

5.2.3 Using external physical disk drives in the Storwize V7000 Unified system

Important: Do not place file systems that are accessed with the Common Internet File System (CIFS) or Server Message Block (SMB) in external storage systems. CIFS or SMB access to file system volumes is supported only for volumes that are placed in internal storage. External storage can be used to provide NFS, iSCSI, and FC protocols to other hosts after the external storage is under the control of the Storwize V7000 Unified system.

Generally, there is a dependency on the specifics of the external storage subsystem that is being used and the features that are built into it. Therefore, these steps provide an overview of how to use external SAN-attached storage. Our only assumption with these steps is that the storage subsystem provides RAID functionality for data protection against physical disk drive failures.

In the external SAN-attached storage system (storage controller), there are internal logical devices that are created, presented, and mapped to the Storwize V7000 system as logical volumes, logical unit numbers (LUNs), or RAID arrays. This depends on the capabilities of the specific storage system. These logical entities are recognized by the Storwize V7000 system as MDisks, which are then added to storage pools in the same fashion as for MDisks based on Storwize V7000 internal disks. Next, volumes are created in these storage pools and mapped to external hosts. This process is the same as before, as shown in Figure 5-4 on page 56.

The following steps are required on the external storage system and controller to map the external storage to the V7000:

1. Select the physical disks.
2. Create the RAID array.
3. Map the entire array to the Storwize V7000. Or, create logical volumes within the array and map these to the Storwize V7000. They are recognized as MDisks in the Storwize V7000 and treated as such in the next logical configuration steps.
4. To set up the volumes on the Storwize V7000 system and make them accessible to a host, use the CLI to initiate the discovery process for the system to detect the MDisks from the external controller.
5. Add the MDisks that you find to the storage pools.
6. Then create volumes in the pools and map the volumes to the external hosts connected through iSCSI or FC.

See Figure 5-4 on page 56.

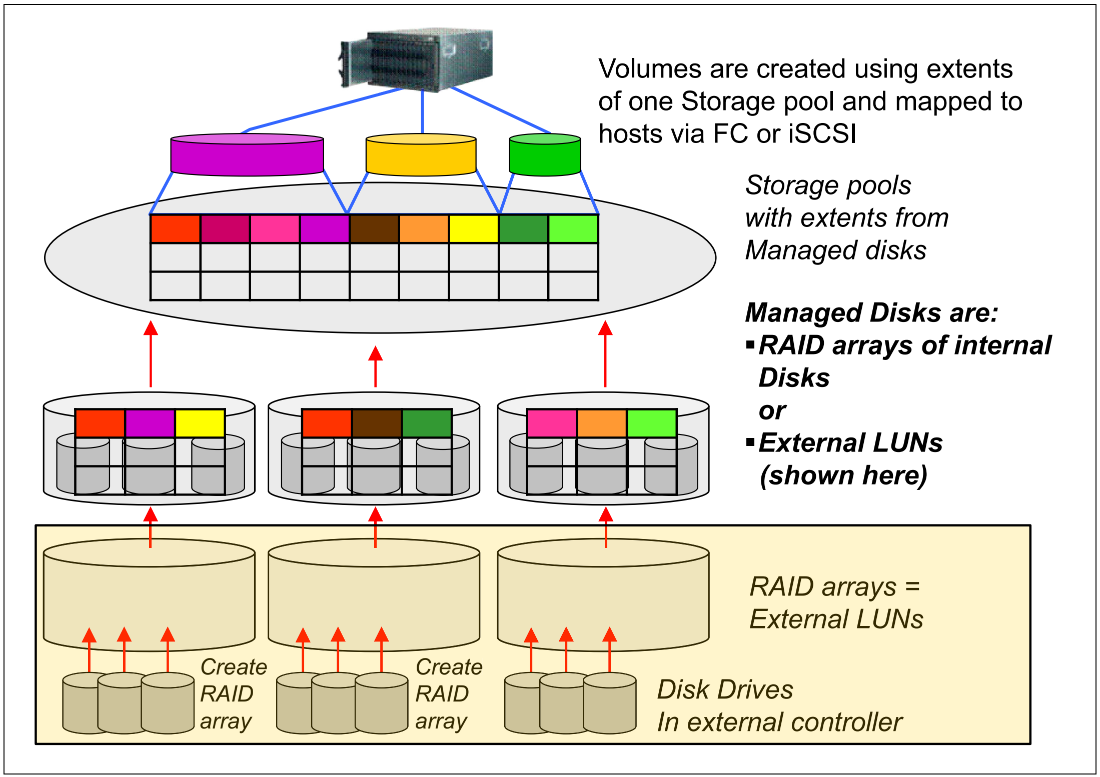

Volumes are created using extents of one Storage pool and mapped to hosts via FC or iSCSI
Storage pools with extents from Managed disks
Managed Disks are:
▪RAID arrays of internal Disks
or
▪External LUNs (shown here)
RAID arrays = External LUNs
Create RAID array
Create RAID array
Disk Drives In external controller

Figure 5-4 Virtualization layers using external disks

In the previous sections, we described how the different layers interact with each other to provide the volume entity that is to be presented to the connected hosts.

In a Storwize V7000 Unified system configuration, the file system storage uses its own set of volumes that are created in Storwize V7000 storage pools. The volumes are created when a file system is created, and that controls the naming, sizing, and number of volumes. Volumes that are intended for other hosts besides the file modules need to be created individually.

5.3 Summary

Storage virtualization is no longer merely a concept or an unproven technology. All major storage vendors offer storage virtualization products. Using storage virtualization as the foundation for flexible and reliable storage helps enterprises to better align business and IT by optimizing the storage infrastructure and storage management to meet business demands.

The IBM System Storage SAN Volume Controller, the IBM Storwize V7000 Disk System, the IBM Storwize V3500, the IBM Storwize V3700, and the IBM Storwize V7000 Unified Disk System are built on a mature, sixth-generation virtualization that uses open standards and is consistent with the Storage Networking Industry Association (SNIA) storage model. The appliance-based in-band block virtualization process (in which intelligence, including advanced storage functions, is migrated from individual storage devices to the storage network) can reduce your total cost of ownership and improve the return on your investment.

It can also improve the use of your storage resources, simplify your storage management, and improve the availability of your applications.

6

NAS use cases and differences: SONAS and Storwize Unified

In this chapter, we build on the network-attached storage (NAS) methods described in Chapter 4, “Access control for file serving clients” on page 37. We describe typical use cases for NAS by using the features and functions that are built into the IBM Storwize V7000 Unified Disk System.

We also list the major differences in the IBM Scale Out Network Attached Storage (SONAS), which is built by using the same software stack that was adopted for the Storwize V7000 Unified system file access component. Therefore, most of the file access methods and functions that are built in are similar. However, because the SONAS hardware is very different, there are also some major differences between the two products. One of the major areas is scalability.

6.1 Use cases for the Storwize V7000 Unified system

The sections that follow provide examples of use cases that benefit from the powerful software features that are built into the Storwize V7000 Unified Disk System. There are many other possibilities, and most of the options can be combined to build a tailored solution that fits your company's needs.

6.1.1 Unified storage with both file and block access

Sometimes, there are requirements for a storage system that can handle providing and blocking file access at the same time. Therefore, an extra benefit is flexibility regarding the storage assignment. There is the flexibility of being able to move storage capacity between these two access methods as required and as needs change over time. With access to the data, simultaneously using both access methods, the methods should not interfere with each other in terms of creating performance dependencies.

The Storwize V7000 Unified system provides the flexibility of storage in both ways in one system. It provides dedicated interfaces for both methods of data access and allows shifting storage capacity between them. It offers the following advantages:

- File access through the IP network, for example, for Server Message Block (SMB), Common Internet File System (CIFS), and Network File System (NFS) exports. File access is handled by the two file modules.
- Block storage access through IP for iSCSI, with block access handled by the Storwize V7000 system.
- Block storage access through the storage area network (SAN) by using Fibre Channel protocol, with block access handled by the Storwize V7000 system.
- Flexibility to use separate or shared storage pools between file access and block access, in both cases internally, using separate volumes for access versus block access.
- Flexible use and moving of storage capacity according to changing needs.

An overview of the different interfaces is shown in Figure 6-1.

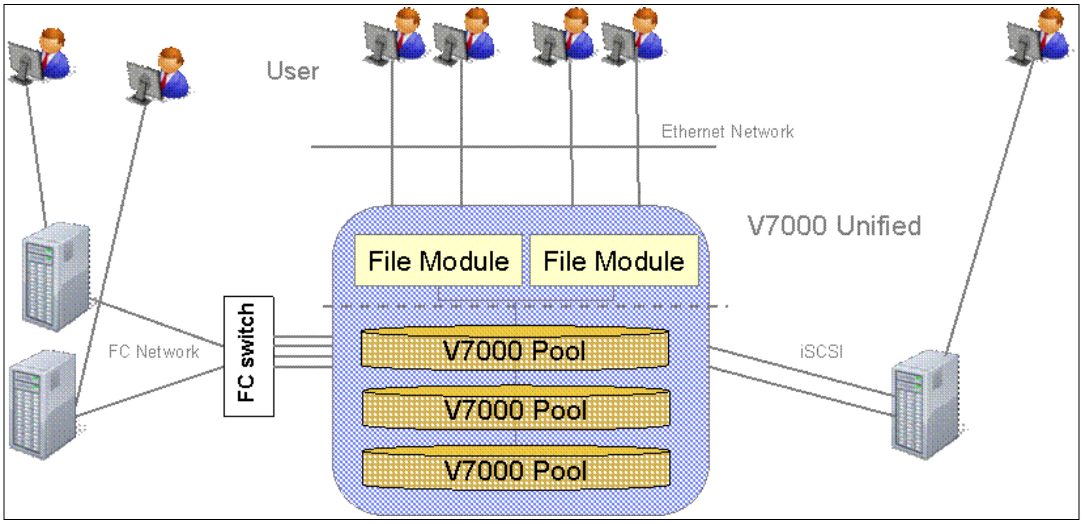


Figure 6-1 Storwize V7000 Unified system: Unified storage for both file access and block access

6.1.2 Multi-user file sharing with centralized snapshots and backup

In many cases, there is a benefit to having centralized storage available to multiple users who are working together, for example in a project workgroup. Although all users have their own home directories and data to work with, there is also a need to share files among the members of the group. This can be set up with the appropriate share and access control structures. In addition, it is more efficient to handle requirements such as data protection and space management at a workgroup level rather than at an individual user level.

The Storwize V7000 Unified system has the appropriate functions to enable centralized management and protection along with individual data access. As shown in Figure 6-2, it also provides enhanced scalability in a single namespace, compared to traditional single NAS file solutions.

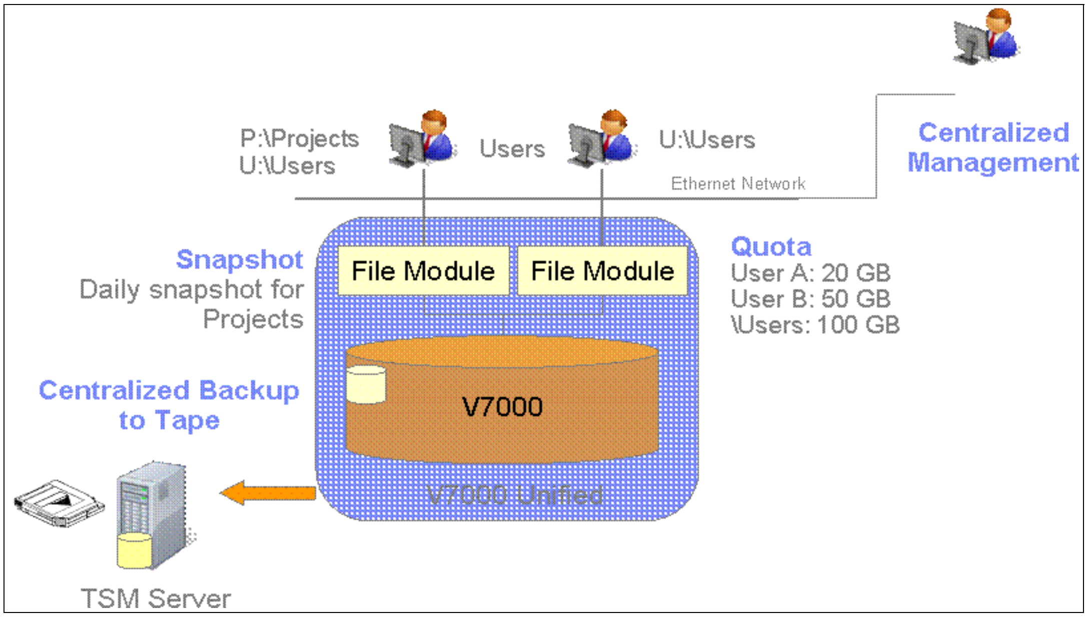


Figure 6-2 Storwize V7000 Unified system centralized management for multiple users sharing files

- Multiple users and groups store and share files on a Storwize V7000 Unified system:
  - Multi-user access to a single file is possible by using IBM General Parallel File System (GPFS) locking mechanisms.
  - File sharing between Microsoft Windows and UNIX client environments is possible.
- The Storwize V7000 Unified system allows you to set quotas for individual users, groups, or at a share level, which offers these advantages:
  - This provides granular space management as needed.
  - It includes warning levels (*soft quota*) and hard limits (*hard quota*).
- The V7000 Unified system allows for centralized snapshots of important data of the entire workgroup.
  - The administrator uses general snapshot rules, such as scope, frequency, and levels of retention, which are tailored to the needs of the workgroup.
  - This removes the need for all users to take care of their own data.
  - It provides easy recovery of multiple versions of files.

- A centralized backup provides data protection for the entire system or at a share level, which offers these benefits:
  - Individual users do not have to define and take care of this themselves.
  - Resource use and scheduling are more efficient.
- Files can be replicated on a second Storwize V7000 Unified system for disaster protection. This works at the file system level and protects the entire file system, including all file shares.

6.1.3 Availability and data protection

The V7000 Unified system has multiple built-in options for availability and data protection of important data:

- Clustered file modules provide high availability and redundancy in case of failures by using a GPFS-based clustered file system to store your data.
- Clustered file modules balance loads for file access from multiple clients simultaneously.
- The system provides a highly available storage back-end.
- Data can and should be protected by backing it up to tape.
- Disaster protection can be augmented by using asynchronous replication of files to a distant site.

Some of these options are illustrated in Figure 6-3.

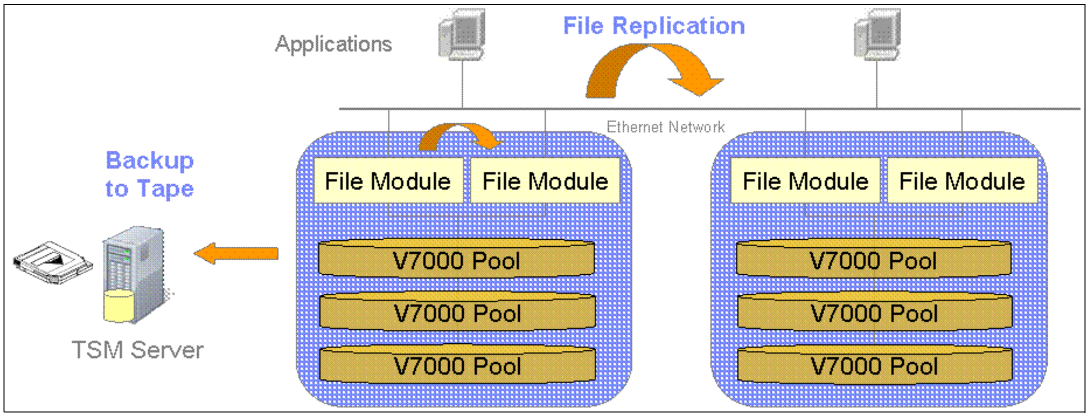


Figure 6-3 Storwize V7000 Unified system availability and data protection

6.1.4 Information lifecycle management (ILM), hierarchical storage management (HSM), and archiving

It is often a requirement to manage the data placement over the lifetime of a file and to place or move it to the appropriate storage tier for a cost-efficient solution. Preferably, this is all done automatically, with minimal administrative tasks and related costs. At the same time, there might be legal requirements to keep certain types of files, for compliance, for an extended period.

The Storwize V7000 Unified system is well-suited to these requirements with its built-in features and through integration with IBM Tivoli Storage Manager for Space Management. Policy-based information lifecycle management (ILM) and hierarchical storage management (HSM) functions meet these requirements, as shown in Figure 6-4 on page 61.

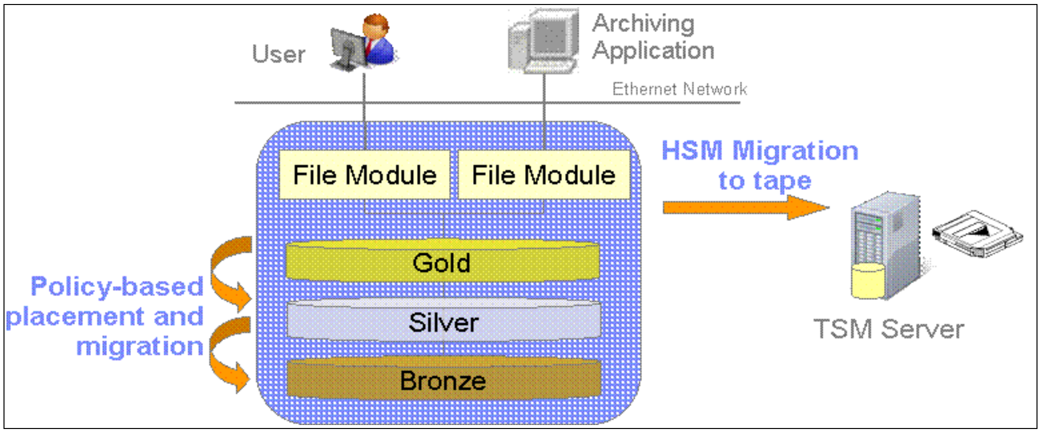


Figure 6-4 Storwize V7000 Unified system as an ILM, HSM, and archiving solution

ILM and HSM provide the following capabilities:

- Archive applications and users can store data in file shares on the Storwize V7000 Unified system.
- You can define policies for automated data placement, migration, and deletion by using these options:
  - A powerful SQL-like policy language that is built in to define individual, tailored policies, as required.
  - Placement policies define the initial placement of a file when it is created.
  - Migration policies move data to its appropriate storage tier over its entire lifetime.
- Static data can be kept for an extended period, as defined by legal requirements, and deleted automatically.
- Tivoli Storage Manager HSM software can automatically migrate rarely used files (according to defined criteria) to tape, with these additional benefits:
  - Both migration to external tape and recall of the data on request are fully visible.
  - A stub file is left in the file system where the original data was located. If there is a request to access that data again, a file recall moves the data back into the file system.

6.2 Storwize V7000 Unified system and SONAS

To describe the differences from the Storwize V7000 Unified system, we first need to introduce and then explain the implementation of the IBM Scale Out Network Attached Storage (SONAS) solution.

6.2.1 SONAS brief overview

SONAS is a scale-out NAS implementation that is built with a focus on scalability. There is significant room for growth until the architectural limitations are hit. However, it still provides a single namespace across the entire configuration. It is a GPFS two-tier implementation (see Chapter 7, “IBM General Parallel File System” on page 65) with two types of nodes:

- The nodes that handle client I/O operations are known as *interface nodes*.
- The nodes that provide the Network Shared Disks (NSDs) and handle the back-end storage tasks are called *storage nodes*.

This process is illustrated in Figure 6-5.

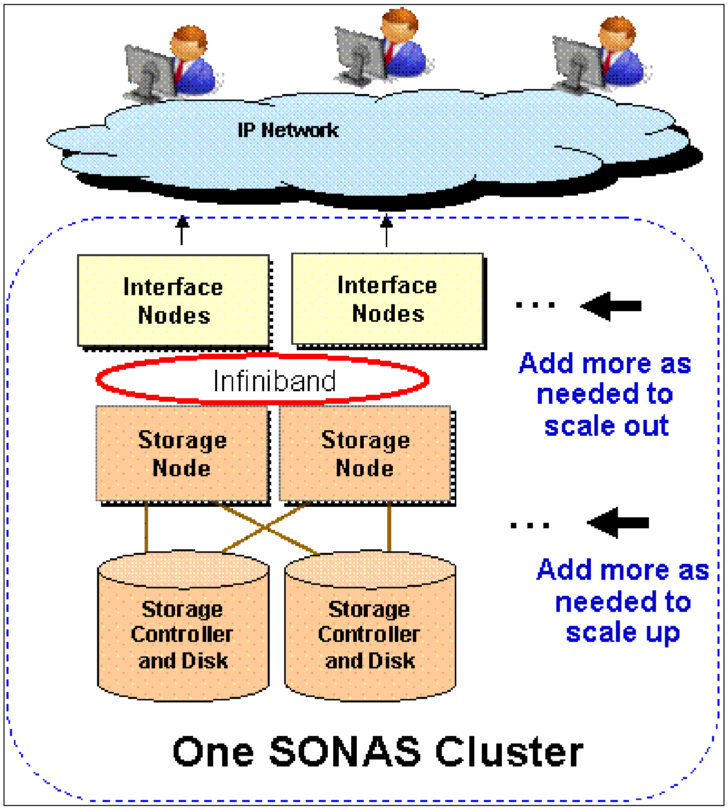

IP Network
Interface Nodes
Interface Nodes
Infiniband
Storage Node
Storage Node
Storage Controller and Disk
Storage Controller and Disk
...
Add more as needed to scale out
...
Add more as needed to scale up
One SONAS Cluster

Figure 6-5 SONAS overview: Two-tier architecture and independent scalability

The scalable, high-speed connectivity between interface nodes and storage nodes is implemented through an internal InfiniBand network.

At the time of writing, the latest SONAS version is 1.4.2. For detailed information on the latest version, see the SONAS Information Center:

http://pic.dhe.ibm.com/infocenter/sonasic/sonas1ic/index.jsp

SONAS 1.4.2 supports up to the following capacities:

- **30 interface nodes.** Each interface node can provide from two to eight Ethernet connections.
- **30 storage pods.** These pods provide up to 7200 disk drives in the storage back-end. With 3 TB hard disk drives (HDDs), this results in a 21.6 PB storage capacity.

SONAS uses a centralized management concept with multiple distributed management node roles. This configuration provides single access by using either a graphical user interface (GUI) or command-line interface (CLI) with the SONAS cluster.

In general, SONAS and the Storwize V7000 Unified system support the same software features. However, there are differences, such as the ones listed in 6.2.2, “Implementation differences between Storwize V7000 Unified and SONAS” on page 63. Therefore, be sure to check both products’ support pages as the official references:

- Support portal for SONAS:

  http://ibm.co/1olkvk0

- Support portal for the Storwize V7000 Unified system:

  http://www.ibm.com/storage/support/storwize/v7000/unified

6.2.2 Implementation differences between Storwize V7000 Unified and SONAS

Although both products use the same NAS software stack, there are differences because of the different hardware implementations, scalability, and supported software features. The following list might be subject to change in future releases, but it reflects differences at current software releases for both products (SONAS Version 1.4.2 and Storwize V7000 Unified Disk System Version 1.4):

- GPFS one-tier architecture in the Storwize V7000 Unified system and two-tier architecture in SONAS.

  SONAS uses dedicated, different servers as interface nodes and storage nodes. The Storwize V7000 Unified system uses a pair of file modules,

- Hardware scalability is limited in the Storwize V7000 Unified system. For example, there is no independent scalability of resources for interface nodes and storage nodes in the Storwize V7000 Unified system. This requires planning for these constraints:

  - There is a fixed number of two file modules in the Storwize V7000 Unified system.
  - Storage capacity limits in the Storwize V7000 Unified system has the following storage capacity limits:

    - Internally, one V7000 control enclosure
    - Up to nine V7000 expansion enclosures (up to 240 disk drives)
    - Support for external virtualized SAN-attached storage

- Local authentication is available only on the Storwize V7000 Unified system. For more information, see Chapter 11, “Implementation” on page 141.

- IBM Real-time Compression is available only on the Storwize V7000 Unified system. For more information, see Chapter 16, “IBM Real-time Compression in the Storwize V7000 Unified system” on page 289.

7

IBM General Parallel File System

In this chapter, we describe the clustered file system that is built into the IBM Storwize V7000 Unified Disk System as part of its foundation: the IBM General Parallel File System (GPFS). In terms of the GPFS implementation within the V7000 Unified system, we advise you to check *V1.4 Configuration Limits and Restrictions for IBM Storwize V7000 Unified*:

`http://www.ibm.com/support/docview.wss?uid=ssg1S1004227`

Asynchronous replication for files, as implemented in the Storwize V7000 Unified system, is not a generic GPFS function. Therefore, we do not describe that feature. See Chapter 8, “Copy services overview” on page 75.

7.1 Overview

GPFS has been available from IBM since the mid-1990s. Its roots are in parallel computing requirements, where a scalable, highly available file system was required. GPFS includes the parallelism of serving data to (many) clients, as well as availability and scalability. Therefore, it has been used for many years in parallel computing, high-performance computing (HPC), digital media solutions and smart analytics, among other uses.

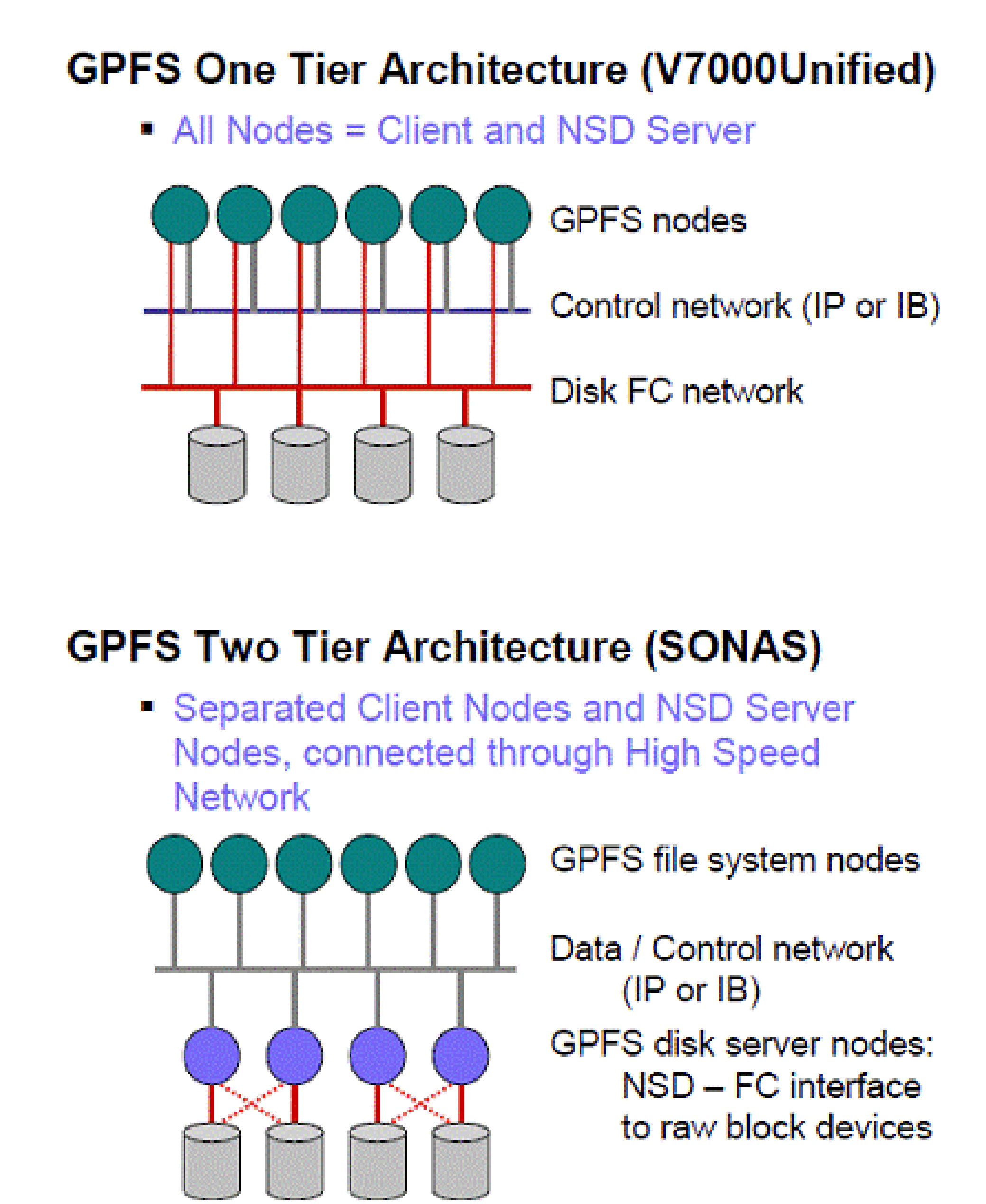
It also inherited functions from other projects over time, such as the IBM SAN file system. GPFS is part of many other IBM offerings, such as IBM Information Archive and Smart Analytics solutions.

7.2 GPFS technical concepts and architecture

The concept of GPFS is a clustered file system that is built on a grid parallel architecture. Parallelism for both host access and data transfers to storage enables its scalability and performance.

The storage entities that GPFS recognizes are called *Network Shared Disks* (NSDs), as shown in Figure 7-1 on page 67. GPFS works with the concept of separate NSD servers and NSD clients in a two-tier architecture. However, if only one tier is used, both NSD server and client roles are in the same machine.

The Storwize V7000 Unified system implementation uses one-tier architecture. In contrast, an IBM Scale Out Network Attached Storage (SONAS) implementation uses two-tier architecture (also shown in Figure 7-1 on page 67).



GPFS One Tier Architecture (V7000Unified)

- All Nodes = Client and NSD Server

GPFS nodes
Control network (IP or IB)
Disk FC network

GPFS Two Tier Architecture (SONAS)

- Separated Client Nodes and NSD Server Nodes, connected through High Speed Network

GPFS file system nodes
Data / Control network (IP or IB)
GPFS disk server nodes: NSD – FC interface to raw block devices

Figure 7-1 GPFS: Examples for one-tier (Storwize V7000 Unified) and two-tier architecture (SONAS)

GPFS stripes all of the data that is written across all available NSDs by using the defined file system block size as the stripe size. This way, GPFS ensures that the maximum number of NSDs is contributing to a particular I/O operation and avoids performance hot spots:

- Supported file system block sizes in the Storwize V7000 Unified system are 256 KB, 1 MB, and 4 MB.
- The minimum I/O size that GPFS is using is called a *sub-block* or *fragment* (which is 1/32 of the file system block size, because GPFS is working with 32 sub-blocks for each file system block, internally).
- Sub-blocks are introduced to combine small files or parts of files into a single block to avoid wasted capacity.

Therefore, for virtualized NSDs presented to the GPFS layer (such as NSDs provided by the Storwize V7000 Unified system), it is beneficial to optimize the NSD I/O characteristics according to the GPFS I/O pattern.

From a storage subsystem perspective, the goal is to get *full stride writes* to its underlying RAID arrays. This means that parity (for RAID 5 and RAID 6) can be calculated immediately without the need to read from disks first, which avoids the RAID penalty. In conjunction with GPFS, this leads to a change of the RAID presets that are built into the Storwize V7000 Unified system, compared to the V7000 stand-alone system. The presets for the RAID arrays in the Storwize V7000 Unified system aim to configure eight data disks plus the required parity disks into one RAID array:

- For RAID 5, it is an 8+P array
- For RAID 6, this is an 8+P+Q array

These new presets are also reflected in the sizing tools, such as IBM Capacity Magic and IBM Disk Magic.

7.2.1 Split brain situations and GPFS

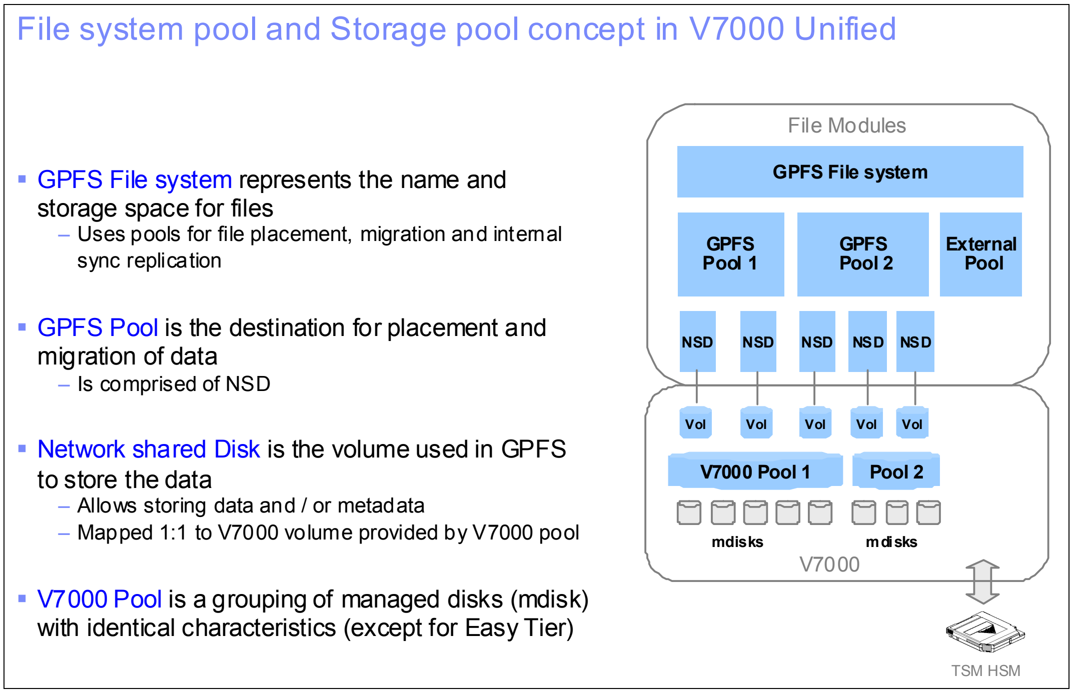
A GPFS cluster normally requires at least three cluster nodes. An uneven number is selected on purpose by most clustering solutions so that there is still a quorum of cluster nodes (or other voting members) available if one of them fails to avoid a *split brain* situation in the cluster. A typical split brain scenario means that there are two parts of the cluster still active, each one with half of the remaining cluster nodes (so that none of the two parts has a quorum), but they cannot communicate any longer. Because of the loss of communication between the two parts, the parts cannot distinguish which one has the most current information and should continue to operate. One solution for this is to have a tiebreaker in the configuration. The IBM SAN Volume Controller and Storwize V7000 clustering implementation uses a quorum disk for that purpose.

In the GPFS cluster implementation in the Storwize V7000 Unified system, it is not possible to have three cluster nodes because there are only two file modules in the configuration. To help with a split brain situation when the file modules lose communication, the Storwize V7000 storage system in the back-end acts as the tiebreaker. If they lose communication, both file modules, as cluster members, communicate with the Storwize V7000 and provide their status.

The Storwize V7000 then determines which file module should continue to operate (*survive*) as the GPFS cluster and sends an **expelmember** command to the other file module, which then must leave the cluster. In addition, the V7000 removes the volume mappings of the expelled file module to guarantee data integrity for the remaining GPFS cluster.

7.2.2 GPFS file system pools and the Storwize V7000 storage pools

GPFS has the internal concept of *pools* (in GPFS internal terminology, they are called *storage pools*, too). To distinguish them from the pools that are used in the Storwize V7000 storage layer, we refer to them as *file system pools*. As described in Chapter 3, “Architecture and functions” on page 19, these GPFS file system pools are mapped to V7000 storage pools within a Storwize V7000 Unified system. Figure 7-2 on page 69 shows n overview of the different internal structures that are involved.



File system pool and Storage pool concept in V7000 Unified

- GPFS File system represents the name and storage space for files
  - Uses pools for file placement, migration and internal sync replication
- GPFS Pool is the destination for placement and migration of data
  - Is comprised of NSD
- Network shared Disk is the volume used in GPFS to store the data
  - Allows storing data and / or metadata
  - Mapped 1:1 to V7000 volume provided by V7000 pool
- V7000 Pool is a grouping of managed disks (mdisk) with identical characteristics (except for Easy Tier)

Figure 7-2 File system pool and storage pool concept in a Storwize V7000 Unified system

Regarding the mapping between file system pools and storage pools, there is one exception: GPFS synchronous internal replication uses a *one to two* mapping (one file system pool to two storage pools), as described in 7.2.6, “GPFS synchronous internal replication” on page 72. Normally, there is a *one to one* mapping between file system pools and storage pools. In either case, this mapping is typically established when the file system is created.

For a standard file system that is not providing information lifecycle management (ILM), there is one file system pool, which is mapped to one storage pool.

For a file system that is providing ILM functions, there are multiple file system pools, for example, *system gold*, *silver*, and *bronze*, where each one is mapped to one storage pool, and where the storage pools have descending tiers. That is, storage classes descend from fast and expensive to slower and more affordable storage, from the tier that is mapped to the system gold file system pool to the one mapped to the bronze file system pool.

> **Note:** To ensure that there is no confusion between these two logical entities within the Storwize V7000 Unified system, we refer to the GPFS pool entity as the *file system pool* and the Storwize V7000 pool as the *storage pool*.

7.2.3 File system pools in GPFS

The following standards exist in file system pools in GPFS:

- There are a maximum of eight internal pools per file system.
- One pool (default) is always required as the system pool.
- Seven optional pools are available as user pools.

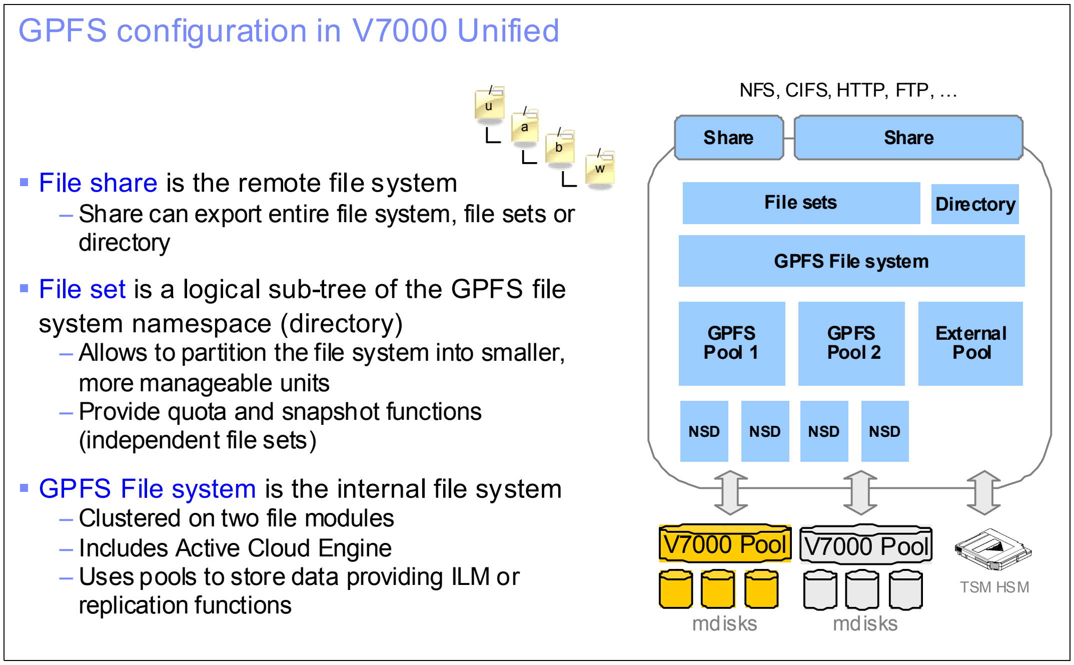
- For configuring ILM, the file system is created with multiple file system pools. Therefore, in addition to the one default file system pool called *system* (which exists for every GPFS file system), there are as many more *file system pools* of descending tiers that are defined (and mapped to storage pools with corresponding descending tiers) as extra storage tiers.
- An example of a typical hierarchy of pools of descending storage tiers and classes is gold, silver, bronze.
- An external file system pool is also possible. This pool is used for offloading data as part of a hierarchical storage management (HSM) solution that is managed by IBM Tivoli Storage Manager for Space Management as the storage manager application for HSM.

GPFS provides a clustered file system layer that can support up to 256 file systems in a native GPFS or SONAS implementation. The supported limit in a Storwize V7000 Unified environment is 64 file systems, currently.

> **Note:** For the most current information for not only the file system limit, but for all specifications, see the *V1.4 Configuration Limits and Restrictions for IBM Storwize V7000 Unified* web page:
>
> `http://www.ibm.com/support/docview.wss?uid=ssg1S1004227`

In addition to the *file system pools* (with NSDs) and *storage pool* (with volumes) concepts that are shown in Figure 7-2 on page 69, there are more configuration layers that are involved to establish and manage access to the data stored in the GPFS file system.

The essential step is the definition of shares (exports) to be able to access the data from file clients. *File sets* and directories provide more granularity for the management of that data access, as shown in Figure 7-3 (file sets also enable quota management, and independent file sets enable snapshots in addition).



GPFS configuration in V7000 Unified

- File share is the remote file system
  - Share can export entire file system, file sets or directory
- File set is a logical sub-tree of the GPFS file system namespace (directory)
  - Allows to partition the file system into smaller, more manageable units
  - Provide quota and snapshot functions (independent file sets)
- GPFS File system is the internal file system
  - Clustered on two file modules
  - Includes Active Cloud Engine
  - Uses pools to store data providing ILM or replication functions

Figure 7-3 Layers that are involved in managing data access in the Storwize V7000 Unified system

7.2.4 GPFS file sets

A *file set* is a subtree of a file system namespace that behaves like a separate file system in many ways. You can use file sets to partition a file system to allow administrative operations at a finer granularity than for the entire file system.

File sets are available in two types: *dependent file sets* and i*ndependent file sets*:

- An independent file set has a separate inode space but shares physical storage with the remainder of the file system.
- A dependent file set shares the inode space and snapshot capability of the independent file set where it is contained.

When the file system is created, only one file set exists, which is called the *root file set*. The root file set contains the root directory and system files, such as quota files.

Knowing file set details and differences helps:

- The default is one root file set per file system.
- Quotas and policies are supported on both *dependent file sets* and *independent file sets*.
- Snapshots are only supported for independent file sets (and at the level of the entire file system) for the following reasons:
  - Only independent file sets provide their own inode space. Dependent file sets use the inodes of the file system.
  - The GPFS limit for snapshots is 256 per file system. However, 32 are reserved for internal use, such as for backup and asynchronous replication. Therefore, a maximum of 224 snapshots per file system are available to the user.
  - A maximum of 256 snapshots are available for each independent file set.
- A maximum of 1000 independent file sets and 3000 dependent file sets is supported for each file system.

For more information about the differences between dependent and independent file sets, see the IBM Storwize V7000 Unified Information Center:

http://ibm.co/1mQZEtt

7.2.5 GPFS parallel access and byte-range locking

GPFS uses a distributed cluster manager and various roles that are distributed across the cluster nodes, both highly scalable and highly available, with adaptive and self-healing capabilities. For that purpose, all nodes or a redundant subset of nodes have equal roles and run the same daemons. If one node fails, another available node can take over its role.

GPFS allows parallel access from different client systems to a single file that is managed by sophisticated locking mechanisms. Those mechanisms operate GPFS cluster-wide on the level of byte range within a file. There is also support for *opportunistic locking* (*oplocks*), which allows client-side caching of data. This caching can provide a performance benefit.

Important: For applications with critical data, disable all of the following non-mirrored caching options in GPFS:

- Caching on the client side: Controlled through the opportunistic locking (oplocks) option
- Caching in file module cache or local interface node cache: Controlled through the *syncio* option

This can be done by using the CLI **`chexport`** command for the relevant shares with these parameters:

- **`oplocks=no`**
- **`syncio=yes`**

7.2.6 GPFS synchronous internal replication

GPFS provides optional redundancy by using a synchronous, file system internal replication, which is based on grouping NSDs into independent *failure groups*:

- A *failure group* is a logical group of NSDs with the same dependencies, for example, failure boundaries:
  - This is typically two independent storage systems that provide NSDs to protect the GPFS data against the failure of an entire storage system.
  - Within the implementation in the Storwize V7000 Unified system, there is one Storwize V7000 in the back-end, which provides redundancy by design and protects against any single failure. Also, virtualized external SAN-attached storage systems are managed and presented by the Storwize V7000 and can also rely on the failure boundary of an entire Storwize V7000 storage system.
- Setting up the GPFS synchronous internal replication requires the definition of two independent storage pools (as failure groups) for each file system pool:
  - This results in a two-to-one mapping between two storage pools and the file system pool.
  - Metadata is always stored in the file system pool.
- The replication process is fully synchronous mirroring over two sets of NSDs from the two independent storage pools.
- The configurable options are to replicate *metadata*, *data*, or *both* between two independent storage pools (failure groups).

Note: GPFS synchronous internal replication of metadata, data, or both is fully supported and configurable on the Storwize V7000 Unified system. The goal is to perform synchronous mirroring between independent failure groups on two separate storage systems, which do not have to provide redundancy. In the Storwize V7000 Unified implementation, there is one highly available main storage subsystem with redundancy by design. This system is the Storwize V7000 itself, managing all of the disk storage in the back-end (Storwize V7000 internal disks and external SAN-attached storage systems that are virtualized by the Storwize V7000 system).

7.2.7 IBM Active Cloud Engine

GPFS provides a fast and scalable scan engine that can scan through all files or subdirectories quickly and that is used for multiple purposes. For example, purposes include identifying files for antivirus scanning, for ILM, and for changed files for incremental backups. Because GPFS is designed for scalability and can grow to a large file system with a single namespace, its scan engine is also scalable. This is a significant competitive advantage of GPFS compared to other large clustered file systems.

In the GPFS implementations in both SONAS and Storwize V7000 Unified systems, this engine is called the IBM Active Cloud Engine®.

The GPFS scan engine is also used to apply user-defined policies to the files that are stored in the GPFS, building the foundation for the ILM of files within GPFS:

- Enables ILM, automated migrations of files that are based on user-defined criteria
- Uses a subset of Structured Query Language (SQL) as policy language
  - Rules that are grouped in different policies can be specified by using this policy language.
  - Can be scripted.
- Policies and rules for file placement (creation), migration, and deletion
- Rules within the policy are evaluated first to last. The first one to match is applied and determines the handling of the relevant files:
  - The recommendation is to add a default rule in every case that gets applied when no other rule matches the criteria.
  - If no default rule exists and no other rule matches the criteria that are defined, no action is taken unless defined otherwise.

For more information, see the *GPFS 3.5: Advanced Administration Guide*, SC23-5182-05:

http://www.ibm.com/support/docview.wss?uid=pub1sc23518205

7.2.8 GPFS and hierarchical storage management (HSM)

GPFS has support for Tivoli Storage Manager HSM built-in, which provides a way to offload data from the file system to external storage with Tivoli Storage Manager for Space Management. Internally, this is handled by a special file system pool that is defined, called the *external pool*. Based on the defined criteria, the GPFS policy engine identifies the files to be moved. GPFS then moves these files into this external storage pool from which the Tivoli Storage Manager HSM server fetches the data and stores it on a Tivoli Storage Manager HSM-supported storage device, usually a tape device.

While the data itself is being offloaded and is saving space within the file system, for every file, a *stub file* is left inside of the file system. This file contains all of the metadata that belongs to that file and that the scan and policy engine need to read. Therefore, for policy scans, the data can remain in the external storage, outside of the GPFS file system, because all metadata information for this file is still available through the stub file. But if a user wants to access the file, or Tivoli Storage Manager wants to back up the file, or the antivirus scanner wants to scan the file, it is recalled and loaded back into the file system by HSM. This process is all apparent to the user.

7.2.9 GPFS snapshots

GPFS also offers a space-efficient snapshot technology, which is described in Chapter 8, “Copy services overview” on page 75. The following list is a summary of the main features of GPFS snapshots as implemented in the Storwize V7000 Unified system:

- Snapshots use pointers to data blocks based on *redirect-on-write*. This means that a new write coming in, and even an update to an existing data block, is written to a new data block because the old data block is still contained in the snapshot. This method provides efficient use of space because it does not use space or capacity when started. Space is only used when data changes and new blocks are written based on redirect-on-write.
- Snapshots are available for file systems and for independent file sets.
- A maximum of 256 snapshots of the entire file system is allowed, plus 256 for independent file sets underneath.
- Of these, 32 of these snapshots are reserved for internal use, so 224 are available for client use.
- Snapshot rules allow scheduling snapshots if required, and retention rules for snapshots can be defined.
- The snapshot manager routine runs once per minute, executing snapshot rules in sequential order of definition.

7.2.10 GPFS quota management

You can specify quotas for file system space management:

- Quotas can be set at a file set, user, or group level.
- Soft quotas, a grace period, and hard quotas are supported:
  - When the soft quota limit is reached, a warning is sent but write access is still possible until either the grace period expires or the hard quota limit is reached, whichever comes first. Writing of data is then inhibited until space is freed up, for example by deleting files.
  - The default grace period is seven days.
  - When the hard quota limit is reached while updating a file, writing and closing the currently open file is still possible to protect the data.

You can find more detailed information about IBM GPFS, including different purpose-built configurations, in these IBM Redbooks publications:

- *GPFS: A Parallel File System*, SG24-5165
- *Implementing the IBM General Parallel File System (GPFS) in a Cross Platform Environment*, SG24-7844

8

Copy services overview

This chapter gives you an overview of the IBM Storwize V7000 Unified Disk System storage copy functions that are provided by the Storwize V7000 storage subsystem. It also describes the file-level copy functions that are provided by the file modules.

For details about storage copy functions, see the IBM Redbooks publication titled *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

8.1 Storage copy services of the Storwize V7000 Unified

The Storwize V7000 Unified system provides storage in the form of logical volumes to the internal file modules and to external storage clients. In addition, it provides the same logical volume-based copy services that the stand-alone Storwize V7000 provides. These services are IBM FlashCopy, IBM Metro Mirror, IBM Global Mirror, and Global Mirror with Change Volumes.

8.1.1 FlashCopy for creating point-in-time copies of volumes

FlashCopy is the point-in-time copy capability of the Storwize V7000. It is used to an create instant, complete, and consistent copy from a source volume to a target volume. Often, this functionality is called *time-zero copy*, *point-in-time copy*, or *snapshot copy*.

Creating a copy without snapshot functionality

Without a function such as FlashCopy, to achieve a consistent copy of data for a specific point in time, the I/O of the application that manipulates the data must be quiescent for the entire time that the physical copy process takes place. The time that the copy process requires is defined by the amount of data to be copied and the capabilities of the infrastructure to copy the data. Only after the copy process is finished can the application that manipulates the data start to access the volume that was involved. Only then is it ensured that the data on the copy target is identical to the data on the source for a specified point in time.

Creating copies with FlashCopy

With FlashCopy, this process is different. FlashCopy enables the creation of a copy of a source volume to a target volume in a very short time. Therefore, the application has to be prevented from changing the data only for a short period. For the FlashCopy function to be executed, a FlashCopy mapping must be created; two ordinary volumes get mapped together for the creation of a point-in-time copy.

After the FlashCopy process is started on the mapping, the target volume represents the contents of the source volume for the point in time when the FlashCopy was started. The target volume does not yet contain all the data of the source volume *physically*. It can be seen as a virtual copy, created by using bitmaps.

After FlashCopy starts but before it finishes physically copying the data to the target, the copy can be accessed in read/write mode. From that point on, data that has to be changed on the source volume (by the applications that manipulate the source volume) is written to the target volume beforehand. This ensures that the representation of the data on the target volume for that point in time is valid.

It is also possible to copy all of the data of the source volume to the target volume through a background copy process. The target volume, although not fully copied yet, represents a clone of the source volume when the relationship between the source and the target exists. When all data has been copied to the target volume, the relationship between the volumes can be removed, and both volumes become normal. The former target volume is now a physical clone of the source volume for the point in time that the FlashCopy was started.

To create consistent copies of data that spans multiple volumes, *consistency groups* can be used. Consistency groups are sets of FlashCopy mappings, that get copied at the same point in time, so they create a consistent snapshot of the data across all volumes.

There are many variations in ways to use FlashCopy:

- **Cascaded FlashCopy.** A volume can be a source volume in one FlashCopy mapping and for that volume to be the target volume in another FlashCopy mapping.
- **Multiple Target FlashCopy.** One volume can be the source volume in multiple FlashCopy mappings with *different* target volumes.
- **Incremental FlashCopy.** You can incrementally update a fully copied target volume with only the changes made to the source volume of the same mapping.
- **Reverse FlashCopy.** You can reverse the direction of the mapping, which makes it possible to restore a source volume from a target volume while retaining the original target volume.
- **Space-Efficient FlashCopy.** This refers to enhancing the flexibility of FlashCopy with the use of thin provisioned volumes.

FlashCopy use cases

FlashCopy has many uses. One obvious use is for backing up a consistent set of data without requiring a long backup time frame. The application that is manipulating the data must ensure that the data is consistent, and the application must be suspended for a short period. When the copy is started, the backup application can access the target, and the applications can resume manipulating the live data. No full volume copy is needed.

One especially useful case for FlashCopy is to create a full consistent copy of production data for a specific time at a remote location. In this case, we combine Metro Mirror, Global Mirror, and FlashCopy, and we take a FlashCopy from the Metro Mirror and Global Mirror secondary volumes. We can make a consistent backup of production data on the second location or create a clone of the data so that it is available if anything happens to the production data.

FlashCopy can also be used as a safety net for operations that make copies of data inconsistent for longer-than-normal periods of time. For example, if Global Mirror gets out of sync, the auxiliary volume is still consistent in itself, but the process of resynchronization renders the auxiliary volume inconsistent if it is not finished. To obtain a consistent copy of the data of the auxiliary volume while it is being synchronized, a FlashCopy of this volume can be created.

Another use for FlashCopy is to create clones of data for application development testing or for application integration testing. FlashCopy is also useful when a set of data needs to be used for different purposes. For example, a FlashCopy database can be used for data mining.

FlashCopy presets

The IBM Storwize V7000 storage subsystem provides three FlashCopy presets, named *Snapshot*, *Clone*, and *Backup*, to simplify the more common FlashCopy operations, as shown in Table 8-1 on page 78.

Table 8-1 FlashCopy presets

Preset	Purpose
Snapshot	Creates a point-in-time view of the production data. The snapshot is not intended to be an independent copy. It is used to maintain a view of the production data at the time the snapshot is created. This preset automatically creates a thin-provisioned target volume with 0% of the capacity that is allocated at the time of creation. The preset uses a FlashCopy mapping with 0% background copy so that only data written to the source or target is copied to the target volume.
Clone	Creates an exact replica of the volume, which can be changed without affecting the original volume. After the copy operation completes, the mapping that was created by the preset is automatically deleted. This preset automatically creates a volume with the same properties as the source volume and creates a FlashCopy mapping with a background copy rate of 50. The FlashCopy mapping is configured to automatically delete itself when the FlashCopy mapping reaches 100% completion.
Backup	Creates a point-in-time replica of the production data. After the copy completes, the backup view can be refreshed from the production data, with minimal copying of data from the production volume to the backup volume. This preset automatically creates a volume with the same properties as the source volume. The preset creates an incremental FlashCopy mapping with a background copy rate of 50.

8.1.2 Metro Mirror and Global Mirror for remote copy of volumes

IBM Metro Mirror and IBM Global Mirror are names for the synchronous remote copy (Metro Mirror) and asynchronous remote copy (Global Mirror) functions. We use the term *Remote Copy* to refer to both functions where the text applies to each equally. These functions are used to maintain a copy of logical volumes that are held by one Storwize V7000 subsystem, Storwize V7000 Unified system, or IBM SAN Volume Controller in another Storwize V7000 subsystem, Storwize V7000 Unified system, or SAN Volume Controller at a remote location. This copy can be either synchronous or asynchronous. You can use Global Mirror with the Change Volumes function for low-bandwidth links (this function was introduced in SAN Volume Controller Version 6.3.0).

IBM Metro Mirror

Metro Mirror works by establishing a *synchronous* copy relationship between two volumes of equal size. This relationship can be an intracluster relationship that is established between two nodes within the same I/O group of one cluster. This is an intercluster relationship, which means a relationship between two clusters that are separated by distance. Those relationships can be stand-alone or in a consistency group.

Metro Mirror ensures that updates are committed to both the primary and secondary volumes before sending confirmation of completion to the server. This ensures that the secondary volume is synchronized with the primary volume. The secondary volume is in a read-only state, and manual intervention is required to change that access to a read/write state. The server administrator must also mount the secondary disk so that the application can start to use that volume.

IBM Global Mirror

Global Mirror copy relationships work similarly to the way that Metro Mirror does but by establishing an *asynchronous* copy relationship between two volumes of equal size. This relationship is intended mostly for intercluster relationships over long distances.

With Global Mirror, a confirmation is sent to the server before copying is complete at the secondary volume. When a write is sent to a primary volume, it is assigned a sequence number. Mirror writes sent to the secondary volume are committed in sequential number order. If a write is issued while another write is outstanding, it might be given the same sequence number.

This function operates to always maintain a consistent image at the secondary volume. It identifies sets of I/O operations that are active concurrently at the primary volume, assigns an order to those sets, and applies the sets of I/Os in the assigned order at the secondary volume. If another write is received from a host while the secondary write is still active for the same block, even though the primary write might be complete, the new host write on the secondary is delayed until the previous write is completed.

Global Mirror with Change Volumes

Global Mirror with Change Volumes is a function added to Global Mirror to help attain consistency on lower-quality network links.

Change Volumes use FlashCopy but cannot be manipulated as FlashCopy volumes, because they are special-purpose only. Change Volumes provide the ability to replicate point-in-time images on a cycling period (the default is 300 seconds.) This means that the change rate needs to include only the condition of the data at the time that the image was taken, rather than all of the updates during the period.

With Change Volumes, a FlashCopy mapping exists between the primary volume and the primary Change Volume. The mapping is updated on the cycling period (60 seconds to one day). The primary Change Volume is then replicated to the secondary Global Mirror volume at the target site, which is then captured in another change volume on the target site. This provides an always consistent image at the target site and protects the data from being inconsistent during resynchronization.

Copy services interoperability with the SAN Volume Controller, Storwize V7000 subsystem, and Storwize V7000 Unified system

With Version 6.3.0, the *layers* concept was introduced in the Storwize V7000 and Storwize V7000 Unified systems. Layers determine how the Storwize V7000 and Storwize V7000 Unified system interact with the SAN Volume Controller. Currently, there are two layers, *replication* and *storage*. All devices must be Version 6.3.0 or later, and the Storwize V7000 and Storwize V7000 Unified systems must be set to be the replication layer when in a copy relationship with the SAN Volume Controller.

The replication layer is for when you want to use the Storwize V7000 or the Storwize V7000 Unified systems with one or more SAN Volume Controllers as a remote copy partner. The storage layer is the default mode of operation for the Storwize V7000 system. It is for when you want to use the Storwize V7000 system to present storage to an SAN Volume Controller.

8.2 File system level copy services of the Storwize V7000 Unified system file modules

This section provides an overview of how the Storwize V7000 Unified system file modules implement copy services. We describe the two main features: *snapshots* and *asynchronous replication*.

8.2.1 Snapshots of file systems and file sets

The Storwize V7000 Unified system implements space-efficient snapshots. Snapshots enable online backups to be maintained to provide near instantaneous access to previous versions of data without requiring complete, separate copies or resorting to offline backups.

In the current version, the Storwize V7000 Unified system offers 256 snapshots per file system and 256 per file set. The snapshots can be scheduled or performed by authorized users or by the Storwize V7000 Unified system administrator. IBM Scale Out Network Attached Storage (SONAS) snapshot technology efficiently uses storage by storing only block-level changes between each successive snapshot. Only the changes that are made to the original file system use extra physical storage. This reduces physical space requirements and maximizes recoverability.

Snapshots also support the integration with Microsoft Volume Shadow Copy Service (VSS). You can use the VSS to display an older file or a folder version in Microsoft Windows Explorer. Snapshots are exported to Windows Server Message Block (SMB) clients by the VSS application programming interface (API). Snapshot data can be accessed and copied back through the Previous Versions option in Microsoft Windows Explorer.

8.2.2 Asynchronous replication

Another important feature of the Storwize V7000 Unified system file module software is asynchronous replication. In this section, we provide an overview of how asynchronous replication is designed to conserve bandwidth.

Asynchronous replication is available for replicating incremental changes to another site at the file system level. This is done by using an IBM-enhanced and IBM-supported version of the open source *rsync* tool. The enhancements include the ability to have more than one file module in parallel that can work on the rsync transfer of the files.

Asynchronous replication is has several advantages:

- First, a central policy engine scan for asynchronous replication is run. The high-performance scan engine is used for this scan. As part of the replication, an internal snapshot is made of both the source file system and the target file system.
- Next, mathematical hashes of the source and target snapshots are taken and compared.
- Both file modules then participate in the transfer of changed blocks to the target remote file systems. The internal snapshot at the source side assures that data that is being transmitted maintains integrity and is from a single point in time. The internal snapshot at the target is there to provide a fallback point-in-time capability in case the transfer of changes from source to target fails before it is complete.

The first step is to run a central policy engine scan for asynchronous replication. The high-performance scan engine is used for this scan. As part of the replication, an internal snapshot is made of both the source file system and the target file system.

The next step is to make a mathematical hash of the source and target snapshots, and compare them.

The final step is to use the parallel data transfer capabilities by having both file modules participate in the transfer of the changed blocks to the target remote file systems. The internal snapshot at the source side assures that data that is being transmitted maintains integrity and is from a single point in time. The internal snapshot at the target is there to provide a fallback point-in-time capability in case the drain of the changes from source to target fails before it is complete.

The basic steps of Storwize V7000 Unified system asynchronous replication are as follows:

1. Take a snapshot of both the local and remote file system. This ensures that you are replicating a frozen and consistent state of the source file system.
2. Collect a file path list with corresponding status information by comparing the two with a mathematical hash to identify changed blocks.
3. Distribute the changed file list to a specified list of source interface nodes.
4. Run a scheduled process that performs rsync operations on both file modules, for a particular file list, to the destination Storwize V7000 Unified system. Rsync is a well-understood open source utility that picks up the changed blocks on the source Storwize V7000 Unified file system and streams those changes in parallel to the remote system. It then writes them to the target Storwize V7000 Unified file system.

The snapshot at the remote Storwize V7000 Unified system ensures that a safety fallback point is available if there is a failure in the drain of the new updates.

When the transfer is finished, the remote file system is ready to use. Both snapshots are automatically deleted after successful replication.

Figure 8-1 shows a simple flow diagram of asynchronous replication.

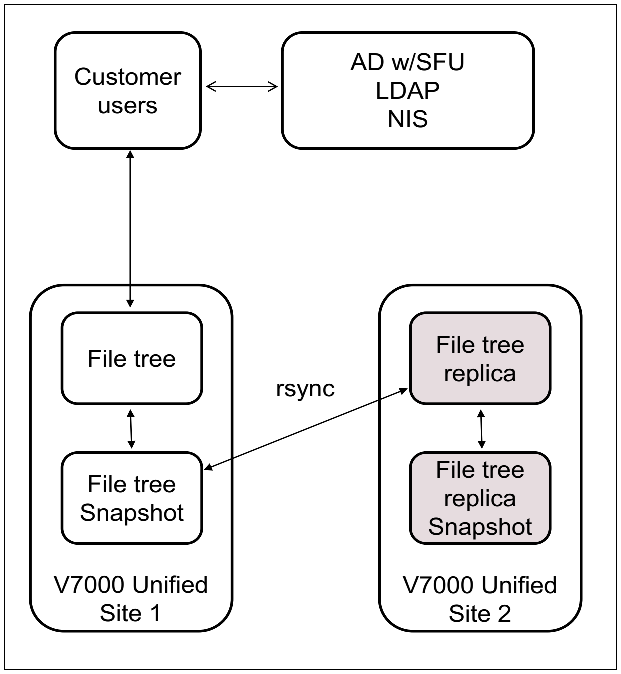


Figure 8-1 Asynchronous replication process

Asynchronous replication limitations

There are limitations that are important to keep in mind when you are using the asynchronous replication function:

- The asynchronous replication relationship is configured as a one-to-one relationship between the source and target.
- The entire file system is replicated in asynchronous replication. Although you can specify paths on the target system, you cannot specify paths on the source system.
- The source and target cannot be in the same system.
- Asynchronous replication processing on a file system can be affected by the number of files that are migrated from the file system. Asynchronous replication of a source file system causes migrated files to be recalled and brought back into the source file system during the replication process.
- File set information about the source system is not copied to the target system. The file tree on the source is replicated to the target, but the fact that it is a file set is not carried forward to the target system’s file tree. File sets must be created and linked on the target system before initial replication because a file set cannot be linked to an existing folder.
- Quota information is also not carried forward to the target system’s file tree. Quotas can be set after initial replication, as required, by using quota settings from the source system.
- Neither Active Directory alone nor Active Directory with NIS, using Storwize V7000 Unified internal UID and GID mapping, are supported by asynchronous replication, because the mapping tables in the Storwize V7000 Unified system clustered trivial database (CTDB) are not transferred by asynchronous replication. If asynchronous replication is used, the user ID mapping must be external to the Storwize V7000 Unified system.

Other considerations

For the first occurrence of running asynchronous replication, consider transporting the data to the remote site physically and have replication take care of changes to the data. Asynchronous replication is no faster than a simple copy operation. Ensure that adequate bandwidth is available to finish replications on time.

There is no mechanism for throttling on asynchronous replication. IBM General Parallel File System (GPFS) balances the load between asynchronous replication and other processes.

Source and target root paths that are passed as parameters must not contain spaces, white space characters, nor any of these characters or marks:

,	comma
(or)	parentheses
‘	apostrophe (single quotation mark)
“	quotation mark (also called double quotation mark)
:	colon
?	question mark
!	exclamation mark
%	percent symbols
/ or \	forward slash or back slash
\n	
\r	
\t	

More detailed information is available in the “Managing asynchronous replication” section of the IBM Storwize V7000 Unified Information Center:

http://ibm.co/1iakvpx

9

GUI and CLI setup and use

The primary interface for the IBM Storwize V7000 Unified Disk System is the graphical user interface (GUI) where all configuration and administration functions can be performed. All functions can also be performed by using the terminal-based command-line interface (CLI). A few specialized commands are available only in the CLI, which might also be required during recovery if the GUI is unavailable. Both methods of issuing commands are required to manage the cluster.

In this chapter, we demonstrate how to set up both methods of access, show how to use them, and explain when each is appropriate.

9.1 Graphical user interface setup

Almost all of the IP addresses in the cluster have a web interface that is running behind them, but each has a specific purpose.

9.1.1 Web server

Each node in the cluster has a web server that is running. What is presented by each of these web servers depends on the functional status and configuration of the particular node at any specific time. All web connections use the HTTPS protocol. If a connection is attempted by using HTTP, it is usually redirected.

Storage node canisters

Both the storage nodes in the control enclosure can be connected to by using their service IP address. This displays the Service Assistant panel for that node. This is a direct connection to the node software and does not require that the cluster is active or operational. It requires only that the node has booted its operating system.

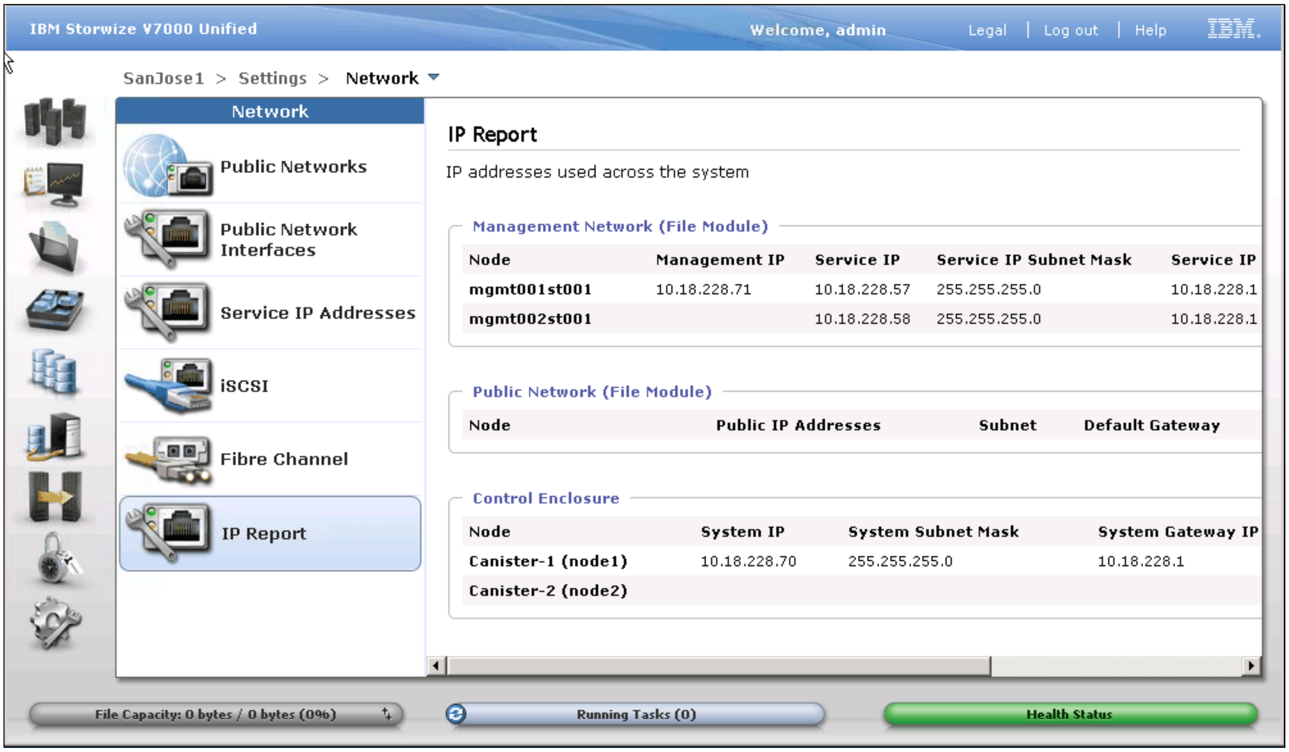
One of the nodes assumes the role of the config node when the cluster is active. That node presents the storage management IP address and presents the storage system management GUI. Only one of the nodes presents this, and there is only one address.

File nodes

For management functions, one of the file nodes is the active management node. This node presents the management IP address and the management GUI for the entire cluster.

Both the file nodes are connectable with HTTPS over the other IP addresses that are assigned to their interfaces. What is presented depends on the configuration of the cluster.

The IP Report that is shown in Figure 9-1 can be accessed by selecting **Settings** → **Network** → **IP Report**. It shows all of the IPs that are being used to manage the file module nodes and control enclosure nodes.



IP Report

IP addresses used across the system

Management Network (File Module)

Node	Management IP	Service IP	Service IP Subnet Mask	Service IP
mgmt001st001	10.18.228.71	10.18.228.57	255.255.255.0	10.18.228.1
mgmt002st001		10.18.228.58	255.255.255.0	10.18.228.1

Public Network (File Module)

Node	Public IP Addresses	Subnet	Default Gateway

Control Enclosure

Node	System IP	System Subnet Mask	System Gateway IP
Canister-1 (node1)	10.18.228.70	255.255.255.0	10.18.228.1
Canister-2 (node2)			

Figure 9-1 IP Report

9.1.2 Management GUI

The primary management interface for the Storwize V7000 Unified cluster is the management IP address that is assigned to the file modules. This GUI combines all management functions and can be used for both file and block storage management.

The storage system or control enclosure also has a management interface, which is the same as the management GUI found on the stand-alone Storwize V7000. This can also be connected to at any time, but provides management of the storage function only. Access to resources directly used by the file modules is prohibited, but normal block configuration can be done. It is suggested that you use only the full cluster GUI presented from the file module to avoid confusion, although there might be times during complex recovery when an IBM Support representative asks you to connect to this interface.

You need to access the storage GUI during implementation to set passwords and to test its functionality.

9.1.3 Web browser and settings

To connect to the GUI, you need a workstation that is running an approved web browser. Generally, any current browser is supported, but to see the current list of supported browsers check the *V1.4 Supported Interoperability List for IBM Storwize V7000 Unified* on the IBM Support portal:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004228

At the time of writing, Mozilla Firefox 3.5 or later and Microsoft Internet Explorer (IE) 8.x or later are supported.

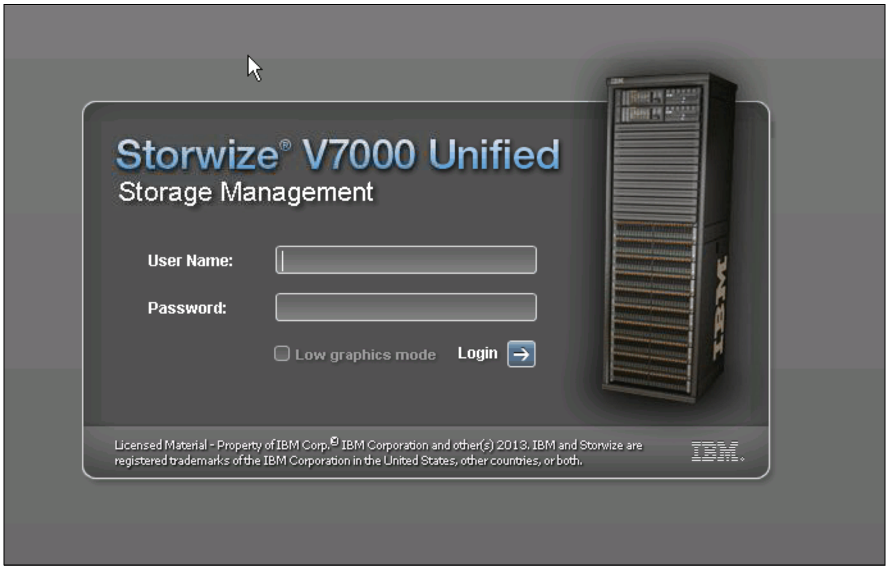
To access the management GUI, you must ensure that your web browser is supported and has the appropriate settings enabled. For browser settings, see *Checking your web browser settings for the management GUI* in the IBM Storwize V7000 Unified Information Center:

http://ibm.biz/BdxFX2

9.1.4 Starting the browser connection

Start the browser application and enter the management IP address that is assigned to the file modules. If you used `http://`*`ip_address`*, you are redirected to `https://`*`ip_address`*. You are then warned that there is a security exception and you need to approve the exception to continue. This step is normal for this type of HTTPS connection.

Next, you see the logon page that is shown in Figure 9-2.



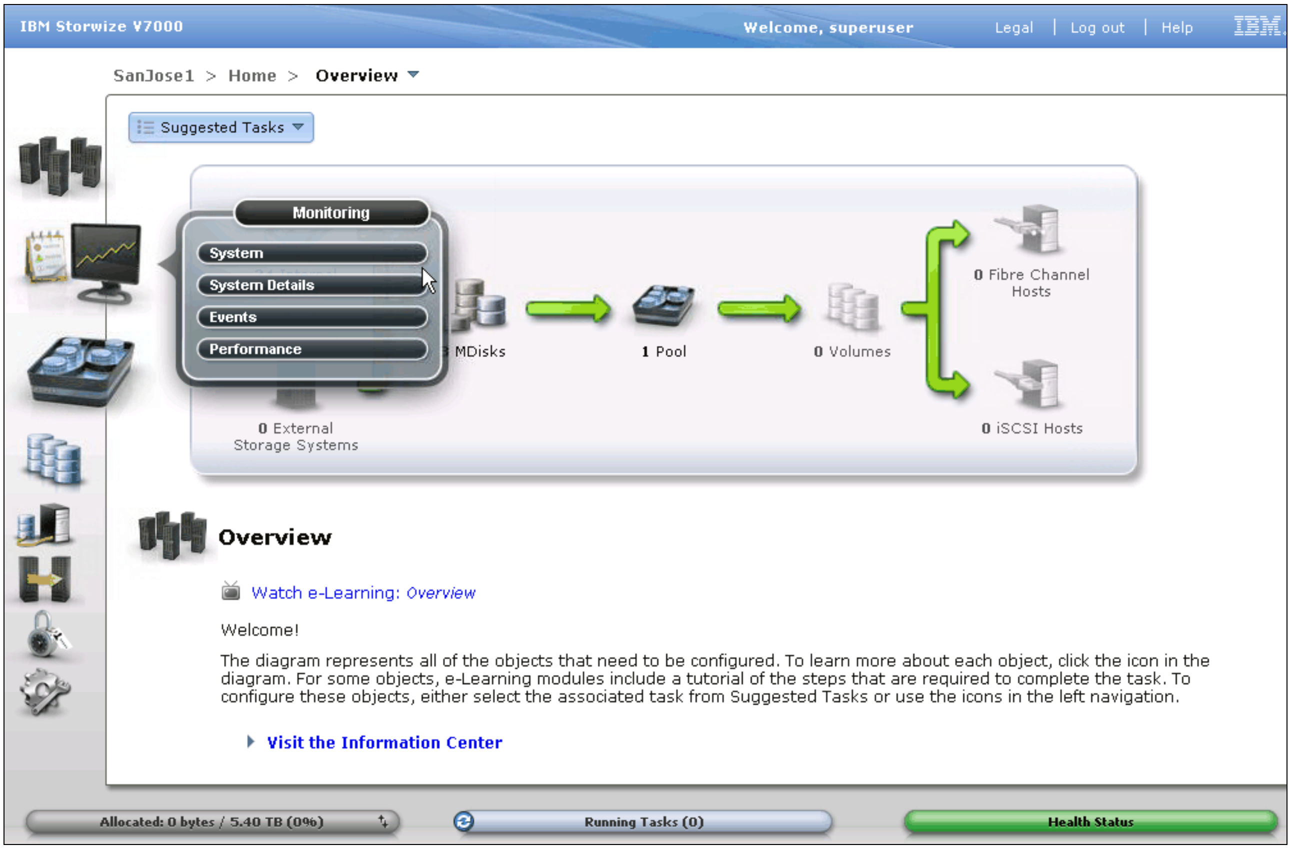
Licensed Material - Property of IBM Corp.© IBM Corporation and other(s) 2013. IBM and Storwize are registered trademarks of the IBM Corporation in the United States, other countries, or both.

Figure 9-2 GUI logon

Enter your user name and password. If this is a new installation, the default is `admin/admin`. Otherwise, you need to use the user ID and password that is assigned to you by your storage administrator.

Notice the box on this window labeled *Low Graphics Mode*. This option disables the animated graphics on the management pages and provides a simplified graphics presentation. This is useful if you are connecting remotely, because it reduces the traffic and improves response time. Some users prefer to disable the animation by using this option.

With animation on, hover over the icons on the left side and the submenu choices are presented. See Figure 9-3. Using the mouse, select the submenu to start the wanted page.

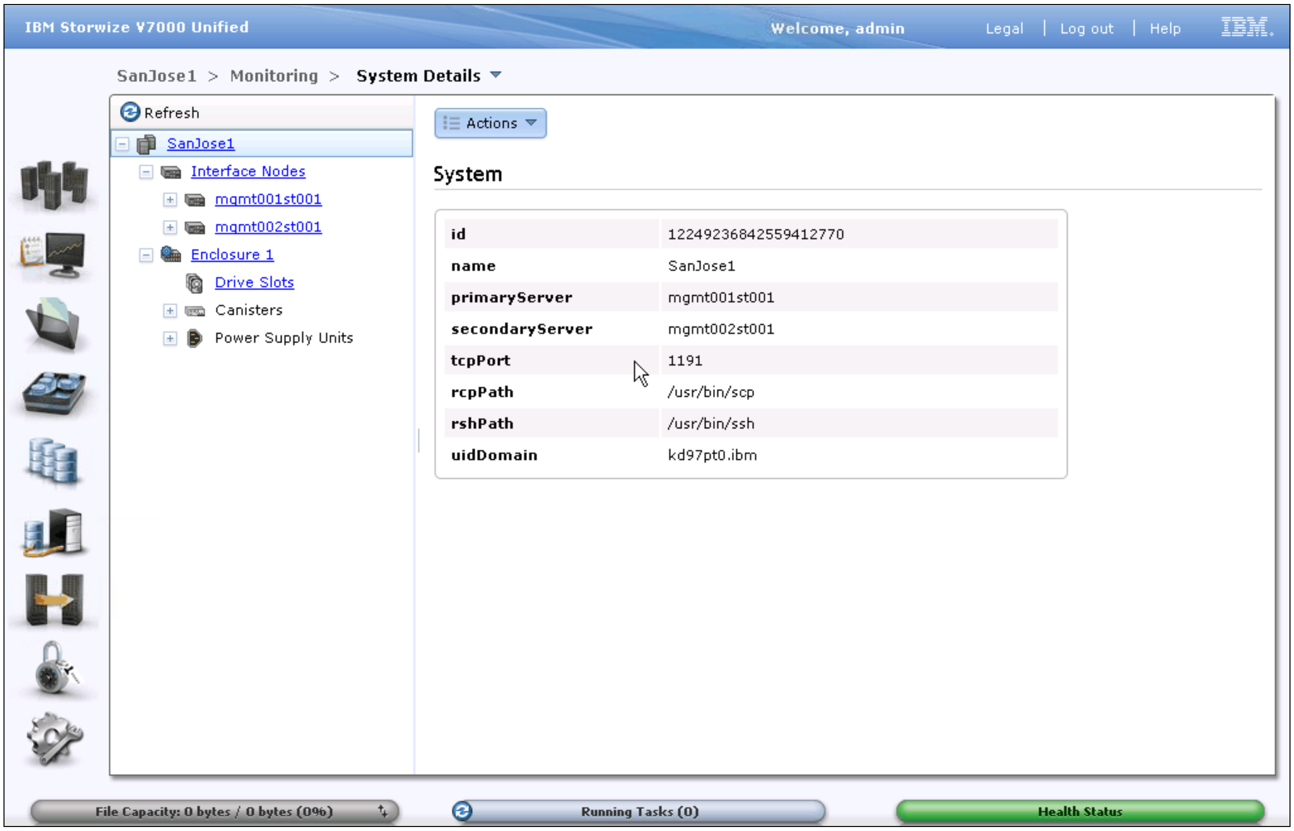


Welcome!

The diagram represents all of the objects that need to be configured. To learn more about each object, click the icon in the diagram. For some objects, e-Learning modules include a tutorial of the steps that are required to complete the task. To configure these objects, either select the associated task from Suggested Tasks or use the icons in the left navigation.

Figure 9-3 GUI animation

Alternatively, if you disabled animation, click the icon to show that section first. Then, by using the pull-down menus for the icons at the top of the page, select the subheading, as shown in Figure 9-4 on page 88.



System

id	12249236842559412770
name	SanJose1
primaryServer	mgmt001st001
secondaryServer	mgmt002st001
tcpPort	1191
rcpPath	/usr/bin/scp
rshPath	/usr/bin/ssh
uidDomain	kd97pt0.ibm

Figure 9-4 GUI no animation

9.2 Using the GUI

All management and configuration functions for both file and block are available through the Storwize V7000 Unified management GUI interface. For this reason, there is no need to connect to the GUI interface of the storage module for normal operations.

When logged on to the Management GUI, the main window is displayed. The window has five main areas:

Top action bar This is the blue bar across the top. It has a Welcome message and includes links to Help and Information. It also has a logout link to close all functions and log off the GUI.

Main section icons Along the left side of the window, there are several icons. Each represents a main section of management. Hovering the mouse cursor over each icon causes it to become larger, and a submenu opens to the right. If low graphics mode was selected, click the icon to display the topic and chose the submenu by using the navigation.

Navigation menu Along the top of the main window, there is a menu that shows the currently displayed panel and which section and menu item it belongs to. If the submenu has multiple choices, it is shown as a pull-down menu that you can use to select the submenu.

Main window The current window is displayed in the right (largest) panel, which is based on the selections made. The contents of this window vary depending on the action being performed.

Bottom status bar At the bottom are three bars:

- The left bar indicates the current file capacity of the cluster and how much is already used.

- The middle bar indicates the number of background tasks that are running.

- On the right side, the bar gives information about the health of the cluster. Normally this bar is colored green, but it changes to yellow or red if there are exceptions. Hover your mouse cursor over the X at the left to open a list of the major components that have unhealthy statuses. It indicates the highest-priority status on those components.

9.2.1 Menus

The following sections describe the menus that are available in the GUI.

Home

The Home menu has only one submenu, the Overview. The Overview window displays a graphical view of the entire cluster from a data point of view. It shows the major areas where data is managed and each icon also gives the number of resources that are defined in that area. The icons are arranged to show the relationship between each and the data flow.

The suggested task button gives a list of shortcuts to common tasks.

Clicking an icon displays a brief description of that resource at the bottom of the window.

Monitoring

The Monitoring menu displays the following submenus:

System This gives a graphical view of the cluster, showing each major component. The graphic for each component indicates its status with colored indicators. Hovering your mouse cursor over the links or clicking the item displays windows that show the status and configuration details. The identify button turns on the attention light on each component to help in locating the physical device.

System Details This option gives a tree view of each component. Clicking the entry in the tree view displays details of that component in the right panel. There is an action pull-down icon to select from the available actions for that component. Each component view is unique and gives detailed information about that component, including its status. Where applicable, the event logs relating to that component are listed.

Events This menu option is described in detail in 15.2, “Event logs” on page 259. There are two tabs that display the two independent event logs, *file* and *block*. The view of each log can be customized and filtered by selecting the wanted options in the filter controls at the top of the panel. Both logs provide an action pull-down icon that acts on the highlighted line entry in the view that follows. It is also possible to right-click a log entry directly to show this action list. The choices in the action list vary between the two logs.

Capacity This menu option gives a window with five tabs. The *File Systems* tab shows a list of the file systems, their total capacity, and use details. Clicking each file system name causes it to be included in the graph that is displayed at the bottom of the window, which tracks historic use over time, or if you prefer, the percentage. The *File System Pools* tab shows a list of the file systems, their total capacity, and specific use details. It also shows details that are related to compression and thin provisioning. The *File Sets* tab lists the file sets defined and gives usage metrics on each one. The *Users* tab gives file use by each user that is defined to the cluster. The *User Groups* tab gives a higher-level view that is based on the groups the users belong to.

Performance The performance option has three tabs and gives a simple view of some key performance indicators. The graphs that are shown are not meant to provide detailed tuning information, but to show at a quick glance the areas of immediate concern that might need further investigation or to quickly identify a problem area during a performance impact on the cluster. The *File* tab shows four graphs, and you can alter the scale by using the pull-down menu on the right side. The *Block* tab shows four graphs with the scale fixed to 5 minutes. The scope can be changed to show the whole cluster or one node. The *File Modules* tab shows graphs, and you can alter the scale by using the pull-down menu on the right side.

Files

All configuration actions for the file services are performed in this menu. These functions are covered in detail in the Chapter 11, “Implementation” on page 141. The Files menu presents the following submenus:

File Systems Use this option to view the status and manage the file systems that are configured on the cluster. Click the **New File System** button to create a file system or the **Actions** pull-down menu to perform management functions on an existing one. You can determine whether the file system is compressed or not and also see capacity information. You can also filter whether to show NSD or storage pool details for each file system.

Shares This option lists all shares or exports that are defined on the cluster, the path for their root, and the protocol they are able to be accessed with. Use the **New Share** button to create a share. This action starts a window to enter the details. Or, use the **Actions** pull-down menu to manage an existing share.

File Sets This menu option shows the defined file sets in a list that gives details about their types, paths, which file system they are in, and statistical details. Click the **New File Set** button to define a new file set and the **Actions** pull-down menu to manage an existing one.

Snapshots In this option, you can create a snapshot, or manage an existing one from the list that is displayed by using the **New Snapshot** and **Actions** pull-down menus.

Quotas In this option, you can create a quota by using the **New Quota** pull-down menu, or manage an existing one from the list that is displayed by using the **Actions** pull-down menu.

Services The services tab is used to configure and manage the additional tools that are provided for the file service:

- Backup selection gives a choice of which backup technology is used to back up the file service. At the time of writing, two options are available: IBM Tivoli Storage Manager and Network Data Management Protocol (NDMP).

- The backup option display is technology-specific and is used to configure the backup process.

- The antivirus selection is used to configure the external antivirus server if antivirus scanning is being used.

Pools

The storage pools are a pool of storage from which volumes are provisioned and used as block storage by servers directly. These pools are also used by the file server to form file systems. Resources that are owned by the file server do not show in all the GUI views, but the capacity that is used is seen. The display for each pool also displays details that are related to compression. This menu gives several views:

Volumes by Pool Clicking the pool (or MDisk group) in the left panel displays the volumes in that group and their details in the right panel. You can use the New Volume tab to create new volumes for blocked storage to servers and also the Actions tab to manage these same volumes. You can monitor only NSDs that are the volumes that are assigned to file systems such as examining properties.

Internal Storage The Storwize V7000 has internal disk drives in the enclosures. This option displays and manages these drives. Click the drive class in the left panel to display the drives in that class.

External Storage Storwize V7000 can also manage external storage subsystems if wanted by using the SAN connection. If any are attached, they are managed in this option. Click the storage system controller in the left panel to display the volumes that are presented in the right panel.

MDisks by Pools This option gives a different view of the pools. Here, you can see which MDisks are in each pool.

System Migration This wizard is to assist with migrating an external storage system to be managed by the Storwize V7000.

Volumes

The volumes are built from extents in the storage pools and presented to hosts as external disks. There are several types of block volumes, such as thin-provisioned, compressed, uncompressed, or generic, and mirrored volumes. In this view, you can create, list, and manage these volumes:

Volumes This is a listing of all volumes.

Volumes by Pool By selecting the pool in the left panel, you can display the volumes that are built from that pool.

Volumes by Host The hosts that are defined to the cluster are listed in the left panel. By clicking a host, you can see which volumes are mapped to that host. The file modules are hosts also, and they use block volumes (NSDs) but do not appear as hosts.

Hosts

Each host that will access block volumes on the cluster needs to be defined. In each definition, there also needs to be defined the worldwide name (WWN) or Internet Small Computer System Interface (iSCSI) details of the ports of that host. When the host is defined, then volumes can be mapped to it, which are then visible to the ports with the WWNs listed:

Hosts This is a list of all defined hosts. Here, you can add and manage these definitions.

Ports by Host In this view, you can see the ports that are defined on each host. The hosts are listed in the left panel. Clicking the host displays the ports in the right panel.

Host Mappings Each mapping that shows the host and the volume that is mapped is listed, one per line.

Volumes by Host In this view, you can select the host from the left panel and see volumes that are mapped to it in the right panel.

Copy services

Storwize V7000 Unified provides several different methods of coping and replicating data. FlashCopy is provided for instant copy of block volumes within the cluster. Remote copy is used to copy block volumes to another location on another cluster, and this can be done synchronously (with Metro Mirror) or asynchronously (with Global Mirror). File systems can be replicated to another file system by using the File Copy Services submenu:

FlashCopy In this option, all the volumes in the cluster are listed. Here, you can create and manage copies and view the status of each volume.

FlashCopy mappings With this option, you can create and view the relationship (mapping) between the FlashCopy source and target volumes.

Consistency groups These are used to group multiple copy operations together if they need to be controlled at the same time. In this way, the group can be controlled by starting, stopping, and so on, with a single operation. The function also ensures that when it stops for any reason, the I/Os to all group members all stop at the same time, in terms of what the host writes to the primary volumes. This ensures time consistency across volumes.

Remote copy In this option, you can create remote copies and consistency groups. You can then view and manage these.

Partnerships For a remote copy to be used, a partnership must be set up between two or more clusters. This option is used to create and manage these partnerships.

File copy services Use this panel to select different methods to replicate data between different file systems.

Access

There are several levels of user access to the cluster, which are managed in this option. The access levels are divided into groups, each having a different level of access and authority. You can define multiple users and assign different access levels to suit the tasks that they perform:

Users This option lists the user groups in the left panel, and the users in that group in the right panel. New users can be added to a group and managed.

Audit log All commands that are issued on the cluster are logged in this log. Even if initiated from the GUI, most actions cause a CLI command to be run, so this is also logged.

Local authentication The system supports user authentication and ID mapping by using a local authentication server for network-attached storage (NAS) data access. Using local authentication eliminates the need for a remote authentication service, such as Active Directory or Samba primary domain controller (PDC). This simplifies configuration and management of authentication policies.

Settings

Use the Settings panel to configure system options for event notifications, Directory Services, IP addresses, and preferences that are related to display options in the management GUI:

Event Notifications This option is used to configure the alerting and logging. Here, you define the email and SNMP servers and the alert levels that you prefer. For more information, see 15.5, "Call Home feature and alerts" on page 276.

Directory Services Directory Services defines the fundamental settings for the file server. You need to define the DNS domain and servers. Define the authentication method that is used by the file server and the authentication server.

Network Protocol If HTTP is configured as an access protocol on any shares, define the HTTPS security. Web access is by HTTPS only; pure HTTP protocol is not allowed. Here, you have the option to set up the authentication method and keys that you prefer.

Network The network setup for all the interfaces in the cluster is configured here. Use the buttons in the left panel to select the interface and view or modify the values in the right panel:

- Public Networks define the file access IP addresses that are presented on the client facing interfaces. These addresses float across the file module ports as needed.

- Service IP Addresses are for the storage enclosure only. Define a unique address for port 1 on each node canister. This address is used only for support and recovery.

- iSCSI defines settings for the cluster to attach iSCSI-attached hosts.

- Use the Fibre Channel panel to display the Fibre Channel connectivity between nodes, storage systems, and hosts.

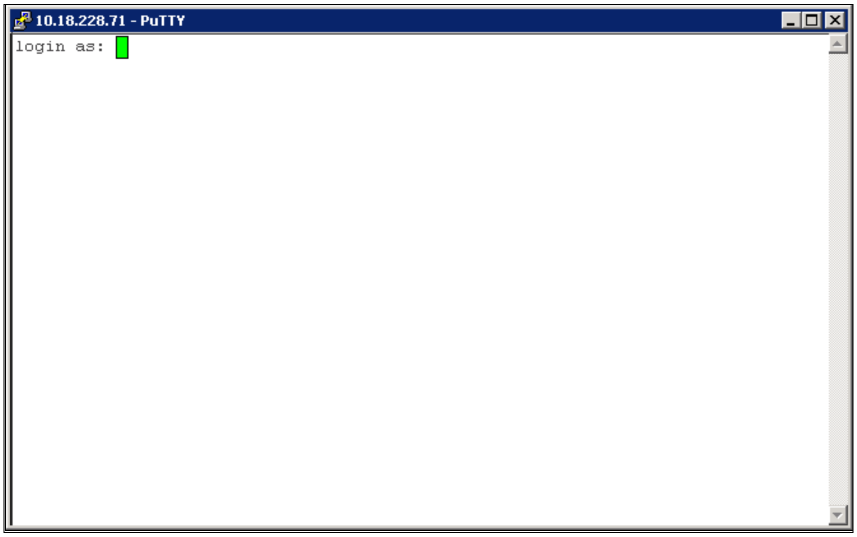
IP Report The IP Report panel displays all the IP addresses that are currently configured on the system.

Support You can use this option to define connectivity for sending alerts to IBM Support and to allow IBM Support representatives to connect to the cluster. You can also create, offload, and manage the data collections IBM Support needs.

General In this option, you can set the time and date for the cluster, enter licensing details if needed, and perform software upgrades for the cluster. The software process is covered in detail in 15.9, "Software" on page 280.

9.3 Command-line interface setup

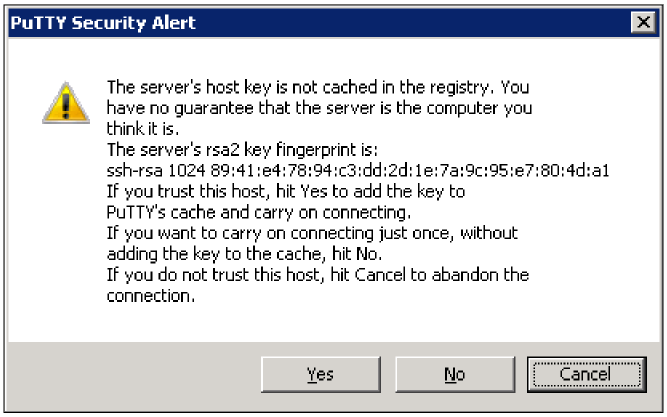
Using a suitable terminal client such as PuTTY, connect to the management IP address by using Secure Shell (SSH) (port 22). You then get a login prompt, as shown in Figure 9-5.



```
login as:
```

Figure 9-5 CLI: Login prompt

If this is the first time that a connection is made from this workstation, you might be asked to accept a security key, as shown in Figure 9-6. Click **Yes** to tell PuTTY to save the rsa key for future connections.



The server's host key is not cached in the registry. You
have no guarantee that the server is the computer you
think it is.
The server's rsa2 key fingerprint is:
ssh-rsa 1024 89:41:e4:78:94:c3:dd:2d:1e:7a:9c:95:e7:80:4d:a1
If you trust this host, hit Yes to add the key to
PuTTY's cache and carry on connecting.
If you want to carry on connecting just once, without
adding the key to the cache, hit No.
If you do not trust this host, hit Cancel to abandon the
connection.

Figure 9-6 PuTTY rsa key

Save the connection definition in PuTTY so it can easily be started in the future.

Also connect, test, and save a session to the storage management IP address. This is used only in a recovery situation, but it is a good practice to have it tested and easy to start beforehand. If you are accustomed to earlier versions of SAN Volume Controller, notice that there is no longer a requirement to create and store a key file. Authentication is now by user ID and password.

9.4 Using the CLI

> **Note:** We suggest using the GUI rather than the CLI. The GUI builds the CLI commands that you need and automatically includes the correct parameters. We also suggest that you use the GUI to determine the best way to use CLI commands. One example is when you are creating a compressed volume that requires some specific parameters, such as **`rsize`** and **`autoexpand`,** to avoid having the volume go offline prematurely because these parameters are missing or misconfigured.

Like the GUI, there is a CLI connection to the Storwize V7000 Unified management address and also to the Storwize V7000 storage enclosure management address. All functions can be performed on the Storwize V7000 Unified, so the only access required for normal operation is this single CLI session. The CLI session to the storage is needed only in recovery situations, but it is a good practice to set it up and test it.

The commands are unique, so storage commands can be issued on the unified CLI by using the same syntax. Most block commands can be prefixed with `svcinfo` or `svctask` as with SAN Volume Controller and Storwize V7000 previously. Where there is ambiguity, this prefix needs to be added. This ensures that the command is unique and gets the correct result.

For example, **`lsnode`** displays information about the file modules, as shown in Example 9-1.

Example 9-1   lsnode file module information

```
[kd97pt0.ibm]$ lsnode
Hostname     IP          Description             Role                          Product version Connection status GPFS
status CTDB status Last updated
mgmt001st001 169.254.8.2 active management node  management,interface,storage 1.4.2.0-27      OK
active      active      10/22/13 1:17 PM
mgmt002st001 169.254.8.3 passive management node management,interface,storage 1.4.2.0-27      OK
active      active      10/22/13 1:17 PM
EFSSG1000I The command completed successfully.
[kd97pt0.ibm]$
```

The **`svcinfo lsnode`** command displays information about the Storwize V7000 nodes, as shown in Example 9-2.

Example 9-2   lsnode Storwize V7000 node information

```
[kd97pt0.ibm]$ svcinfo lsnode
id name  UPS_serial_number WWNN             status IO_group_id IO_group_name config_node UPS_unique_id    hardware
iscsi_name                              iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number
1  node1                   5005076802002B6C online 0           io_grp0       yes         5005076802002B6C 300
iqn.1986-03.com.ibm:2145.sanjose1.node1             01-1       1            1           78G06N1
2  node2                   5005076802002B6D online 0           io_grp0       no          5005076802002B6D 300
iqn.1986-03.com.ibm:2145.sanjose1.node2             01-2       1            2           78G06N1
[kd97pt0.ibm]$
```

The information center has detailed information about the use and syntax of all commands. Most commands are available to all users, but some commands depend on the authority level of the user ID that is logged on.

Scripting of CLI commands is supported if the scripting tool supports SSH calls. See *Adding SSH keys for hosts other than the IBM System Storage Productivity Center* in the information center for details about generating a key and using scripts:

`http://ibm.co/1cXm0jt`

The list that follows includes commands that you might find useful during recovery. Always check the information center for syntax and expected results.

9.4.1 File commands

`lscluster` Lists the clusters that are managed

`lsnode` Lists the nodes in the cluster

`lsnwmgt` Shows the configuration of the management ports

`lsnwinterface` Lists the physical client-facing interfaces

`lsnw` Lists the networks (or subnets) that are defined

`chnwmgt` Sets or changes the addressing of the file module management ports

`chrootpwd` Changes the root password across all nodes (requires root logon)

`initnode` Stops or restarts a file node

`resumenode` Resumes a node that has was suspended or banned

`stopcluster` Shuts down a cluster or node

`suspendnode` Suspends a node

`lsfs` Lists the file systems

`lsmount` Lists the mount status of all file systems

`mountfs` Used to mount a file system but only during recovery

`unmountfs` Unmounts a file system

9.4.2 Block commands

`svc_snap` Gathers a data collection from the block storage Storwize V7000

`lssystem` Lists the Storwize V7000 storage system

`svcinfo lsnode` Lists the nodes in the storage system

`lsdumps` Lists dump files that are saved on the storage system

`lsfabric` Produces a list (often very long) of all of the Fibre Channel paths that are known to the storage system

`lsmdisk` Lists all the MDisks that are visible to the storage system (useful if you need to also see MDisks owned by the file storage, which are hidden in the GUI)

`detectmdisk` Rescans and rebalances Fibre Channel paths (use with care, because this reconfigures the pathing to the current visible paths and drop failed paths)

`chsystemip` Changes or sets the IP addresses of the storage system

`stopsystem` Stops a node or the entire storage system

10

Planning for implementation

In this chapter, we describe the planning steps that are required to prepare for a successful implementation of the IBM Storwize V7000 Unified system. We strongly suggest considering all the solution requirements in advance in order to achieve the best results, rather than an ad hoc, unplanned implementation. Be sure to record these requirements as part of the planning process.

The IBM SONAS and V7000 Unified Disk System questionnaire is a vital resource to complete during the planning phase. We describe the questionnaire in section 10.1, “IBM SONAS and V7000 Unified Questionnaire” on page 98.

The “Planning” section of the IBM Storwize V7000 Unified Information Center is another good resource to consult:

http://ibm.co/1fqJ0vT

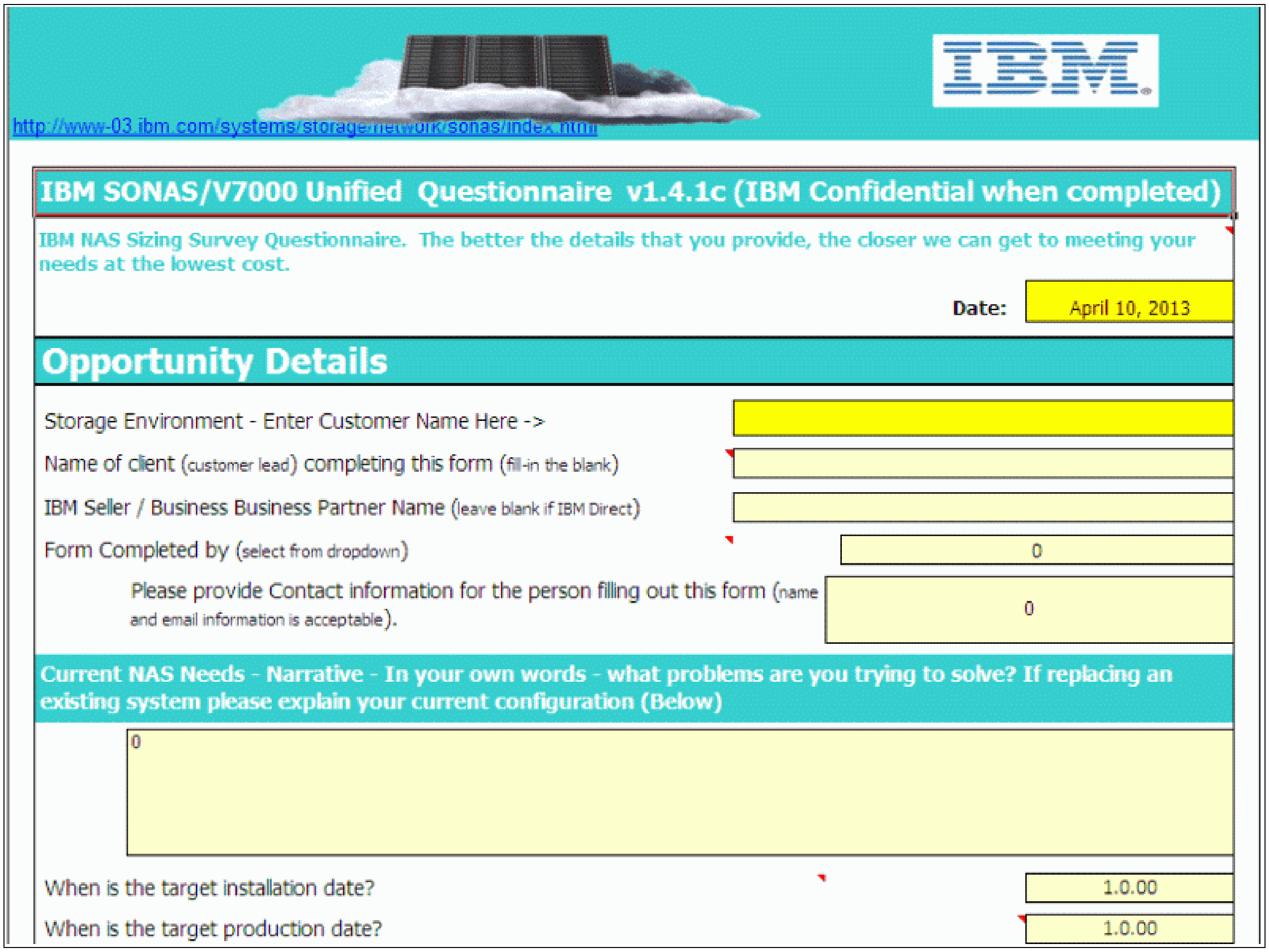
10.1 IBM SONAS and V7000 Unified Questionnaire

Adequate and careful planning is a very important part of solution design and implementation. Without thorough planning, issues will arise during implementation phase and different functional, performance, and capacity problems are likely to occur when the V7000 Unified system is put into production use.

The IBM SONAS/V7000 Unified Questionnaire is a useful tool to assist you during your research activities. The IBM account team gives you the questionnaire so that they can collect and understand all relevant information and requirements for your solution. It is important to take the time and respond as completely and as accurately as possible to ensure the optimal solution planning.

Figure 10-1 shows the beginning section of the IBM SONAS/V7000 Unified Questionnaire.

Important: When complete, the questionnaire is classified as IBM Confidential.



http://www-03.ibm.com/systems/storage/network/sonas/index.html

IBM SONAS/V7000 Unified Questionnaire v1.4.1c (IBM Confidential when completed)

IBM NAS Sizing Survey Questionnaire. The better the details that you provide, the closer we can get to meeting your needs at the lowest cost.

Date: April 10, 2013

Opportunity Details

Storage Environment - Enter Customer Name Here ->	
Name of client (customer lead) completing this form (fill-in the blank)	
IBM Seller / Business Business Partner Name (leave blank if IBM Direct)	
Form Completed by (select from dropdown)	0
Please provide Contact information for the person filling out this form (name and email information is acceptable).	0

Current NAS Needs - Narrative - In your own words - what problems are you trying to solve? If replacing an existing system please explain your current configuration (Below)

0

When is the target installation date?	1.0.00
When is the target production date?	1.0.00

Figure 10-1   IBM SONAS/V7000 Unified questionnaire

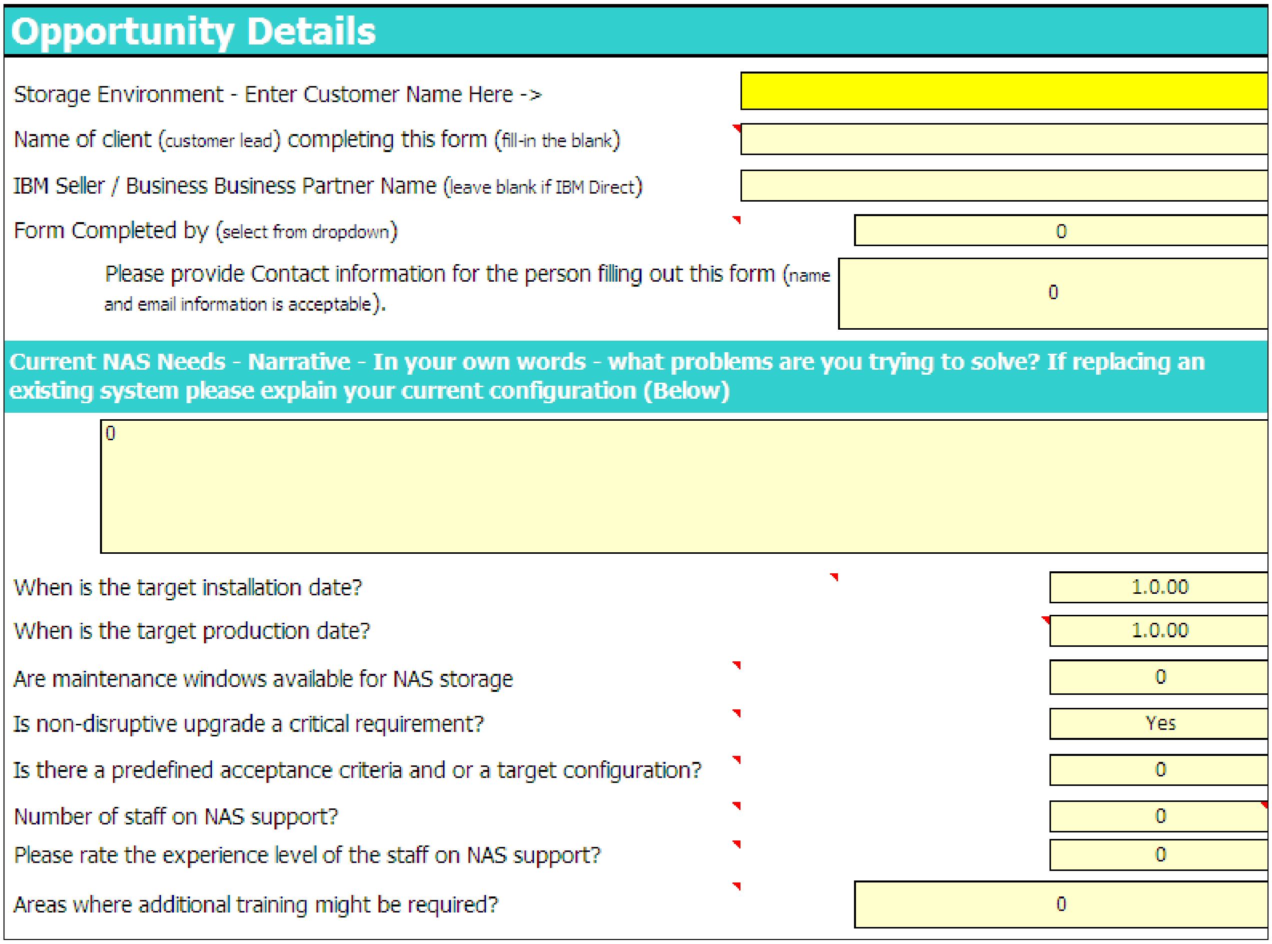
The questionnaire contains the following key areas:

- Opportunity details
- Capacity requirements
- Performance requirements
- Client communication protocols
- Data protection
- Data center requirements
- Storwize V7000 Unified specific questions
- Service requirements

We explain these key areas in more detail in the following sections.

10.1.1 Opportunity details

Figure 10-2 shows the opportunity details section in the questionnaire.



Opportunity Details

Storage Environment - Enter Customer Name Here ->	
Name of client (customer lead) completing this form (fill-in the blank)	
IBM Seller / Business Business Partner Name (leave blank if IBM Direct)	
Form Completed by (select from dropdown)	0
Please provide Contact information for the person filling out this form (name and email information is acceptable).	0

Current NAS Needs - Narrative - In your own words - what problems are you trying to solve? If replacing an existing system please explain your current configuration (Below)

0

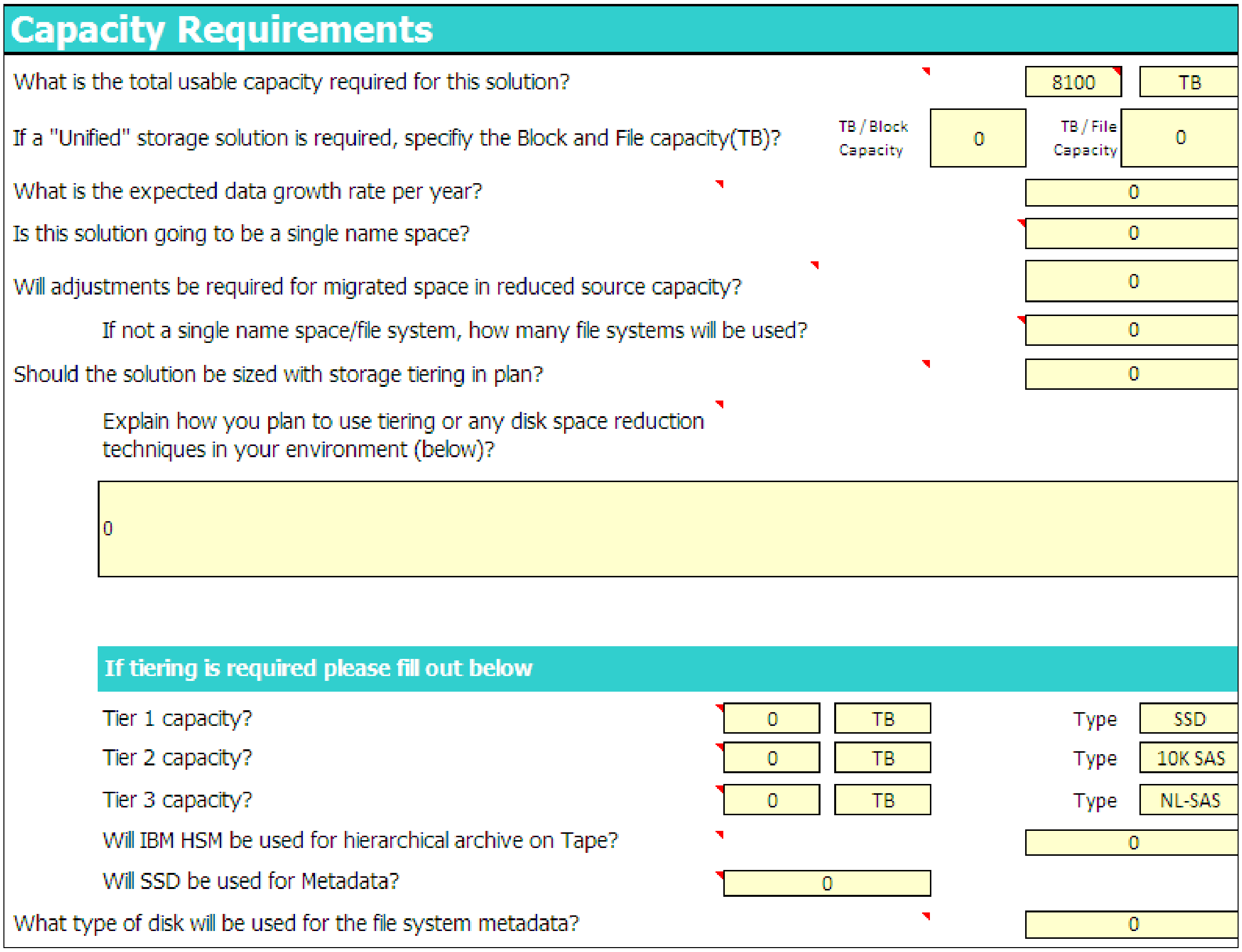
When is the target installation date?	1.0.00
When is the target production date?	1.0.00
Are maintenance windows available for NAS storage	0
Is non-disruptive upgrade a critical requirement?	Yes
Is there a predefined acceptance criteria and or a target configuration?	0
Number of staff on NAS support?	0
Please rate the experience level of the staff on NAS support?	0
Areas where additional training might be required?	0

Figure 10-2 Opportunity details

When completed, this section provides details such as the client name, reseller or IBM Business Partner details, target installation, production dates, and so on.

10.1.2 Capacity requirements

This part of questionnaire is shown in Figure 10-3.



Capacity Requirements

What is the total usable capacity required for this solution? 8100 TB

If a "Unified" storage solution is required, specifiy the Block and File capacity(TB)? TB / Block Capacity 0 TB / File Capacity 0

What is the expected data growth rate per year? 0

Is this solution going to be a single name space? 0

Will adjustments be required for migrated space in reduced source capacity? 0

If not a single name space/file system, how many file systems will be used? 0

Should the solution be sized with storage tiering in plan? 0

Explain how you plan to use tiering or any disk space reduction techniques in your environment (below)?

0

If tiering is required please fill out below

Tier 1 capacity? 0 TB Type SSD

Tier 2 capacity? 0 TB Type 10K SAS

Tier 3 capacity? 0 TB Type NL-SAS

Will IBM HSM be used for hierarchical archive on Tape? 0

Will SSD be used for Metadata? 0

What type of disk will be used for the file system metadata? 0

Figure 10-3 Capacity requirements

In this section of the questionnaire, we gather information about your required capacity for block and file I/O, expected data growth rate, file systems, and tiering details.

You can use the Capacity Magic modeling tool to determine the configuration for your required capacity, based on the data that you enter.

Information lifecycle management

File systems for information lifecycle management (ILM) require multiple internal file system pools to be defined in different tiers. These tiers are then mapped to different storage pools, which it is best to base on storage tiers, for example, drive classes and drive technology.

Create a plan for the lifecycle of a file. Base it on file type, time since last modification, or time since last access. Then, based on the file capacities that you need for the different tiers, provide the corresponding capacity in storage tiers. This determines the type and number of disk drives to order for the back-end storage.

Determine the policy definitions that you want for the following components:

- Data placement at file creation
- Data migration between the tiered pools during the lifetime of the file
- Data deletion after the file’s specified expiration criteria are met

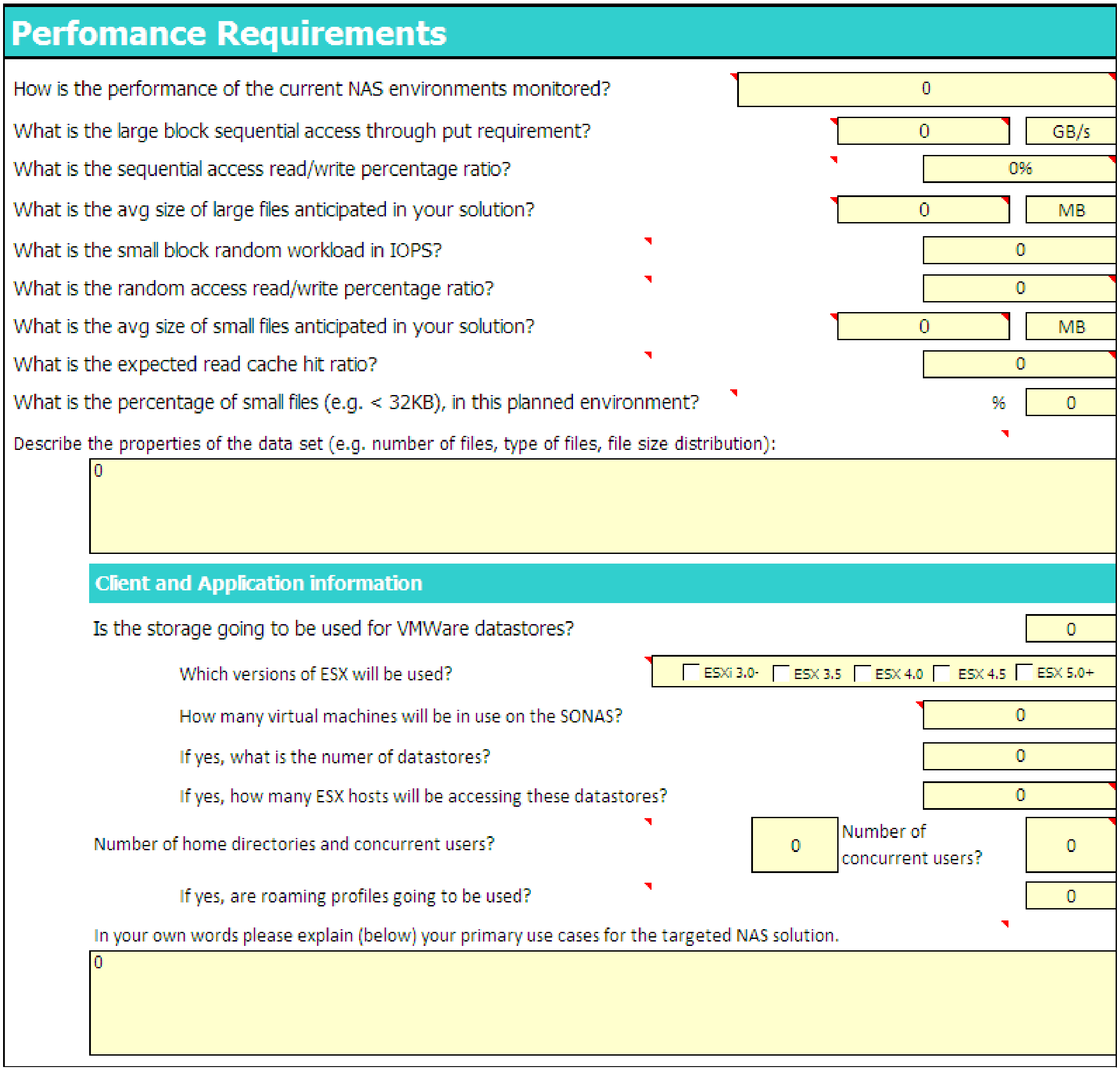
Hierarchical storage management

With hierarchical storage management, consider the following factors:

- Hierarchical storage management (HSM) works with IBM Tivoli Storage Manager as the backup method
- HSM is not supported by Network Data Management Protocol (NDMP) backup
- Requires software and a license for Tivoli for Space Management
- Requires an external file system pool to be configured
- Requires external storage that is supported by Tivoli Storage Manager HSM

10.1.3 Performance requirements

We show the performance requirements section of the questionnaire in Figure 10-4 on page 102.



Perfomance Requirements

How is the performance of the current NAS environments monitored? 0

What is the large block sequential access through put requirement? 0 GB/s

What is the sequential access read/write percentage ratio? 0%

What is the avg size of large files anticipated in your solution? 0 MB

What is the small block random workload in IOPS? 0

What is the random access read/write percentage ratio? 0

What is the avg size of small files anticipated in your solution? 0 MB

What is the expected read cache hit ratio? 0

What is the percentage of small files (e.g. < 32KB), in this planned environment? % 0

Describe the properties of the data set (e.g. number of files, type of files, file size distribution):

0

Client and Application information

Is the storage going to be used for VMWare datastores? 0

Which versions of ESX will be used? □ ESXi 3.0- □ ESX 3.5 □ ESX 4.0 □ ESX 4.5 □ ESX 5.0+

How many virtual machines will be in use on the SONAS? 0

If yes, what is the numer of datastores? 0

If yes, how many ESX hosts will be accessing these datastores? 0

Number of home directories and concurrent users? 0 Number of concurrent users? 0

If yes, are roaming profiles going to be used? 0

In your own words please explain (below) your primary use cases for the targeted NAS solution.

0

Figure 10-4 Performance requirements

The questions in this part try to determine the factors important for good performance. These factors include the expected type of I/O (sequential access I/O versus random access I/O), large files versus small files, expected cache hit ratio, number of concurrent users, and so on. If you plan to use storage in a VMware environment, you also need to identify VMware vSphere version, number of virtual machines in use, number of data stores, and number of VMware ESXi hosts.

If this is an existing environment or if a test environment is available, the workloads that you have experienced can be measured to project future ones. Tools are available at no cost to analyze workloads and gather the necessary information about the system components, I/O patterns, and network traffic. For Windows environments, *perfmon* can be used. In IBM AIX and Linux environments, *nmon* is one of the options to use. There are other options, such as: traceroute, netstat, tcptrace, tcpdump, iozone, iorate, netperf, nfsstat, iostat, and others.

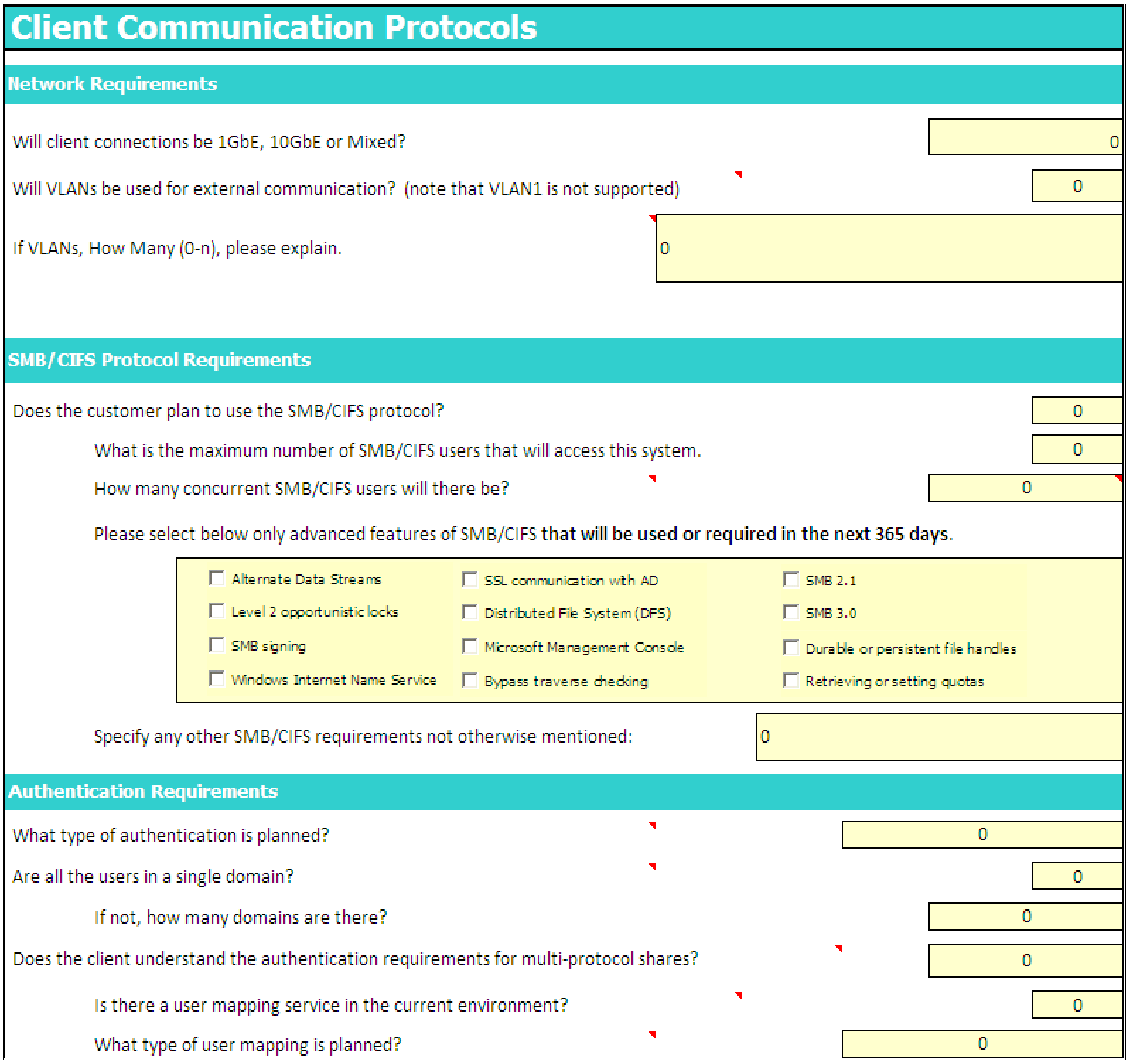
The Disk Magic modeling tool can be used to verify that performance requirements can be met with the system configuration determined by Capacity Magic. You can adjust drive types, Redundant Array of Independent Disks (RAID) levels, and the number of drives required, accordingly.

10.1.4 Client communication protocols

This is the largest part of the questionnaire, and its purpose is to collect information about communication protocols that you intend to use. The subsections cover these topics:

- Network requirements
- SMB/CIFS protocol requirements
- Authentication requirements
- NFS protocol requirements
- Other client and protocol requirements

We show the first part of the IBM client communication protocols section in Figure 10-5 on page 104.



Client Communication Protocols

Network Requirements

Will client connections be 1GbE, 10GbE or Mixed? 0

Will VLANs be used for external communication? (note that VLAN1 is not supported) 0

If VLANs, How Many (0-n), please explain. 0

SMB/CIFS Protocol Requirements

Does the customer plan to use the SMB/CIFS protocol? 0

What is the maximum number of SMB/CIFS users that will access this system. 0

How many concurrent SMB/CIFS users will there be? 0

Please select below only advanced features of SMB/CIFS **that will be used or required in the next 365 days**.

□ Alternate Data Streams	□ SSL communication with AD	□ SMB 2.1
□ Level 2 opportunistic locks	□ Distributed File System (DFS)	□ SMB 3.0
□ SMB signing	□ Microsoft Management Console	□ Durable or persistent file handles
□ Windows Internet Name Service	□ Bypass traverse checking	□ Retrieving or setting quotas

Specify any other SMB/CIFS requirements not otherwise mentioned: 0

Authentication Requirements

What type of authentication is planned? 0

Are all the users in a single domain? 0

If not, how many domains are there? 0

Does the client understand the authentication requirements for multi-protocol shares? 0

Is there a user mapping service in the current environment? 0

What type of user mapping is planned? 0

Figure 10-5 Client communication protocols (first part)

Figure 10-5 shows fields that belong to the following subsections:

- Network requirements

  This subsection contains fields for entering information about file I/O network parameters, such as network speed and VLAN details.

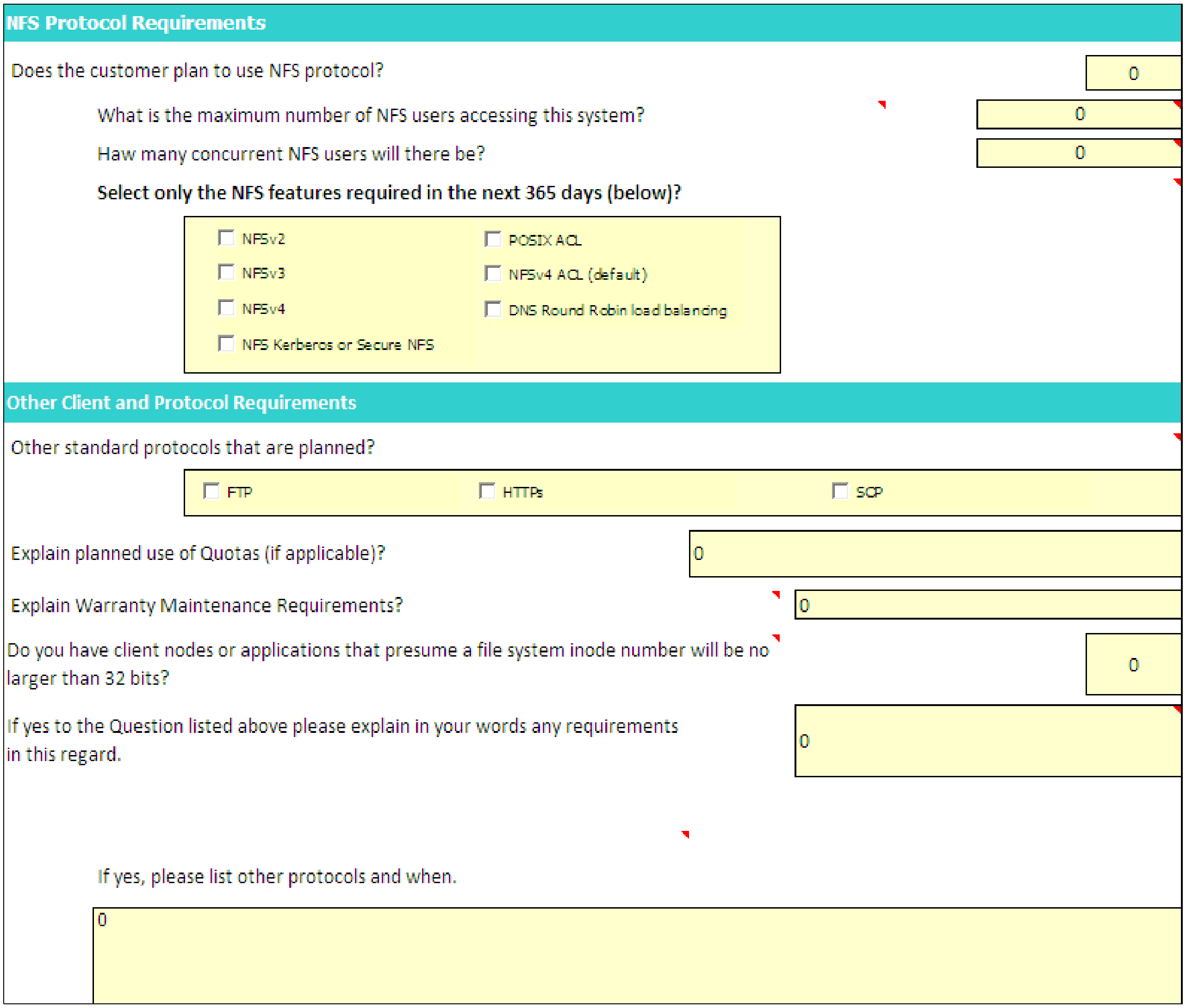
- SMB/CIFS requirements

  If you plan to use SMB/CIFS protocol, complete this subsection.

- Authentication requirements

  In this subsection, authentication type, number of user domains and other pertinent information can be provided.

The second part of client communication protocols section is shown in Figure 10-6 on page 105.



NFS Protocol Requirements

Does the customer plan to use NFS protocol? 0

What is the maximum number of NFS users accessing this system? 0

Haw many concurrent NFS users will there be? 0

Select only the NFS features required in the next 365 days (below)?

□ NFSv2 □ POSIX ACL

□ NFSv3 □ NFSv4 ACL (default)

□ NFSv4 □ DNS Round Robin load balancing

□ NFS Kerberos or Secure NFS

Other Client and Protocol Requirements

Other standard protocols that are planned?

□ FTP □ HTTPs □ SCP

Explain planned use of Quotas (if applicable)? 0

Explain Warranty Maintenance Requirements? 0

Do you have client nodes or applications that presume a file system inode number will be no larger than 32 bits? 0

If yes to the Question listed above please explain in your words any requirements in this regard. 0

If yes, please list other protocols and when.

0

Figure 10-6 Client communication protocols (second part)

This part shows the following subsections:

- NFS protocol requirements

  This subsection needs to be completed if you plan to use NFS. Provide details such as the maximum number of NFS users and the number of concurrent NFS users.

- Other client and protocol requirements

  The use of protocols such as FTP, HTTPS, and SCP can be indicated here, along with other relevant information.

Network considerations

The Storwize V7000 Unified system uses the Storwize V7000 Storage System as the back-end storage system for internal storage. Therefore, all of the LAN considerations for the Storwize V7000 system apply to the back-end part also:

- In contrast to the ports on the file modules, the network ports on the Storwize V7000 system are *not* bonded by default.

- The 1 Gb Ethernet Port 1 on both node canisters is used for management access through the Storwize V7000 cluster IP address by default:
  - Optionally, use Port 2 to define the second management IP address for redundancy.
  - As with the stand-alone Storwize V7000, the management IP address is active on Port 1 of the current configuration node canister. Either one of the two node canisters in a V7000 can act as the configuration node canister. Configuration node role changes between the two in case of problems, changes, or during V7000 code updates.

  > **Note:** The management communication between the Storwize V7000 and the two file modules runs through these 1 GbE ports on the Storwize V7000 system. They must be configured to be in the same subnet as the management ports on the file modules. The file modules can optionally use 10 GbE ports for management, but the 1 GbE ports are the default and must be used for the initial configuration.
  >
  > The Storwize V7000 Unified system uses only IPv4 (at the time of writing).

- Both 1 GbE ports can be used and configured simultaneously for iSCSI access from iSCSI hosts.
- The 10 GbE ports in V7000 models 312 and 324 can be configured for iSCSI access only.
- Configure every node canister in the V7000 system with a service IP address that is accessible in your network environment, in case the Storwize V7000 cluster management IP address cannot be reached or the cluster is no longer working. Then, access through the Service Assistant interface to the individual node canisters might be required to debug and resolve the situation. The default service IP addresses of node canisters are `192.168.70.121` and `192.168.70.122`, with a subnet mask of `255.255.255.0` and a default gateway of `192.168.70.1`. These can be used, if appropriate, or changed to other valid addresses on the network.

> **Note:** In contrast to the stand-alone Storwize V7000 system, the service IP addresses of the Storwize V7000 node canisters can no longer be changed as part of the USB key initialization process of a Storwize V7000 Unified system. The init tool screens for Storwize V7000 Unified allow you to set only the service IP addresses of the two file modules during the initial installation. Therefore, when GUI access is available, we suggest that you change the service IP addresses of the V7000 node canisters to the ones that you use.

In addition, there are two file modules that provide their own, different interfaces, which creates more considerations:

- All interfaces are bonded. That is, they use a virtual interface that is bonded on two physical ports.

  Therefore, the two ports that belong to one bonded interface cannot be attached to separate networks. This configuration is also described in the installation documentation.

  The file modules use two bonded 1 GbE ports for a direct cluster connection between them.
- Similar to the Storwize V7000, the two remaining 1 GbE ports can be used for both management access and data traffic (difference: the V7000 supports iSCSI traffic only) through TCP/IP.

The default bonds are configured:

  - `ethX0` for data traffic on the 1 GbE ports
  - `mgmt0` for management traffic on the same 1 GbE ports
  - `ethX1` for the 10 GbE ports of the file modules.

- Default management access is through 1 GbE ports, and this is required for initial installation:
  - Management through 10 GbE ports is optional and can be configured later.
  - Ensure that communication with the Storwize V7000 management IP through 1 GbE continues to work, because management of the Storwize V7000 storage system is always through 1 GbE.
- VLANs are supported and can be configured for both 1 GbE and 10 GbE after initial installation and Easy Setup steps are complete. There is no VLAN support during initial installation and Easy Setup.

Note: Currently, there is an open problem when you have both a 1 GbE data network and 10 GbE data network on the same subnet. The Storwize V7000 Unified then responds only on the 1 GbE interfaces. This issue will be fixed in one of the next maintenance releases.

SAN considerations

The storage area network (SAN) considerations of the Storwize V7000 Unified system are similar to the ones of a stand-alone Storwize V7000 because all of the FC access-related functions are the same. The only difference is that the Storwize V7000 Unified system has only four FC ports available on its V7000 node canisters for SAN connectivity, because the other four ports are dedicated for, and directly connected to, the two file modules.

We suggest that you have a redundant SAN configuration with two independent fabrics to provide redundancy for the V7000 connections, FC host port connections, and connections for external SAN virtualized storage systems. Distribute all connections evenly between both fabrics to provide redundancy in case a fabric, host, or storage adapter goes offline.

Zoning considerations

For the Fibre Channel connectivity of the Storwize V7000 Unified, the same zoning considerations apply as for a stand-alone V7000. The one difference is that there are only four FC ports available (two ports per node canister: Port 3 and port 4).

We suggest that you create a node zone in every fabric with two of the four V7000 ports (one per node canister: Port 3 in one fabric, Port 4 in the other fabric) for a redundant communication path between the two node canisters (if there is a problem with the communication through the midplane within the V7000).

For the Storwize V7000 host attachment through Fibre Channel, create a host zone for each host in each fabric, assign the FC connections of this host in a redundant fashion, and zone the V7000 node canisters to ensure redundancy.

If there is external SAN-attached storage to be virtualized, create a storage zone in each fabric with half of the ports of the external storage system and one port per node canister in the same way.

File access protocols

All data subsets need to be accessed through only one file protocol:

- NFS exports: The default owner of an export is the *root* user. You need root user access to the NFS client for initial access to the export and to create the directory structures and access permissions for other users as wanted.
- CIFS shares: It is mandatory to specify an owner when you create a new share to allow access to the share from the Active Directory (AD) domain. Otherwise, the default owner is the root user as with NFS, but this user typically does not exist on CIFS client side. The initial owner that is specified is the one used for initial access from CIFS client side to create the directory structures and access control lists (ACLs) for all users, as required. It is important, for example, that the *traverse folder* has the right to be able to access directories below the original home directory of the share. When the directory structure for other users and appropriate ACLs are defined, necessary shares can be defined in the Storwize V7000 Unified afterward. If their access works as needed, the initial owner can be deleted if wanted or the initial owner's access right can be minimized as needed.

Multiple simultaneous exports of same subset of data via different protocols

This is fully supported by the Storwize V7000 Unified Disk System. Most likely, this multiprotocol export is using NFS and CIFS. The difficulty is to ensure that the access rights and ACLs set are compatible from both client sides.

Notes:

First-time creation of a CIFS share: It is mandatory to specify an owner for first-time access when you create a CIFS share. Only the root user has access to that share if no other permissions are set from the client side already (which is not the case if it is truly first-time access). An owner of a share can be set or changed on the Storwize V7000 Unified system only when there is no data that is stored in the share yet.

For managing CIFS ACLs: Authorization setting to *Bypass traversal check* is not supported by Storwize V7000 Unified. Therefore, *traverse folder* rights must be explicitly granted to all users (or to *Everyone*) who needs access to directory structures below the level of the current directory. These users do not see any contents when they traverse directories that allow *only traverse-folder* rights.

Simultaneous export of the same data through both CIFS and NFS: Changing ACLs from the NFS side is likely to destroy the CIFS ACLs, because NFS uses the much simpler Portable Operating System Interface (POSIX) bits for users, groups, and others to manage access rights. Because CIFS provides much more sophisticated ACL management options, it is recommended to manage ACLs for the common share on the CIFS client side.

Note: Volumes for file access (equivalent to NSDs, as seen by GPFS) are not explicitly visible in the GUI and cannot be modified in the standard GUI windows, such as volumes or pools. Only volumes for block I/O can be created and modified in these GUI windows. This is intentional, because the *file volumes* are created during file system creation and are always associated with a file system. Therefore, they must not be manipulated separately and are hidden from the standard GUI panels that involve volumes. They are managed on the file system layer instead (typical CLI commands: `chfs` and `mkdisk`).

The *file volumes* can be displayed explicitly by listing the details about the appropriate file system in the GUI or through CLI.

Authentication requirements

Decide which implementation of the authentication service to use. If only external (as suggested) or a mixed external and internal, user ID mapping is used.

If there is an existing authentication infrastructure, it often includes only one version (for example, Active Directory or Lightweight Directory Access Protocol [LDAP]). That determines the decision for the implementation of Storwize V7000 Unified system.

> **Important:** The Storwize V7000 Unified supports only *one* authentication method at a time. Changing it later is *not recommended*. Therefore, it is important to carefully decide and select the method at the start.
>
> Ensure that long-term goals are taken into account. Also, the potential use of certain functions, such asynchronous replication, and the planned future enhancements for WAN caching (see Statement of Direction that is published at the announcement time of Storwize V7000 Unified), require an *external-only* ID mapping. Therefore, if there are chances that this might be needed in the future, ensure that an external only user ID mapping is used from the start.

The details of each of the following authentication options are described in Chapter 4, “Access control for file serving clients” on page 37.

A summary of the available options follows. Select *one* method from these options.

Active Directory (includes Kerberos)

The following are the Active Directory options and considerations:

- Standard (provides ID mapping for Windows only)

  In this case, the Storwize V7000 Unified system uses an internal ID mapping, both for UNIX type users (using a User ID, Group ID scheme) and for mapping Microsoft Windows subject identifiers (SIDs) to local user identifiers (UIDs) and group identifiers (GIDs).
- With Services for UNIX (RFC2307 schema)
  - Available on domain controllers that are running Windows 2003 SP2 R2 and later.
  - This option provides full external user ID mapping for both Windows and UNIX users inside the Active Directory server.
- With Services for UNIX (SFU schema)
  - Available on domain controllers that are running Windows 2000 and 2003
  - Similar option for older domain controllers to provide full external user ID mapping for both Windows and UNIX users inside the Active Directory server.
- With Network Information Service (NIS) (netgroup support only)

  Adding netgroup support through NIS. Netgroup is an option to group hosts and manage them as one group.
- With NIS (netgroup support and user ID mapping)

  Using NIS for the UNIX user ID mapping.

When using Active Directory, we suggest that you provide full external user ID mapping on the Active Directory server.

Lightweight Directory Access Protocol

The following are the Lightweight Directory Access Protocol (LDAP) options:

- LDAP
- Secure LDAP (with Kerberos)
- Secure LDAP (with Kerberos) and Secure Sockets Layer/Transport Layer Security (SSL/TLS) encrypted communication (available through CLI only)

Secure LDAP (with Kerberos) and SSL/TLS encrypted communication provides the most security and is therefore the recommended option in an LDAP environment.

Samba primary domain controller (NT4 mode)

The Samba primary domain controller is a legacy implementation that not many environments require. However, it is still supported by the Storwize V7000 Unified in the following modes:

- Standalone Samba PDC
- Samba PDC with NIS extension (netgroup support only)
- Samba PDC with NIS extension (netgroup support and ID mapping)

This implementation does not provide user-based authentication. Instead, it provides client-based (host or IP address) authentication because all users who connect from the same NFS client machine get access.

Local authentication

Local authentication was added in Version 4.1. This gives you the ability to create an open LDAP server on the node and replicate the configuration between the nodes by using the LDAP MirrorMode.

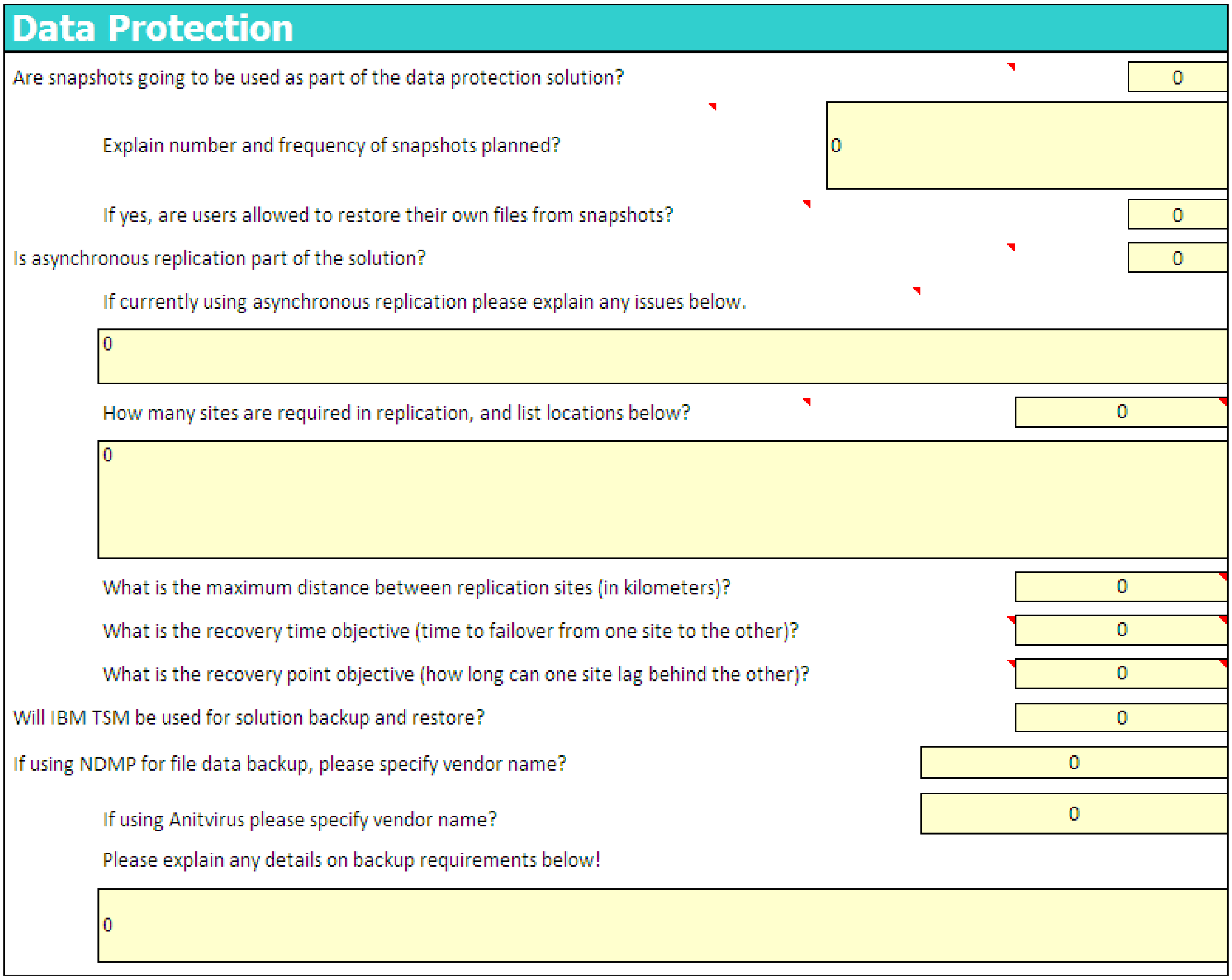
For more information, see Chapter 11, “Implementation” on page 141.

10.1.5 Data protection

We use this section to gather planning information about the following data protection methods:

- Snapshots
- Asynchronous replication
- Back up and restore with IBM Tivoli Storage Manager or NDMP
- Antivirus

Figure 10-7 on page 111 shows data protection section of questionnaire.



Data Protection

Are snapshots going to be used as part of the data protection solution? 0

Explain number and frequency of snapshots planned? 0

If yes, are users allowed to restore their own files from snapshots? 0

Is asynchronous replication part of the solution? 0

If currently using asynchronous replication please explain any issues below.

0

How many sites are required in replication, and list locations below? 0

0

What is the maximum distance between replication sites (in kilometers)? 0

What is the recovery time objective (time to failover from one site to the other)? 0

What is the recovery point objective (how long can one site lag behind the other)? 0

Will IBM TSM be used for solution backup and restore? 0

If using NDMP for file data backup, please specify vendor name? 0

If using Anitvirus please specify vendor name? 0

Please explain any details on backup requirements below!

0

Figure 10-7 Data protection

Snapshots

Snapshots are, by design, space-efficient, because they work with pointers and redirect on write for updates to existing data blocks, which are part of a snapshot. Therefore, the rate of changes to the data, as part of a snapshot, and the frequency of creating snapshots and the retention period for the existing snapshots determine the capacity that is required for snapshots. If a snapshot is used only for backup or asynchronous replication and deleted afterward, there is usually no need to include significant extra capacity for snapshots in your plan. But if several previous versions of files are kept in snapshots in case of operational failures (for easy file restoration), take this into account, along with the expected change rate for this data.

Asynchronous replication

With asynchronous replication, consider the following factors:

- Requires a second Storwize V7000 Unified system, because it is not supported within a single Storwize V7000 Unified system
- Requires full external user ID mapping, for example, Active Directory with Services for UNIX (SFU)
- Operates at the file system level, with a 1:1 relation between the local and remote file system

- Sequence of operations:
  - First, a snapshot is taken on the source file system
  - The source file system is scanned to identify the files and directories that were changed (created, modified, or deleted) since the last asynchronous replication
  - Changes are identified and replicated to the target file system
  - When the replication is complete, a snapshot is taken on the target file system and the source snapshot is deleted
- Timing:

  > **Note:** To accurately size replication for the Storwize V7000 Unified ensure you follow the instructions in the *V7000 Unified Asynchronous Replication Performance Reference Guide* PDF available at:
  >
  > `http://www.ibm.com/support/docview.wss?uid=ssg1S7004539&aid=1`

  Frequency is determined by the interval that is defined in a scheduled task for asynchronous replication. The minimum interval that can be defined is 1 minute. Duration of one run is determined by the time required to take a snapshot, scan for changed files, and transfer the changes to the remote site, given the network bandwidth that is available. Also factored into this is the time that is required to take a snapshot at the remote site and the time to delete the source snapshot. In case the first asynchronous replication cycle is not finished before the next scheduled asynchronous replication is triggered, the subsequent replication does not start. This is to enable the first one to complete successfully (an error is logged). After its completion, a new asynchronous replication will start at the next scheduled replication cycle.

> **Note:** File set and share definitions, quota, and snapshot rules that are defined are not contained within the replicated data. This information is kept only at the source and is not transferred to the replication target.
>
> These definitions must be applied to the target file system as needed for a failover scenario (which might be different from the scenario at the source).
>
> For testing disaster recovery, the target file system can be mounted as read-only to clients on the target side. If writes are allowed and occur on the target file system, there is a potential data integrity issue, because the source file system does not reflect these updates (changes are tracked only at the source).
>
> If write access is needed on the target side (as part of a disaster recovery test, for example), file clones must be created for the affected data files within the target file systems. This cannot be done by using the snapshot on the target side (which can be accessed only as read-only), because it is not possible to create file clones from a snapshot.

Back up and restore with IBM Tivoli Storage Manager or NDMP

With Tivoli Storage Manager and NDMP, remember the following factors:

- Only one method is supported. Therefore, you must choose Tivoli Storage Manager or NDMP.
- If NDMP is selected, HSM is not supported.
- NDMP backup is supported by Netbackup, Commvault Simpana, EMC Networker, and Tivoli Storage Manager as Data Management Application (DMA).

- The NDMP data service runs on the Storwize V7000 Unified file modules.
- Different NDMP topologies are supported: Two-way or three-way (local is not supported):
  - Two-way: DMA and tape service running on the same system.
  - Three-way: DMA and tape service running on different systems, whereby the metadata information is sent to the DMA and the data containers sent to the tape service.
- If Tivoli Storage Manager is selected as the backup method, the preinstalled Tivoli Storage Manager client on the file modules is used.
- Selection of Tivoli Storage Manager enables the option to use HSM also.

Antivirus

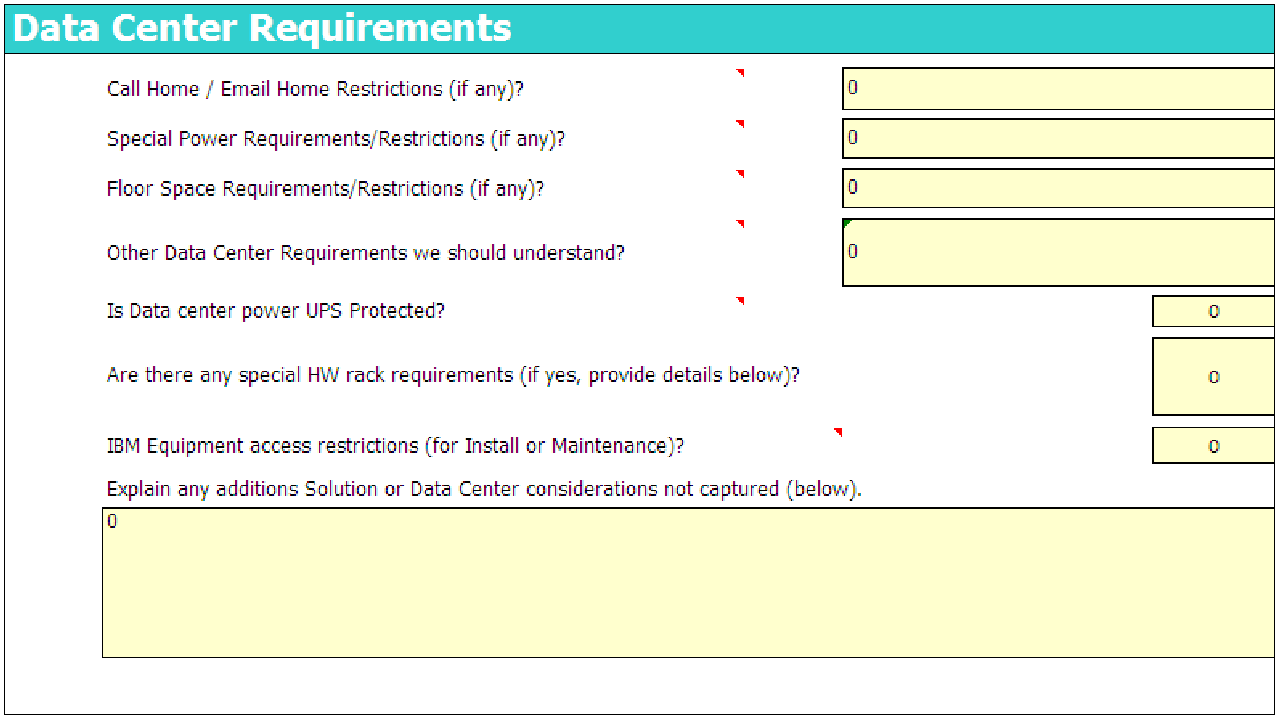
With antivirus software, consider these factors:

- Supported antivirus product families or access schemes are Symantec and McAfee (through Internet Content Adaptation Protocol, or ICAP, on port 1344)
- Requires external scan engines
- Configurable options: Scan on file open, scan on file close after write, scheduled batch scans (also known as *bulk scan*)

> **Note:** Bulk scans do not rescan HSM-migrated files. No file recall is therefore required.

10.1.6 Data center requirements

This section can be used to provide details about special data center requirements, such as power, floor space, and rack requirements. See Figure 10-8 for details.



Data Center Requirements

Call Home / Email Home Restrictions (if any)?	0
Special Power Requirements/Restrictions (if any)?	0
Floor Space Requirements/Restrictions (if any)?	0
Other Data Center Requirements we should understand?	0
Is Data center power UPS Protected?	0
Are there any special HW rack requirements (if yes, provide details below)?	0
IBM Equipment access restrictions (for Install or Maintenance)?	0

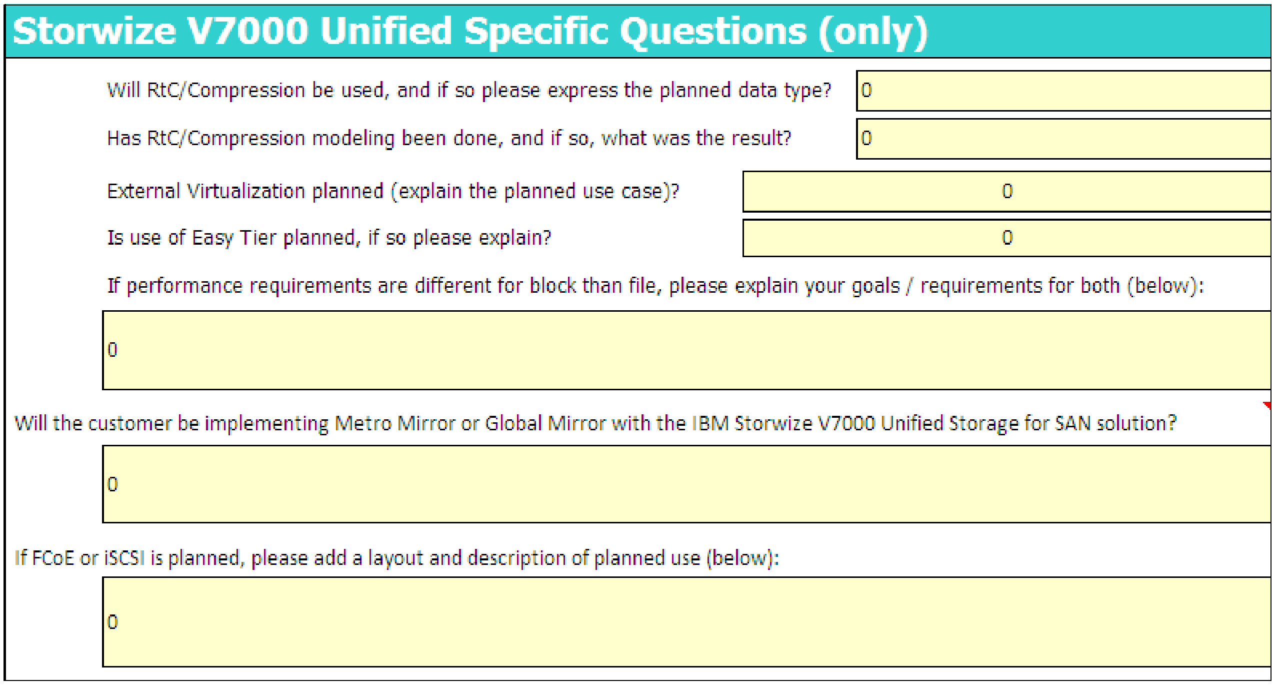
Explain any additions Solution or Data Center considerations not captured (below).

0

Figure 10-8 Data center requirements

10.1.7 Storwize V7000 Unified specific questions

Figure 10-9 on page 114 shows the fields in this section that are specific to the Storwize V7000 Unified system. Use this section to provide planning details relevant to real-time compression, the use of IBM Easy Tier, IBM Metro Mirror, or BM Global Mirror, and the use of FCoE or iSCSI.



Storwize V7000 Unified Specific Questions (only)

Will RtC/Compression be used, and if so please express the planned data type? 0

Has RtC/Compression modeling been done, and if so, what was the result? 0

External Virtualization planned (explain the planned use case)? 0

Is use of Easy Tier planned, if so please explain? 0

If performance requirements are different for block than file, please explain your goals / requirements for both (below):

0

Will the customer be implementing Metro Mirror or Global Mirror with the IBM Storwize V7000 Unified Storage for SAN solution?

0

If FCoE or iSCSI is planned, please add a layout and description of planned use (below):

0

Figure 10-9 Storwize V7000 Unified specific questions

It is important to understand the requirements for using compression before enabling it:

- **Hardware requirements:** Compression requires dedicated hardware resources within the node, which is assigned and de-assigned when compression is enabled and disabled. When you create the first compressed volume in an I/O group, hardware resources are assigned because there are fewer cores available for the fast path I/O code. Therefore, do not create a compressed volume or file system if the CPU use is consistently above 25%.
- **Data type:** The best candidates for data compression are data types that are not compressed by nature. Do not use compression in volumes or file systems that contain data that is compressed by nature. Selecting such data to becompressed provides little or no savings, yet uses CPUresources by generating extra I/Os. Avoid compressing data with less than a 25% compression ratio. Data with at least a 45% compression ratio is the best candidate for compression.
- **Compression ratio estimation**: To estimate the compression ratio of a volume, use the Comprestimator tool. This is a command-line, host-based utility that scans the volume and returns the compression ratio that can be achieved by using compression.

Comprestimator: Comprestimator can be used only on devices that are mapped to hosts as block devices. Therefore, it cannot be used in file servers and file systems in the V7000 Unified. For more information about estimating the compression ratio of files, see Chapter 16, “IBM Real-time Compression in the Storwize V7000 Unified system” on page 289.

- **Placement policy for mixing compressible and noncompressible data in a file system:** When a file system contains a mixture of compressible and noncompressible data, it is possible to create a file system with two file system pools. One for the compressible files, which are configured with compression enabled, and the other for the noncompressible files. The *placement policy* option is configured to place the compressible files in the compressed file system pool. The policy is based on a list of file extensions of compressed file types that are defined as an exclude list. This extensions list is edited manually when you are configuring the policy. Using the placement policy avoids spending system resources on noncompressible data. The policy is affecting only files that are created after the management policy has changed.

  For more information about the placement policy and file types, see Chapter 16, "IBM Real-time Compression in the Storwize V7000 Unified system" on page 289.

- **License:** Compression has limited access in the current version. Therefore, enter a code when you configure a new compressed file system. To get the code to enable compression, contact IBM at NEWDISK@us.ibm.com, and a specialist will contact you to provide the code.
- **Number of compressed volumes:** The number of compressed volumes is limited to 200 per I/O group. This number includes compressed file systems. When a file system is created, three compressed volumes are created for it, and they are counted in the 200 compressed volumes limitation. For example, if you create 195 compressed volumes and then create one compressed file system, you have 198 compressed volumes in use, and you cannot create another compressed file system (only two compressed volumes are left, and the process requires three). Therefore, it is important to understand this limitation and plan the number of compressed volumes and file systems in the entire system before using compression.
- **Plan the number of pools you need to use compression:** When using compression, considerations about the MDisks to create are different. There are several items to consider first:
  - Do not store compressed and noncompressed volumes in the same MDisk group. In mixed volume types, the compressed and noncompressed volumes share cache and it might increase the response time. Therefore, it is not recommended to create compressed volumes in an MDisk group that contains noncompressed volumes.
  - Create different pools for data and metadata. To use compression, separate the data and metadata, because the metadata should not be compressed. Therefore, create at least two storage pools. For more information about configuration refer to Chapter 11, "Implementation" on page 141.
- **Balanced system:** When you are creating the first compressed volume, CPU and memory resources are allocated for compression. When the system contains more than one I/O group, it is recommended to create a balanced system. If you create a low number of compressed volumes, it is recommended to create them all in one I/O group. For larger numbers of compressed volumes, the general recommendation (in systems with more than one I/O group) is to distribute compressed volumes across I/O groups. For example, a clustered pair of Storwize V7000 control enclosures requires 100 compressed volumes. It is better to configure 50 volumes per I/O group, instead of 100 compressed volumes in one I/O group. Also ensure that the preferred nodes are evenly distributed.
- **IBM Easy Tier:** Real-time Compression does not support Easy Tier, which is a performance function that automatically migrates or moves extents of avolume to, or from, one MDisk storage tier to another MDisk storage tier. Easy Tier monitors the host I/O activity and latency on the extents of all volumes that have the Easy Tier function turned on in a multi-tiered storage pool, over a 24-hour period. Compressed volumes have a unique MDisk write pattern, which would trigger unnecessary data migration. For this reason, Easy Tier is *disabled* for compressed volumes and you cannot enable it. You can

however create a compressed volume in an Easy Tier storage pool but automatic data placement is not active.

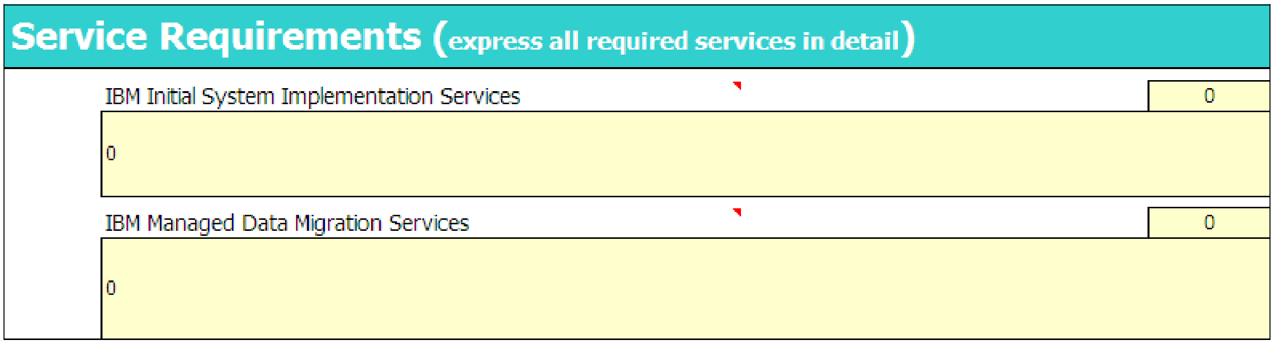
For more information about Real-time Compression, see Chapter 16, “IBM Real-time Compression in the Storwize V7000 Unified system” on page 289.

For more information about compression technology, see *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859:

http://www.redbooks.ibm.com/redpieces/abstracts/redp4859.html

10.1.8 Service requirements

The final section of questionnaire can be used to document service requirement details. As shown in Figure 10-10, this section is to list required system implementation and data migration services.



Service Requirements (express all required services in detail)

IBM Initial System Implementation Services	0
0	
IBM Managed Data Migration Services	0
0	

Figure 10-10 Service requirements

If data migration is needed, there are different ways available to achieve it. In general, if you need IBM assistance with the migration of data to the Storwize V7000 Unified system, contact IBM regarding the Data Migration Services offerings that match your requirements:

http://www.ibm.com/services/us/en/it-services/data-migration-services.html

For SAN-attached block storage, a migration wizard is built into the GUI, which helps in migrating existing data into the Storwize V7000 managed storage. This wizard is described in more detail in *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

For data migration from existing file storage and network-attached storage to the Storwize V7000 Unified system, the migration must happen on a file level to keep all files aware and the software up-to-date with the changes and to maintain the ACLs of files and directories. Therefore, the block level migration options that are built into the V7000 cannot be used for that purpose. We recommend contacting your IBM representative to decide on the best migration policy in this case.

10.2 Planning steps sequence

In the following sections, we list and explain the correct sequence of planning steps.

10.2.1 Perform the physical hardware planning

It is important to take into account the physical components that are needed to ensure that everything that is required is ordered ahead of time. Consider the following components:

- Storwize V7000 Unified system, cables, and connectors
- Storage area network (SAN) and network switches that are required, cables, and connectors
- File access clients that are required, and I/O adapters
- Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) hosts that are required, and I/O adapters
- Power and cooling requirements for all hardware involved
- Plan for the lab floor space and rack layout for all the hardware identified

10.2.2 Define the environment and services needed

Plan for the environment and the services that are required:

- IP addresses needed for management and service of Storwize V7000 and both file modules, public IP addresses to serve file I/O, and client IP addresses
- Authentication service: Servers that are needed according to the selected method and netgroup or ID mapping support
- Time synchronization: Network Time Protocol (NTP) servers
- Domain Name System (DNS): DNS servers
- Copy services and async replication, including required connectivity and remote target systems
- Back up servers according to the method chosen and storage
- Tivoli Storage Manager hierarchical storage management (HSM) servers and storage, if required
- Antivirus scan engines

10.2.3 Plan for system implementation

The closer you get to the actual implementation, the more important it is that you consider the following requirements:

- Define the local and remote (if needed) SAN zoning requirements
- Define the network requirements for management and data access
- Define the network interfaces of V7000 and file modules, including subnets and VLANs
- Define the logical configuration of the system (both file and block access)
- Define the pools and LUN layout for block access
- Define the pools, exports and shares, file systems, file sets, and directory structures for file access

- Define users that are required for management and monitoring roles of the Storwize V7000 Unified system and for file-based access that requires authentication, and configure them within the authentication service and directory server
- Plan for the user ID mapping method (external or mixed)
- Define authorizations that are required for every file access user within the file system/file set/directory structures

10.2.4 Plan for data migration

Based on the features of the Storwize V7000 Unified, there are two different options for data migration:

- Migrating data from existing SAN-attached storage to the Storwize V7000 Unified by using the built-in migration wizard and image mode volumes
- Migrate data from existing NAS systems to the Storwize V7000 Unified by using file-based migration options

10.3 Support, limitations, and tools

Always verify your environment with the latest support information for the Storwize V7000 Unified system, and be sure to use the latest versions of modeling tools for capacity and performance (Capacity Magic and Disk Magic).

Determine lists of hosts and platforms to be attached and verify interoperability support and any restrictions for the following components:

- FC attachments
- Network attachments and file access
- iSCSI attachments

Determine your requirements and verify that they are within the capabilities and limitations of the system. The Technical Delivery Assessment (TDA) checklist provides more useful considerations. Use the modeling tools that are available (with help from your IBM Support or IBM Business Partner Support, if needed) to determine the system configuration that can fulfill your capacity and performance requirements.

Here are several useful links for these purposes:

- Support portal for Storwize V7000 Unified:

  `http://www.ibm.com/storage/support/storwize/v7000/unified`

- Interoperability support pages for Storwize V7000 Unified:

  `http://www.ibm.com/support/docview.wss?uid=ssg1S1004228`

- Configuration Limits and Restrictions:

  `http://www.ibm.com/support/docview.wss?uid=ssg1S1004227`

- The general Limitations section in the information center is useful as preparation for the planning and implementation decisions:

  `http://pic.dhe.ibm.com/infocenter/storwize/unified_ic/index.jsp?topic=%2Fcom.ibm.storwize.v7000.unified.142.doc%2Fadm_limitations.html`

- Verify the planned setup and environment by using the Pre-Sales Technical Delivery Assessment (TDA) checklist. You can find the checklists for TDA for both Pre-Sales and Pre-Installation on this web page (this is an IBM and IBM Business Partner internal link only, so contact your IBM or Business Partner Support for help if you do not have access):

  http://w3.ibm.com/support/assure/assur30i.nsf/WebIndex/SA986

- Determine your capacity requirements, including asynchronous replication for files, snapshots, IBM FlashCopy, Remote Copy for block I/O, and GPFS internal replication requirements, and verify the system configuration by using Capacity Magic.
- Determine all of the workload parameters, such as the number of I/Os per second (IOPS), throughput in MBps, I/O transfer sizes for both file I/O and block I/O workloads, number of clients (for file access), number of hosts (for block I/O access, both iSCSI and FC), copy services requirements for both block I/O and file I/O. Also, verify the system configuration by using Disk Magic modeling.

The accuracy of the input data determines the quality of the output regarding the system configuration required.

It is also important to note that there are influencing factors outside of the Storwize V7000 Unified system that can lead to a different performance experience after implementation, like network setup and I/O capabilities of the clients used.

The IBM modeling tools for capacity (Capacity Magic) and performance (Disk Magic) can be found at this website (This is an IBM internal link only. Contact your IBM or IBM Business Partner support for help with the modeling if you do not have access. IBM Business Partners have access to these tools through IBM PartnerWorld®):

http://w3.ibm.com/sales/support/ShowDoc.wss?docid=SSPQ048068H83479I86

10.4 Storwize V7000 Unified advanced features and functions

In the following sections, we describe planning considerations for the Storwize V7000 Unified advanced functions.

10.4.1 Licensing for advanced functions

With licensing for advanced functions, consider the following factors:

- Almost all advanced functions are included in the two base licenses that are required for Storwize V7000 Unified. The following licenses are required: 5639-VM1 (V7000 Base, one license per V7000 enclosure required) and 5639-VF1 (File Module Base, two licenses required)
- Exception: External virtualization requires a 5639-EV1 license by enclosure
- Exception: Remote Copy Services for block I/O access requires a 5639-RM1 license by enclosure
- Exception: Real-time Compression requires license by enclosure

10.4.2 External virtualization of SAN-attached back-end storage

With external virtualization of SAN-attached back-end storage, consider these factors:

- Provides scalability beyond the Storwize V7000 limit for internal storage, which is 360 TB currently
- Maximum capacity that can be addressed is determined by Storwize V7000 extent sizes that are defined at the storage pool layer, with a maximum of 2^22 extents managed
- External storage is licensed by storage enclosure
- Same support matrix as Storwize V7000 and the IBM SAN Volume Controller

10.4.3 Remote Copy Services (for block I/O access only)

Remote Copy Services (not including asynchronous file-based replication) are the same as available with a Storwize V7000 with the same minimum code release of V6.4. They are not applicable for Storwize V7000 volumes that are used for file systems.

An important consideration in the Storwize V7000 Unified is the reduced FC fabric connections because of the required direct connections between the file modules and the canisters.

See more information in *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

With Remote Copy Services, consider the following factors:

- Remote copy partnerships are supported by other IBM SAN Volume Controller, IBM Storwize V7000 Storage System, or Storwize V7000 Unified Disk Systems (with the StorwizeV7000 in the back-end of a Storwize V7000 Unified system)
- Fibre Channel Protocol support only
- Licensed by enclosure
- SAN and SAN Volume Controller or Storwize V7000 Copy Services distance rules apply (maximum 80 ms per round trip)
- Needs partnerships that are defined to remote system (SAN Volume Controller and Storwize of V7000 and Storwize V7000 Unified systems)
- A maximum of three partnerships at a time are supported, and that means a maximum of four systems can be in one copy configuration. Not all topologies that are possible are supported (for example, all four systems configured in a string: A-B-C-D)
- Within the partnerships that are defined between systems, the copy services relationships are established at a volume level as a 1:1 relationship between volumes. Each volume can be in only one copy services relationship at a time
- Consistency groups are supported

10.4.4 FlashCopy (block volumes only)

The IBM FlashCopy implementation (not including snapshots as used for file sets and file systems) is the same as available with a stand-alone Storwize V7000 system with the same minimum code release of V6.3. FlashCopy operations are not applicable for Storwize V7000 volumes that are used for file systems.

See more about this topic in *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

With FlashCopy, consider the following factors:

- Need to take volumes and capacity that is needed for FlashCopy copies into account
- All SAN Volume Controller and V7000 FlashCopy options are fully supported on standard Storwize V7000 volumes that are not used in file systems
- Consistency groups are supported

10.4.5 General GPFS recommendation

Every file operation requires access to the associated metadata. Therefore, for GPFS, it is best to place the metadata on the fastest drive type that is available. This can be achieved by creating Network Shared Disks (NSDs) (Storwize V7000 volumes that are associated with a file system) based on the fastest drive type and add them to the *system* file system pool with the specific use type of *metadataonly.* In addition, the use type of the other, slower NSDs must be set to *dataonly*. This ensures that only the fastest disks host the metadata of the file system.

10.4.6 GPFS internal synchronous replication (NSD failure groups)

As described in Chapter 7, “IBM General Parallel File System” on page 65, this provides an extra copy of the selected data type (data, metadata, or both) in a different storage pool. Therefore, the pool configuration and additional capacity that is required needs to be taken into account:

- Synchronous replication operates within a file system and provides duplication of the selected data type
- This section describes the needs for additional space. When creating a new file system, define two storage pools: One for the metadata and at least one for the data. You can also add multiple pools for the data in the same file system
- Ideally, file system pools that use this functionality are replicated between storage pools in independent failure boundaries:
  - This independence defines the level of protection, for example, against storage subsystem failure
  - This independence is compromised here, because there is only one Storwize V7000 storage system that manages the back-end storage
- If *metadata*, *data,* or *both are* to be replicated between the file system pools

  Defines level of protection
- Capacity that is used must be included in planning for total file system capacity

  Approximately 5 - 10% of file system capacity is used for metadata, so you need to adjust overall capacity accordingly

10.4.7 Manage write-caching options in the V7000 Unified and on client side

There are different options to enable and disable caching within the layers inside the Storwize V7000 Unified and also outside, for example, on the client side.

On *NFS clients*, the options that are specified with the **`mount`** command determine whether client-side caching is allowed. Therefore, this can be changed at the level of each individual export. The Storwize V7000 Unified has no control of which option each NFS client is using.

Mounting an export with the **sync** parameter disables the client-side caching and assures the data is sent to the Storwize V7000 Unified after each update immediately.

For *CIFS clients*, the Storwize V7000 Unified supports opportunistic locking (oplocks), which enables client-side caching for the CIFS clients. That means by default, the client-side caching is granted by the Storwize V7000 Unified to every client that requests opportunistic locking. This can be changed for every individual export and share by using a the **chexport** command (see shaded information box that follows).

Inside the Storwize V7000 Unified system, there can be write-caching on the file modules that are managed by the NFS and CIFS server layer. This happens by default for all open files for CIFS access. For NFS access, the default is already set to `syncio=yes`. As soon as there is a **sync** command or a file gets closed, the updates are written to the NSDs (volumes in the V7000 pools) and are stored in the mirrored write cache of the V7000 before it is destaged to disk. This is safe because there is a second copy of the data. The caching in the file modules can be controlled for every export or share by using the **chexport** command options (see the following information box).

> **Important:** For applications with critical data, all non-mirrored caching options in GPFS need to be disabled:
>
> - Caching on the client side: controlled by the opportunistic locking (*oplocks*) option
> - Caching in file module cache: controlled *with the syncio* option
>
> This can be done by using the command-line interface (CLI) chexport command for the relevant shares, with the parameters `of oplocks=no` and `syncio=yes`.

10.4.8 Redundancy

The Storwize V7000 Unified is a highly available system, so it provides redundancy. In order to achieve high availability for the data access and operations, it is required that other parts of the environment provide redundancy as well. This is essential for services such as authentication, NTP, and DNS. There is a similar requirement for the networks to be redundant and for the power sources of all these components as well.

If there is no redundancy at even one of these levels, there is an exposure to not being able to continue operations when there is just a single failure in the environment. Having redundancy at all these levels ensures that at least a double failure is necessary to create an outage to the operations.

10.5 Miscellaneous configuration planning

In the sections that follow, we offer other planning considerations.

10.5.1 Set up local users to manage the Storwize V7000 Unified system

A number of predefined roles are available to define users with different accesses and to tailor access levels to your requirements:

- Security Administrator: Security administration rights plus user management
- Administrator: Full administration of the system except user management
- Export Administrator: Export- and share-related administration only
- System administrator: System-related administration only

- Storage Administrator: Storage-related administration only
- Snapshot Administrator: Snapshot-related administration only
- Backup Administrator: Backup and replication-related administration only
- Operator: Has only read access to the system

The default user of a Storwize V7000 Unified system is `admin`, which has the Security Administrator role and can manage other users.

You can create other Security Administrator users as required. Optionally, you can increase security by changing the default access, for example, by changing the password for the `admin` user.

10.5.2 Define call home and event notifications

Call Home requires a Simple Mail Transfer Protocol (SMTP) or email server address on the client LAN that can forward email to the default IBM service address. Details about the system, client and administrator contact, and phone numbers are needed to establish contact from IBM Support personnel in case of problems.

Event notification is supported by the following channels:

- **Email:** Requires SMTP or email server address to be specified. Multiple levels of notifications can be specified (for example, problems and informational events).
- **SNMP:** Defines IP address of server and which kinds of events, for example, status changes. Use triggers a notification.
- **Syslog server:** Defines IP address of server to receive information. Currently, only information about the V7000 is sent.

10.5.3 Storage pool layout

In general, there are useful default settings, referred to as *presets*, built into the system, which are used for the automated configuration steps as offered by Easy Setup. If there are standard performance requirements for either the file I/O or block I/O workload, these can conveniently be used, creating shared pools that contain volumes for both workloads. If there is a significant workload on either the file I/O or block I/O side, it is best to separate these by using separate pools. The separate pools enable fault isolation, performance predictability by using different physical disk drives in the back-end, and easier performance analysis.

Another criterion is the file protocol that is used to access the data: Data that is accessed by CIFS clients *must not* be on SAN-attached, external virtualized storage. Besides the reason for having separate failure boundaries on storage pool level, this is another reason to manage separate storage pools for external storage and assign them only to file systems that do not have CIFS shares defined.

Here is a checklist for the storage pool layout:

- Block, file, or mixed storage/workload required:
  - Block workload only: No dependencies to file workloads. Use GUI or CLI to configure
    - No special performance requirements. Then, use the presets and best practices built into Storwize V7000 by checking **Auto-configure storage** in Easy Setup.
    - Special consideration regarding performance and placement optimization: The CLI allows for specially tailored configurations.

- File workload only: Operating outside of the general positioning of Storwize V7000 Unified system, but there might be good reasons for that:
  - No special performance requirements: Use **Auto-configure storage** in Easy Setup and GUI to configure. This includes the presets and best practices built-in.
  - Special consideration regarding performance and placement optimization: The CLI allows for specially tailored configurations.
- If mixed block and file workload: Plan storage layout between the two, including a manual configuration of MDisks, pools, and volumes as needed. Configure storage layout, first for the file systems (generating file volumes), then block volumes. General recommendation: Although supported, do not use mixed pools. Use separate pools for file access and block I/O for better performance control and bottleneck analysis if required.

10.6 Physical hardware planning

Based on the results of the configuration-sizing steps, determine the list of hardware items to order. Table 10-1 tries to pre-empt the main questions to ensure that all areas have been considered. However, because of the complexity and multitude of options, this might not be complete in every case, and specific items might need to be added, as required.

Table 10-1 Checklist for required hardware

Hardware area	Components and details	Your items or numbers
Storwize V7000 Unified configuration	- Base configuration - V7000 expansions - Connectivity for the different interfaces and protocols for all locations and sites involved	
Network components and connectivity	- Ethernet switches - 1 GbE and 10 GbE and connectivity for all locations or sites involved	
SAN connectivity	- FC switches and directors - Ports and connectivity for all locations or sites involved	
Clients for file access	- Server HW - 1 GbE NIC, or 10 GbE CNA connectivity for all locations or sites involved	
Hosts for FC or iSCSI access	- Server HW - FC HBAs - 1 Gb and 10 Gb network cards - Connectivity for all locations or sites involved	
Services	- Servers for NTP, DNS, authentication, backup, HSM, antivirus: All, including connectivity	

Hardware area	Components and details	Your items or numbers
Miscellaneous	- SAN-attached storage and connectivity - Remote V7000 and Storwize V7000 Unified systems for remote copy or async replication	

Ensure that the required power and cooling are verified and provided as well.

10.6.1 Plan for space and layout

Physical space and layout planning considerations are as follows:

- An appropriate 19-inch rack with 6U - 24U of space is required, depending on the number of expansion enclosures to be installed. Each V7000 enclosure and each file module measures 2U in height. The minimum configuration is one V7000 control enclosure and two file modules with a total height of 6U.
- Redundant power outlets in the rack are required to connect the two power cords per V7000 enclosure and per file module to independent power sources. The number of power outlets that are required ranges from 6 to 24 per Storwize V7000 Unified system, depending on the number of V7000 expansion enclosures.
- Regarding the physical hardware placement, layout, and connectivity, there is detailed information in Chapter 11, “Implementation” on page 141.
- Two serial-attached SCSI (SAS) cables of the appropriate length are required per V7000 expansion enclosure. The individual lengths that are required are determined by the rack layout and placement that is chosen for the V7000 control and expansion enclosures.

> **Note:** There are two independent SAS chains to connect the V7000 control enclosure to the expansion enclosures. A symmetrical, balanced way to distribute the expansion enclosures on both SAS chains is recommended for performance and availability. The internal disk drives of the control enclosure belong to SAS Chain 2. Therefore, a maximum of four expansion enclosures can be connected to this chain. On SAS chain 1, a maximum of five expansion enclosures can be connected. To ensure a symmetrical, balanced distribution, the first expansion enclosure is connected to SAS Chain 1, the second one to SAS Chain 2, the third one to SAS Chain 1, and so on.

The “Storwize V7000 Unified physical installation planning” section of the Storwize V7000 Unified Information Center provides an overview about aspects of the physical implementation planning:

http://ibm.co/1epTwPC1

Hardware installation is described in the *Storwize V7000 Unified Quick Installation Guide*, GA32-1056:

http://bit.ly/1m44iRs

10.6.2 Planning for Storwize V7000 Unified environment

Here is a list of the minimum prerequisites to set up and use a Storwize V7000 Unified system. Several services are essential for operating and accessing the system, so they must be provided in a highly available fashion. These include NTP, authentication, and DNS:

Note: All of these services are required but do not necessarily require a dedicated server. For example, in case of Active Directory for authentication, the server can provide the NTP and DNS service also.

- Time servers for synchronization according to the Network Time Protocol (NTP)

  To guarantee common date and time across the environment, especially between the authentication server and the Storwize V7000 Unified system and Tivoli Storage Manager backups. Provide two servers for redundancy.

- Domain Name System servers
  - Required for DNS round robin for file access. Provide two servers for redundancy.
  - Required for Active Directory authentication (if this is used)
- Authentication servers

  Depends on the decision made in Step 10.5.1, “Set up local users to manage the Storwize V7000 Unified system” on page 122. Provide two servers for redundancy.

  - Select one of these services:
    Active Directory, LDAP, Samba PDC, local authentication, or NIS
  - Optional:
    - NIS server for Active Directory or Samba PDC
    - Kerberos or KDC server for LDAP
- Connectivity
  - Total of 4 x 1 GbE ports for file modules, min. 2 x 1 GbE ports for V7000
  - Optional: 10 GbE ports for file service (4x) or iSCSI attachment (minimum 2x)
- IP addresses for management and service access
- Minimum of six IP addresses are required for system management and service:
  - 1 x Storwize V7000 Unified cluster management IP
  - 1 x V7000 cluster management IP
  - 2 x service IP addresses for the two file modules (one each)
  - 2 x service IP addresses for the two V7000 node canisters (one each)
- Optional: 10 GbE for management of the Storwize V7000 Unified cluster and file modules
- Storwize V7000 requires 1 GbE for management

  Initial setup of Storwize V7000 Unified always requires 1 GbE for management and a dedicated port. VLANs are not currently supported for initial setup if you are using Easy Setup.

- VLANs are not supported for the initial setup, but can be configured later. VLAN ID of 1 must not be used.

> **Important:** The management and service IP addresses of a Storwize V7000 Unified must all be on the same subnet.
>
> All management and service IP addresses must be active, and the network must be configured correctly at initial installation. If there are connectivity problems between the file modules and the Storwize V7000 system, the initial installation fails.

- Optional, IP addresses for iSCSI if required:
  - A range of 1 - 4 IP addresses for 1 Gb iSCSI and 1 - 4 IP addresses for 10 Gb iSCSI
  - Recommended: Use minimum of 2 addresses per required interface 1 Gb or 10 Gb
- IP addresses for serving file I/O to clients:
  - Minimum of two public IP addresses to be used to have both file modules active in serving I/O:
    - For each interface used (1 GbE, 10 GbE) for file I/O
    - 1 GbE uses *ethX0* bond, 10 GbE uses *ethX1* bond
    - All network ports on the file modules are bonded by default. Both ports must be connected to the same subnet as documented.

      The difference: Network ports on Storwize V7000 node canisters are not bonded.
  - Minimum is one public IP address for each interface used, but then only one file module serves I/O. The second file module remains passive.
- Optional prerequisites, needed only if these features and functions are to be used:
  - Backup servers: Tivoli Storage Manager or NDMP, supported storage, licensed by Tivoli Storage Manager Server Value Units
  - Tivoli Storage Manager HSM servers, supported storage, licensed by Tivoli Storage Manager for Space Management
  - Antivirus scan engines

10.7 System implementation planning

If the Storwize V7000 Unified system is to be added to an existing environment, planning is required for only the physical location, power, and connectivity. This is because all external services, such as time synchronization through NTP servers, DNS, and authentication using the existing method that is set up in the environment, are already available. It is possible that an add-on, such as Services for UNIX (SFU), to an existing Active Directory infrastructure is required.

In the same sense, it is necessary to start building the infrastructure (physical location, power, cooling, connectivity) and the required external services first (NTP, DNS, authentication, and so on) before implementing the Storwize V7000 Unified Disk System. As Chapter 11, “Implementation” on page 141 and Table 10-2 on page 128 shows, the relevant, correct, and perhaps detailed information must be entered while using the Easy Setup wizard. Steps such as specifying the NTP servers are mandatory. During Easy Setup, the Storwize V7000 Unified checks whether there is an existing connection. If there is no response (from NTP or DNS servers, for example), the Easy Setup process fails.

If no authentication method is defined during Easy Setup, there is no data access from the file client side. This is because the required protocol daemons or services start within the Storwize V7000 Unified only after authentication is configured.

10.7.1 Configuration details and settings that are required for setup

There are many different options that are involved in implementing the Storwize V7000 Unified. Therefore, it is difficult to provide complete details for all options and combinations. There are also two classes of settings: *optional* and *mandatory*. In this section, we summarize all of the mandatory information and most of the optional areas, so that you can have it available as a comprehensive overview when you are implementing the system. However, there might be some optional areas that are not covered here in detail.

You can find a detailed description of the steps that are covered during the Easy Setup wizard and its related configuration information fields in Chapter 11, "Implementation" on page 141.

The following tables list the required system setup information chronologically, starting with the information that required for the init tool to prepare the USB key, followed by the information that is requested when you are using the Easy Setup wizard. See Table 10-2.

Table 10-2 Information for Storwize V7000 Unified setup

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Init Tool: V7000 IP, gateway, subnet mask	► V7000 management IP address ► Subnet mask ► Gateway IP address	**Mandatory:** ► IP for the V7000 storage cluster (not accessed directly in normal operations) ► Gateway and Subnet mask for entire Unified system **Note:** Service IP addresses for the V7000 node canisters cannot be set here; need access to GUI or CLI. Recommendation: set them first, after completing Easy Setup, and get GUI access the first time.	
Init Tool: Storwize V7000 Unified IP and file module details	► Storwize V7000 Unified cluster management IP address ► File Module 1 service IP address ► File Module 2 service IP address ► Internal network IP address range	**Mandatory:** ► IP for the Unified cluster: Needed and used for all management operations of the Storwize V7000 Unified system ► Individual service IP for direct access to a file module for troubleshooting ► Internal network for direct communication, troubleshooting	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Easy Setup: System attributes	- System name - NetBIOS name - Time zone - NTP server IP address - Alternate NTP server IP address	**Mandatory:** - Name of the Storwize V7000 Unified cluster - Name by which this cluster is seen on the network (SMB protocol) and known to an Active Directory domain - Continent and City: different scheme, not sorted by GMT+/-Xh, see http://ibm.co/1dNZXPJ - NTP server is required **Optional:** It is recommended to have a backup or alternate NTP server	
Easy Setup: Licenses	- External virtualization (by enclosures) - Remote Copy (by enclosures) - Real-time Compression by enclosures	**Mandatory:** - Number of storage enclosures of virtualized SAN storage behind V7000 - Number of storage enclosures that are used for block I/O based Remote Copy functions of V7000 (via Fibre Channel SAN) - Number of storage enclosures that are using IBM Real-time Compression	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Easy Setup Step 4: Support Notifications	Step 1: 1. Email server IP address 2. Company name 3. Customer email 4. Customer telephone number 5. Off-shift telephone number 6. IBM Support email address Step 2: Enable a proxy server to access the Internet	**Optional**, but strongly encouraged: - Storwize V7000 Unified cluster uses this email server to send email - Company name to appear in the email sent - Customer contact to receive email from Storwize V7000 Unified - Prime shift telephone number, which is called by IBM Support - Off-shift telephone number if prime shift number is not answered 24 hours - Leave at default: It is the default address in IBM for Call Home alerts Step 2: If access through proxy server is required, click **enable** and provide the proxy detail information	
Easy Setup Step 5: Domain Name System	- DNS domain name - DNS servers - DNS search domains	**Mandatory:** - Name of public network domain that is associated with Storwize V7000 Unified operations - IP addresses of your DNS servers. One is required; more are recommended for availability or redundancy **Optional:** Extra domain names to be searched in	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Easy Setup Step 6: Authentication	► Active Directory ► LDAP ► Samba PDC ► NIS (NFS only) ► Extended NIS ► Local authentication	**Mandatory** (if not specified here, use the GUI or CLI to configure authentication later): Radio buttons, choice between AD, LDAP, Samba PDC, and NIS **Optional:** Extended NIS can be chosen with Active Directory or Samba PDC	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Authentication Details for **Active Directory**	**Required only if choice is Active Directory:** ▸ Server ▸ User ID ▸ Password ▸ Enable Services for UNIX (SFU) ▸ Domain name, ranges, schema mode If Extended NIS in addition: ▸ Primary NIS domain ▸ Server map ▸ Enable user ID mapping ▸ Domain map ▸ User map ▸ User ID range ▸ Group ID range	**Required only if choice is Active Directory:** ▸ IP address of Active Directory server ▸ Administrative user ID ▸ Password for administrative user ID ▸ Check box: Select if support for UNIX is required ▸ Only if SFU selected: Name of domain SFU belongs to, lower to upper limit of the range for user and group IDs, SFU schema mode used (SFU or RFC2307) If extended NIS in addition: ▸ Name of the primary NIS domain ▸ NIS server to NIS domain map ▸ Check **Enable** if NIS user ID mapping to be used. This enables the next four topics: – Mapping of the Active Directory domain to the NIS domains – Define how to deal with user exceptions (DENY, AUTO, or DEFAULT) – Specify user ID range to be used with AUTO option – Specified group ID range to be used with AUTO option	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Authentication Details for **LDAP**	**(only required if choice is LDAP)** ► Specify one or more LDAP servers ► Search base for users and groups ► Bind distinguished name (DN) ► Bind password ► User suffix ► Group suffix ► Workgroup ► Security method ► Enable Kerberos ► Server name ► Realm	**(only required if choice is LDAP):** ► IP addresses of LDAP servers ► Search base as defined in LDAP server ► DN as defined in the LDAP servers ► Password for this DN ► User suffix as defined by the LDAP server ► Group suffix as defined by the LDAP server ► Domain name ► If SSL or TLS is used, a window to specify certificate appears. If the setting is *off*, the option for Kerberos appears (GUI). Use of SSL/TLS *and* Kerberos can be configured by using the CLI ► Check box: Check to enable Kerberos ► Only if Kerberos is enabled: Name of Kerberos server ► Only if Kerberos is enabled: Kerberos realm	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Authentication Details for **Samba PDC**	**Required only if choice is Samba PDC:** ► Server host ► Administrative user ID ► Administrative password ► Domain name ► NetBIOS name If extended NIS in addition: ► Primary NIS domain ► Server map ► Enable user ID mapping ► Domain map ► User map ► User ID range ► Group ID range	**Required only if choice is Samba PDC:** ► IP address of the NT4 PDC server ► User ID with admin authority to access the NT4 PDC server ► Password for this user ID ► NT4 domain name ► NT4 NetBIOS name if extended NIS in addition: ► Name of the primary NIS domain ► NIS server to NIS domain map ► Check **Enable** if NIS user ID mapping to be used. This enables the next four entry topics: – Mapping of the NT4 domain to the NIS domains – Define how to deal with user exceptions (DENY, AUTO, or DEFAULT) – Specify user ID range to be used with AUTO option – Specified group ID range to be used with AUTO option	
Authentication Details for **NIS (NFS only)**	**Required only if choice is NIS (NFS only), also known as Basic NIS:** ► Primary NIS domain ► Server Map	**Required only if choice is NIS (NFS only), also known as Basic NIS:** ► Name of primary NIS domain ► NIS server to NIS domain map	
Authentication Details for local authentication	► User or group name ► Password	**Optional:** Group ID. Otherwise, it is set automatically	

Step or purpose	**Entry field and information**	**Comment or explanation**	**Your data or selection**
Easy Setup Step 8: Configure storage	Automatically configure internal storage now	**Optional here in Easy Setup**, but **mandatory** to be configured to have V7000 provide storage capacity for the Storwize V7000 Unified system: ► Click **Yes** to configure internal storage as specified in the Configuration Summary ► If not, use GUI or CLI later to configure the internal storage that is provided by the V7000. If external, SAN-attached storage is used, use its appropriate GUI or CLI interfaces to configure	

Step or purpose	Entry field and information	Comment or explanation	Your data or selection
Easy Setup Step 9: Public Networks	► New network ► Subnet ► VLAN ID ► Default gateway ► IP address pool ► Additional gateways ► Interface	**Optional in Easy Setup** but **mandatory** to be configured to enable access to data for file clients (if not configured here, use GUI or CLI to configure later to enable file I/O): ► Select new network to get to next windows ► Subnet with network mask in CIDR syntax (that is, number of bits reserved for network mask), see Table 3: http://ibm.co/1fooKWO ► VLANs cannot be configured in Easy Setup. Later step: Enter VLAN number if VLANs are to be used (Note: VLAN 1 is not supported) ► IP address of default gateway for this subnet ► Pool of public IP addresses that are used to serve file I/O by using DNS round-robin; minimum is one, but need at least two to have both file modules serving I/O ► Optional, add if there are additional gateways ► Select **ethX1** for 10 GbE, **ethX0** for 1 GbE interface bond	

10.7.2 Configuration options for file access only

In Table 10-3, we show the file access-specific configuration options.

Table 10-3 File access-specific configuration options

Step/Purpose	Entry field or information	Comment or explanation	Your data or selection
Client systems	▸ 10 GbE attached ▸ 1 GbE attached	List systems, IP addresses, users	IP addresses, users
Users	▸ Local users in Storwize V7000 Unified for management and monitoring ▸ Users by client systems for data access	▸ Create local users and select their roles ▸ Create users on client systems, specify access rights, create same users within authentication servers if applicable	Local users, Storwize V7000 Unified Users for data access by client
File systems	Sizes and capacity needed	Include ILM and policies if required	
ILM	▸ Define tiered storage pools ▸ Define tiered file system pools ▸ Define ILM policies		
File sets	▸ Independent ▸ Dependent	Add more granularity: ▸ To define snapshot rules and quota ▸ To define quota	
Exports/Shares	By Protocol: ▸ CIFS ▸ NFS ▸ HTTPS ▸ FTP ▸ SCP Mixed exports if required, for example: CIFS and NFS	▸ Define owner at initial CIFS share creation ▸ Define extended ACLs for CIFS if required ▸ Define authorization or access rights from the client side	
Snapshots	▸ Creation rules ▸ Retention rules	▸ By independent file set ▸ By file system	
Quota	▸ By user ▸ By group ▸ By file set	For each entity required: Define soft limit and hard limit	

Step/Purpose	Entry field or information	Comment or explanation	Your data or selection
Backup	► Method ► Server IP addresses	► Choose Tivoli Storage Manager or NDMP ► Specify IP addresses of backup servers	
HSM	► Define external file system pool ► Define Tivoli Storage Manager HSM settings		
Async Replication	► Prepare remote partner systems ► Define file system to be replicated; create target file system on remote system; define replication settings ► Define schedule and task		
GPFS internal replication	► Define multiple storage pools per file system pools, min. for *system* pool, other data pools if required ► Define if metadata, data, or both to be replicated		
Real-time Compression	Define separate pools for the data and metadata		

10.7.3 Configuration options for block I/O access only

Table 10-4 shows the block I/O specific configuration options.

Table 10-4 Block I/O specific configuration options

Step/Purpose	Entry field or information	Comment or explanation	Your data or selection
Hosts	► Host names for FC or iSCSI ► WWPNs	► Create host objects with associated FC WWPNs or iSCSI IQN ► Create SAN zoning	
Storage configuration by host	► Capacity ► Storage pool layout ► Volumes ► Thin provisioning ► Easy Tier	► Pool and volume layout that is based on overall planning and modeling results ► Define thin provisioning arameters ► - Define Easy Tier start configurations (hybrid storage pools)	
Copy Services partnerships	► IBM Metro Mirror ► IBM Global Mirror ► Volumes and volume pairs ► Consistency groups	► To V7000, SAN Volume Controller, or Storwize V7000 Unified systems (but V7000 to V7000 Fibre Channel-attached thereof)	
FlashCopy requirements	► Volumes ► Type and options used ► Consistency Groups		

11

Implementation

This chapter describes the steps to implement the IBM Storwize V7000 Unified Disk System, from hardware setup to providing host storage. It is not expected that one person performs all of the steps, because several different skill sets are likely to be required during the process.

11.1 Process overview

The installation, implementation, and configuration tasks are grouped into major steps in the following processes. Each step is performed sequentially and, in most cases, must be completed before the next step can begin.

A task checklist is included to help in the implementation. It helps ensure that all steps are completed and serves as a quick reference for experienced implementers. The checklist can also be useful in planning a timeline and in identifying the resources and skills needed.

Important: Although the intent is to provide a complete checklist and procedures for implementation, always consult the latest product documentation in the information center the Help files. Where possible, references are included.

11.2 Task checklist

The major steps for installing and configuring the Storwize V7000 Unified system, which follow, give you a quick overview and an aid in planning. These steps are covered in detail in the following sections. Table 11-1 shows the implementation checklist.

Table 11-1 Implementation checklist

Task	Steps	Complete
Hardware rack and stack		
Preparation	Complete the planning checklist, including IP addresses, protocols, and server names and addresses.	
Packing slips	Check all of the items that you receive against the packing lists.	
Environmentals	Verify cooling and power, room access, and safety.	
Rack storage control enclosure	Rack mount the storage enclosure.	
Rack expansion enclosures	If any expansion enclosures are shipped, rack mount these now by using the recommended layout.	
Rack file modules	Rack the two file modules.	
Cabling	Power Control enclosures Expansion enclosures: SAS cables File modules Ethernet	
Power on	*Power on and check in this order:*	
Network	Switches, routers, and devices	
Power on storage enclosures	Storage expansions Storage control enclosure	
Power on file modules	Both	

Task	Steps	Complete
Software (skip if software is preinstalled)		
Prepare for reload if required		
Reinstall software if required		
Initialize		
Configure USB key	Set storage service IPs. Run the Init Tool and enter settings.	
Initialize the storage	Insert key into storage enclosure. Verify success.	
Initialize the file modules	Insert key into one file module. Verify success.	
Base configuration		
Configure and connect to GUI	Setup browser access and PuTTY.	
Easy Setup	Log on to run Easy Setup. Complete as much of the configuration as possible with information provided.	
Backups	Set up scheduled backup.	
Health check		
Run health checks	Verify that system is healthy.	
Security		
Change passwords	Change admin password on file modules and superuser password on block storage.	
Create users	Create more user logons as wanted.	
Storage controller (more configuration is using block storage)		
SAN requirements	Connect to SAN and zone.	
Configure storage	Configure and discover any external storage that is being used. Discover MDisks and build pools.	
Block storage		
Volumes	Configure volumes as required.	
Hosts	Define hosts and host ports to cluster.	
Mapping	Map volumes to hosts.	
Copy services	Configure copy services: - FlashCopy - Intercluster relationships - Remote copy, Global Mirror and Metro Mirror	

Task	Steps	Complete
File storage		
File systems	Create file systems from the pools.	
Files sets	Define file sets if you want them.	
Shares	Create shares. Add authorized user to shares.	

11.3 Hardware unpack, rack, and cable

For the following sections, see the IBM Storwize V7000 Unified Information Center for details and the updates:

http://pic.dhe.ibm.com/infocenter/storwize/unified_ic/index.jsp

11.3.1 Preparation

Be sure to read all of the topics in Chapter 10, “Planning for implementation” on page 97. This includes the physical environment, allocation of IP addresses and names, identification, preparation of network resources (for example, Domain Name System [DNS] and Network File System [NFS]), and access to the online information center.

Also, ensure that all personnel are familiar with safety information.

Tip: You can download the IBM Storwize V7000 Unified Model 2073-720 Quick Start Guide:

http://ibm.co/1orrgRe

11.3.2 Review packing slips and check components

Locate the packing slip in the shipping boxes and review it. Verify that you received all of the components and features that you ordered.

As a minimum, you need the items listed in the following subsections.

Control enclosure

At a minimum, you need the following control enclosure components:

- Control enclosure (Models 2076-112, 2076-124, 2076-312, or 2076-324). The last two digits of the model number identify the number of drive slots, either 12 or 24.
- Expansion enclosure (Models 2076-212 or 2076-224), if ordered.
- Rack-mounting hardware kit, including these components for each enclosure:
  - Two rails (right and left assembly)
  - Two M5 x 15 hex Phillips screws per rail (two rails)
  - Two M5 x 15 hex Phillips screws per chassis

  Note: Two parts of the rail kit are attached to each side of the enclosure.

- Two power cords.
- Drive assemblies or blank carriers (installed in the enclosure).

  Verify the number of drives and the size of the drives.

Other items that are shipped with control enclosure

The following support items are shipped with the control enclosure:

- Documents
  - Read First flyer
  - Quality Hotline flyer
  - Environmental flyers
  - Safety notices
  - Limited Warranty information
  - License information
  - License Function authorization document
  - IBM Storwize V7000 Quick Installation Guide
  - IBM Storwize V7000 Troubleshooting, Recovery, and Maintenance Guide
- Environmental notices CD
- Software CD that contains the publication PDF files and the information center content.
- One USB key, also known as a flash drive, is included with the publications.

Additional components for control enclosures

The following additional components, if ordered, are for the control enclosures:

- Fibre Channel cables, if ordered
- Small form-factor pluggable (SFP) transceivers that are preinstalled in the enclosure
- Longwave SFP transceivers, if ordered

Additional components for expansion enclosures

Two serial-attached SCSI (SAS) cables are needed for each expansion enclosure.

Two file modules

Each file module box contains the following components:

- File module (server)
- Rack-mounting hardware kit, including the following items:
  - Two sets of two rails (right and left assembly)
  - Large cable tie
  - Cable ties
  - Two sets of four M6 screws per rail (two rails)
  - Two sets of two 10-32 screws per chassis
  - Cable management support arm
  - Cable management arm mounting bracket
  - Cable management arm stop bracket
  - Cable management arm assembly

  Note: The rail kits for the servers differ from the control enclosure

- Two power cords

Additional components for file modules

The following additional components come with the file modules:

- Documents
  - Read First flyer
  - Quality Hotline flyer
  - Environmental flyers
  - Safety notices
  - Limited Warranty information
  - IBM Storwize V7000 Quick Installation Guide
  - IBM Storwize V7000 Troubleshooting, Recovery, and Maintenance Guide
  - License information
  - License Function authorization document
- Environmental notices CD
- Software CD that contains the publication PDFs and the information center content
- Small form-factor pluggable (SFP) transceivers that are preinstalled in the enclosure
- One USB key

11.3.3 Verify environmentals and planning

Review Chapter 10, “Planning for implementation” on page 97. Ensure that you understand locations for each component, that the environment has sufficient capacity in terms of power and cooling, and that rack space is available. Also, verify that the planning worksheet is complete.

You need two people to rack the modules. It is also recommended that two people connect the cables, because it makes the task much easier.

When racking, do not block any air vents. Usually 15 cm (6 inches) of space provides the appropriate airflow.

Do not leave open spaces above or below an installed module in the rack cabinet. To help prevent damage to module components, always install a blank filler panel to cover the open space to help ensure the appropriate air circulation.

For the following hardware installation tasks, see the online IBM Storwize V7000 Unified Information Center for the latest details or clarification:

http://ibm.co/1lzcUBs

11.3.4 Rack controller enclosures

Perform these tasks for the rack controller enclosures:

1. Install the rails in the rack for the controller enclosure. Allow 2 U for this module.
2. Now, you might find it easier to install the rails for the expansion enclosures and also the file modules if they are in the same rack. This is because there is more room if you do this step before installing any modules.
3. Remove the enclosure end caps by squeezing the middle of the cap and pulling.
4. From the front of the rack, with two people, lift the enclosure into position and align the rails. Carefully slide it into the rack until it is fully seated.
5. Insert the last two screws (one each side) to secure the enclosure to the rack, and then replace the end caps.

11.3.5 Rack expansion enclosures

Repeat these steps for each expansion enclosure:

1. Install the rails in the rack for the expansion enclosure if that is not already done. Allow 2 U.
2. Remove the enclosure end caps by squeezing the middle of the cap and pulling.
3. From the front of the rack, with two people, lift the enclosure into position and align the rails. Carefully slide it into the rack until it is fully seated.
4. Insert the last two screws (one each side) to secure the enclosure to the rack, and then replace the end caps.

11.3.6 Rack file modules

Follow these steps for each rack file module:

1. Install the rails in the rack for the expansion enclosure if that is not already done. Allow 2 U.
2. Extend the rails fully from the front of the rack.
3. From the front of the rack, with two people, lift the enclosure into position with the front slightly higher, and align the pins at the rear. Then, lower the front and align the front pins. Carefully slide it into the rack until it is fully seated and the latches are secure.
4. If required, insert the two screws (one each side) to secure the enclosure to the rack.
5. Install the cable management arm at the rear by using the instructions included with the arm. You can fit it on either side.
6. Repeat the preceding steps for the second file module.

11.3.7 Cabling

You need several cables to connect the modules of the Storwize V7000 Unified system and to connect to the network and servers. Most are required before you proceed with the installation. Connect the cables by following the steps in this section.

> **Tip:** Install the heaviest cables first (power) and the lightest cables last (fiber) to minimize the risk of damage.

Power

Each module has two power cords. Connect them to diverse power supplies. Electrically separate them as much as possible, preferably to different power strips in the rack that are powered from different distribution boards.

Control enclosures

Follow these steps for each control enclosure:

1. Ensure that the power switches on both power supplies are turned off.
2. On one power supply, prepare the cable retention bracket, release, and extend the clip and hold to one side.
3. Attach the power cable and ensure that it is pushed all the way in.
4. Push the retention clip onto the cable. Then, slide the clip down so that it fits snugly behind the plug.
5. Tighten the fastener around the plug.
6. Route the cable neatly to the power source and connect. Dress the cable away from the rear of the enclosure and secure any excess so it does not interfere with data cabling.
7. Repeat the preceding steps for the other power supply and install the second power cable.

Expansion enclosures (if present)

The power supplies are the same as the control enclosures. Connect the power cables for each expansion enclosure, two per enclosure, by using the same procedure.

File modules

Follow these steps for each file module:

1. Attach the first power cable to the file module and ensure that it is pushed all the way in.
2. Route the cable through the cable management arm, allowing plenty of slack so that the cable does not become tight when the arm and module are extended.
3. Connect to the power source and connect. Dress the cable and secure any excess so it does not interfere with data cabling.
4. Repeat for the second power cable and for the two power cables in the second module.

Ethernet

There are six Ethernet ports on each file module and four on each control enclosure, with an optional four more if an Internet Small Computer System Interface (iSCSI) is specified. The 1 Gb ports require a copper cable with a minimum of CAT5 UTP. The 10 Gb ports are connected by using fiber cables. They need multimode (MM) fiber cables with LC connectors. Connect them by following these instructions and illustrations.

File modules

Use Figure 11-1 as a reference for plug locations. Route each cable through the cable management arm. Connect the following cables for *each* file module.

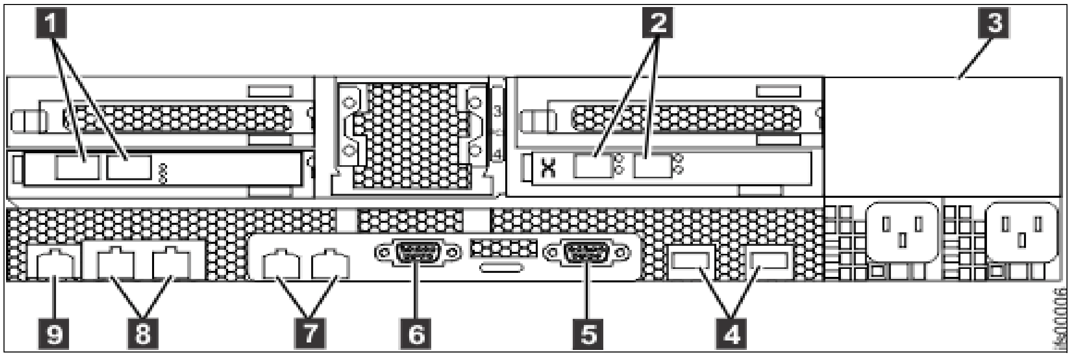


Figure 11-1 File module rear

Port 1: 1 Gb **7** left (required) Internal connection between file modules. Connect a short cable from this port to Port 1 on the other file module.

Port 2: 1 Gb **7** right (required) Internal connection between file modules. Connect a short cable from this port to Port 2 on the other file module.

Port 3: 1 Gb **8** left (required) Provides management connection and optional data.

Port 4: 1 Gb **8** right (optional) Alternate management connection and optional data.

Slot 4-0: 10 Gb **2** right (optional) Data connection only.

Slot 4-1: 10 Gb **2** left (optional) Data connection only.

Control enclosure

Use Figure 11-2 as a reference for plug locations. Connect the cable to the Ethernet port and route them neatly to the rack cable management system.

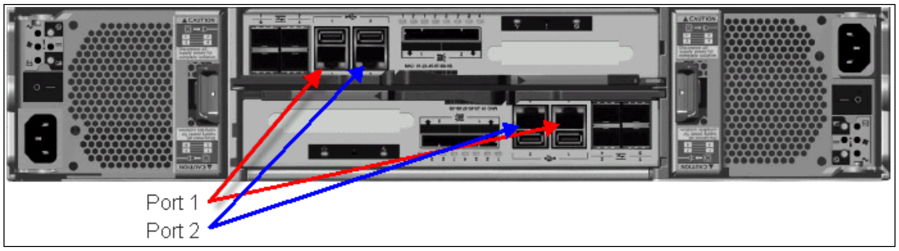


Figure 11-2 Control module rear

Port 1: 1 Gb (required) Management and service connection.

Port 2: 1 Gb (optional) Alternate management connection.

SAS cables

If expansion enclosures are installed, they are connected to the control enclosure by using the SAS cables that shipped with your system. If no expansion enclosures are installed, skip this step.

The control enclosure has two SAS ports on each node canister. Port 1 from each canister forms a pair of chains. These normally connect to the expansion enclosures racked below the control enclosure. Port 2 from each node canister forms the second chain, which normally connects to the upper expansions. The top canister always connects to the top canister of the next enclosure and the bottom to the bottom.

Connect the expansion enclosures by using Table 11-2 on page 151. See Figure 11-3 for port location.

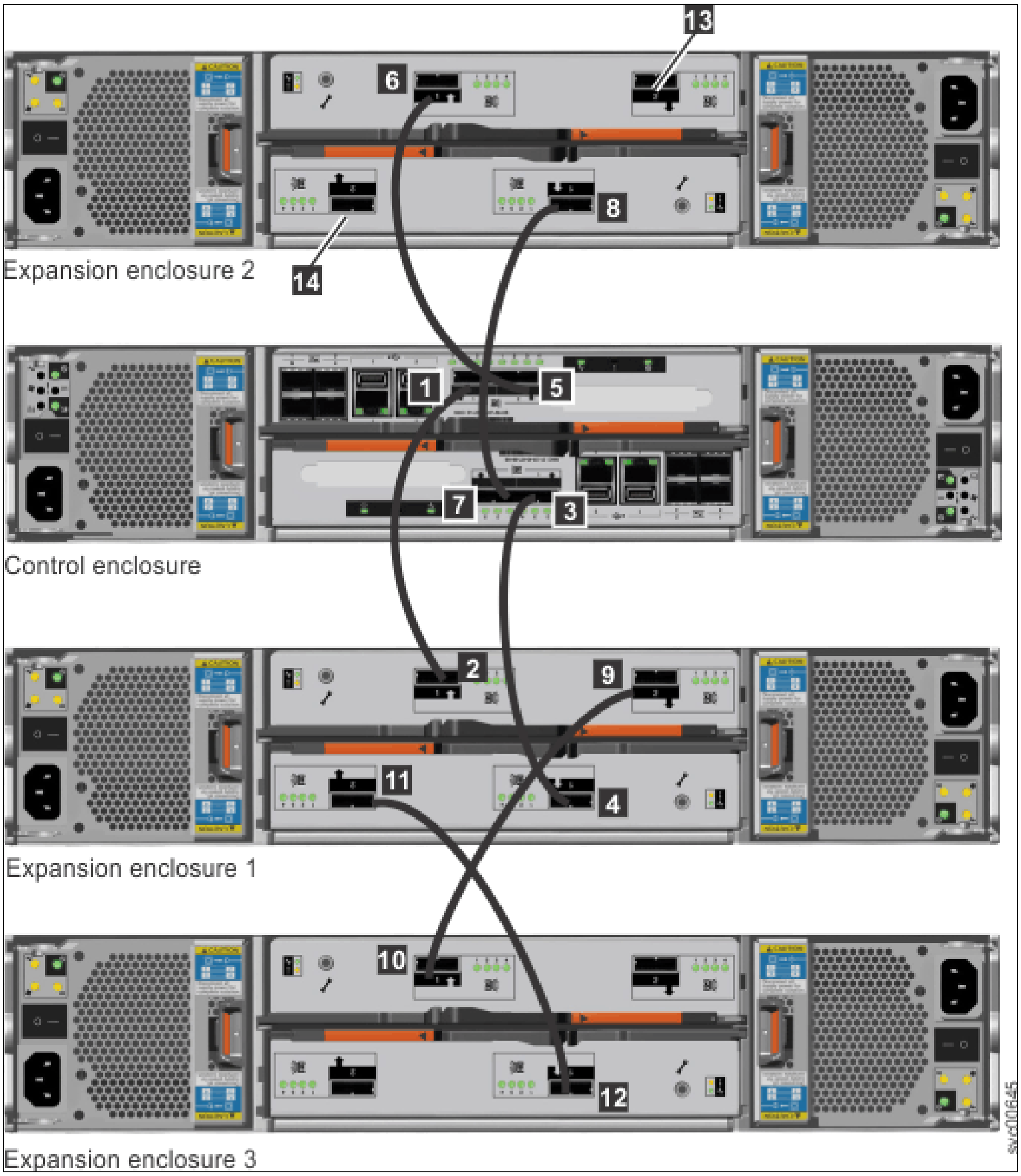


Figure 11-3 SAS cabling for three expansions

Table 11-2 shows the SAS connections.

Table 11-2 SAS connections

SAS connections: How each unit connects to the next unit in the chain		
First unit	**Second unit**	**Number of expansions**
Controller	**Expansion 1**	**1 Expansion**
Upper canister Port 1	Upper canister Port 1	
Lower canister Port 1	Lower canister Port 1	
Controller	**Expansion 2**	**2 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 1	**Expansion 3**	**3 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 2	**Expansion 4**	**4 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 3	**Expansion 5**	**5 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 4	**Expansion 6**	**6 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 5	**Expansion 7**	**7 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 6	**Expansion 8**	**8 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	
Expansion 7	**Expansion 9**	**9 Expansions**
Upper canister Port 2	Upper canister Port 1	
Lower canister Port 2	Lower canister Port 1	

Fiber optic cables

The default configuration is shortwave SFPs, which require MM cables with LC connectors. If a long distance is involved, the SFPs can be specified as longwave. In that case, it is important to use the appropriate single mode (SM) cable.

Dress the cable gently and ensure that there are no kinks or tight bends. Do not use cable ties or any other hard material to hold cables because these cause kinks and signal loss. Hook-and-loop fastener cable wraps provide the most economical and safe method of tying cables.

File module

The use of 10 Gb Ethernet cables is covered in “Ethernet” on page 149. The other fiber optic cables that are connected to the file module are Fibre Channel (FC) cables that connect to the control enclosure.

There are four FC cables required, with two from each file module to each node canister in the control enclosure. These are connected in a mesh pattern to provide full redundancy of pathing.

Select short cables because the file module and control enclosure are normally close to each other in the same rack. Connect the cables as shown in Table 11-3.

Table 11-3 Cable connections

File module	Controller
File module 1: Fibre Channel Slot 2, Port 1	Upper canister Fibre Channel Port 1
File module 1: Fibre Channel Slot 2, Port 2	Lower canister Fibre Channel Port 1
File module 2: Fibre Channel Slot 2, Port 1	Upper canister Fibre Channel Port 2
File module 2: Fibre Channel Slot 2, Port 2	Lower canister Fibre Channel Port 2

See Figure 11-4 on page 153 for connector locations.

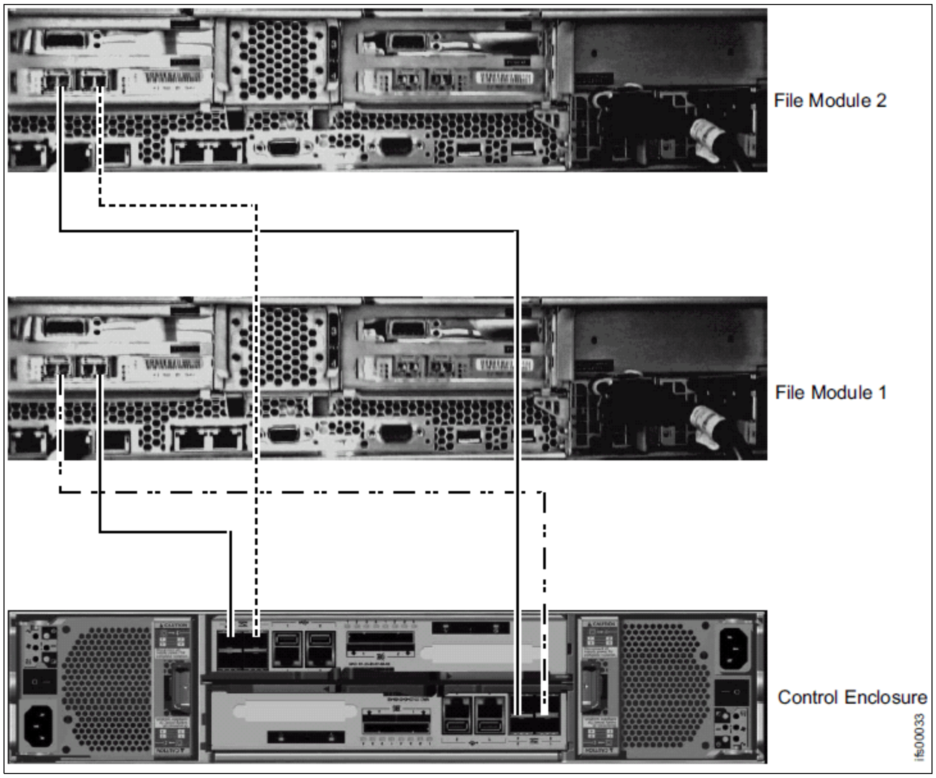


Figure 11-4 File to control modules, Fibre Channel cables

Control enclosure

If you are configuring block storage, FC connections are required to your storage area network (SAN) fabrics. Two FC ports are available on each node canister for connection to the SAN. Connect these in a mesh arrangement to provide fabric path redundancy.

11.4 Power on and check-out procedures

In the topics that follow, we describe how to power on and check out the system.

11.4.1 Network readiness

Ensure that all network devices are powered on and ready, including Ethernet and SAN switches. Access to network resources also needs to be available from the Storwize V7000 Unified system, including DNS, Network Time Protocol (NTP), and email server.

11.4.2 Power on expansions and controllers

Power on the disk storage first. Start with the expansion enclosures, and then the control enclosure, followed by the file modules. Ensure that each group of enclosures is successfully powered on before you begin the next group.

Expansion enclosures

Power on both power supplies by using the switch on each one.

Successful power-on is indicated by the following status lights:

- Power supply LEDs: Power LED on, three fault LEDs off on both power supplies
- Front left-end cap: Power LED on, all others off
- Canister LEDs: Status LED on, fault LED off on both canisters
- Drive LEDs: These come on as each drive becomes ready

Control enclosure

Power on both power supplies by using the switch on each one.

Successful power-on is indicated by the following status lights:

- Power supply LEDs: Power LED on, battery good LED on or flashing (if charging), four fault LEDs on both power supplies are off
- Front left-end cap: Power LED on, ignore others
- Canister LEDs on both canisters: Power status LED on, fault LED and system status depend on current state and can be ignored at this time
- Drive LEDs: These come on as each drive becomes ready

11.4.3 Power on file modules

The power management and support functions are running if there is AC power on one or both of the power cables, but the main processor is not powered on. When power is applied to the cable, the management functions start. This process takes approximately 3 minutes. When complete, as indicated by a flashing power LED on the front of the module, the unit is ready to be powered on.

Connect a keyboard and display to the module and connect a mouse, if it is a new installation. Monitor the boot process on the display.

Power on by pressing the power button on the front of the module. The power LED light shows solid and the boot process starts. If the file module has code that is loaded, the module boots and completion is indicated by a flashing blue attention LED. If no code is present, a message is displayed on the display to indicate that the boot failed.

11.5 Install the latest software

Ensure that as part of the implementation, the latest software levels are installed. In most cases, a more recent level is available than that included on the module, so an upgrade is required. Or, if the file modules have no software installed, the full package can be installed by using the procedure that follows. The control enclosure (Storwize V7000 system) is automatically upgraded, if required, as part of the file module upgrade.

In most cases, the Storwize V7000 Unified system is preloaded, and a software upgrade can be performed when the cluster initialization and configuration is complete. In this case, you can skip the rest of this section and proceed to 11.6, “Initialize the system” on page 157.

If the previous state of the cluster is not known or the integrity of the software is suspect, perform the full restore procedure to begin with a clean software installation.

For details about the software structure and concurrent upgrade processes, see 11.5, “Install the latest software” on page 154. The full DVD restore process is rarely required, because it is done as part of an initialization, so it is included here.

11.5.1 Determine current firmware and versions

This process can be difficult to determine because it generally requires the system to be booted and running and the logon user and password details known. If the intent is to restore the code, regardless, the previous level is not important and the restore can proceed.

To determine the file module software level, connect a keyboard and display to the module. Boot and wait for the login prompt. The initial prompt is preceded by the host name. If console messages were displayed, press **Enter** to restore the prompt message:

```
login:
```

The default user/password is `admin/admin`. If the cluster was in use previously, the user settings might have been changed, so you need to know a valid user login combination.

When logged in, issue the **`lsnode`** command. This displays the file nodes. The software level is listed under the “Product version” heading. If the modules are not currently a cluster, you might need to repeat the process on the other module. If the modules are at different software levels, the node that is performing the initialization (the one that you put the key in) reloads the other node to the same level, so they match.

If you want to know the level of the Storwize V7000 control enclosure, connect to the Service Assistant directly by using the service IP address of one of the nodes, or connect to the management address followed by **`/service`** and log on by using the superuser password. If you do not know the password or IP or if they are not the defaults, wait until the initialization step, 11.6.2, “Initialize the Storwize V7000 controller” on page 160, is finished. Then, you can connect with the management IP address that you defined and the default superuser password of `passw0rd`.

11.5.2 Preparation for reload

We describe preparing to reload the code in the topics that follow.

Download and prepare software

To do a full DVD restore, you must obtain the International Organization for Standardization (ISO) image from IBM Support.

Using a suitable application, burn the DVD from the *`xxx`*`.iso` file. This file is large, over 4 GB, and creates a bootable DVD that contains all the required software.

The file modules are restored individually. If time is critical, you might choose to burn two DVDs and perform the installations at the same time, although the process ejects the DVD after about 25 minutes. Therefore, the second module can be started then, which works well.

Prepare the hardware

> **Important:** This process reinstalls the software from scratch. The following steps are not required with a new installation. They are necessary only if the file module was redeployed or the installation was aborted and cleanup is necessary. These steps might also be required as part of a recovery. Perform these steps only if you determine that reinstallation required. We include them here merely for reference.

The file modules must be installed in the rack and the power must be connected. The power indicator on the front of the module flashes to indicate that the power control module is booted and the server is powered off and ready to start.

Connect a standard PC monitor and keyboard (or an equivalent KVM tool) to the module.

1. If this module was used as a file module, you must remove the previous configuration to achieve a clean installation. It is important to remove the entire `/persist/` directory.
2. If any hardware was replaced in the server or if the basic input/output system (BIOS) settings are in doubt, perform the BIOS reset procedure, which is explained in the information center. This explains the following process:
   a. Power on the server by using the power button on the front.
   b. When the BIOS window opens, press **F1** to start the BIOS setup.
   c. Select **Load Default Settings**.
   d. Select **Boot Manager**.
   e. Select **Add device**.
   f. Select **Legacy only**.
   g. Use the Escape key to exit the panels and save the configuration on exit.
3. If necessary, configure the Redundant Array of Independent Disks (RAID) controller on the server for the mirrored disk drive that is used for the system by using the procedure described in the information center.

11.5.3 Reinstall the software

Power on the server. If you have just connected the AC power, you must wait a few minutes for the power controls to initialize before you can power on the system. If it is already on, reboot the server by using **Ctrl-Alt-Del**, or if in the BIOS window, by exiting the BIOS setup utility.

Power must be on to open the DVD drive. Therefore, as soon as the server is powered up, insert the DVD in the drive.

Watch the video monitor and wait for the DVD to boot. The software installation utility immediately prompts for confirmation with this message:

```
- To install software on a node, press the <ENTER> key.
*NOTE* - this will destroy all data on the node

- Use the Function keys listed below for more information

[F1 - Main] [F5 - Rescue]
boot:
```

Press **Enter** to begin the installation process.

The utility now builds the disk partitions, if they are needed, and begins loading the Linux operating system and the Storwize V7000 Unified system software. There are at least two reboots to complete the installation. The first phase copies the software to the disk. Progress can be monitored on the display. After about 25 minutes, the tool ejects the DVD and reboots the server.

> **Note:** After 25 minutes, the DVD is ejected and can be used to begin the installation process on the other server, if required.

Then, the server is booted from its disk, and Linux installs its components and builds the Linux operating system and installs the Storwize V7000 Unified software. This phase takes about 30 minutes.

When this process is complete, the server is booted by using the operational code and is ready for use. This boot takes less than 5 minutes.

Successful installation and preparation of the file module is indicated by the blue attention light flashing on the server.

Repeat the process for the other file module.

11.6 Initialize the system

The next step is to initialize the system and build a cluster that incorporates the Storwize V7000 storage (control enclosure and any expansion enclosures) and the two file modules.

11.6.1 Configure the USB key

A USB key is shipped with the Storwize V7000 Unified system. If the key is misplaced, any USB mass storage key can be used. However, some models of keys are not recognized, so you might need to try a few. Take care not to use a large capacity key because that can keep the key from being recognized.

The key must be formatted with an FAT32, EXT2, or EXT3 file system on its first partition.

The shipped key is formatted and preinstalled with the initialization tool. If this is missing or you are using a replacement key, you can download this tool from the IBM Support site:

http://www.ibm.com/storage/support/storwize/v7000/unified

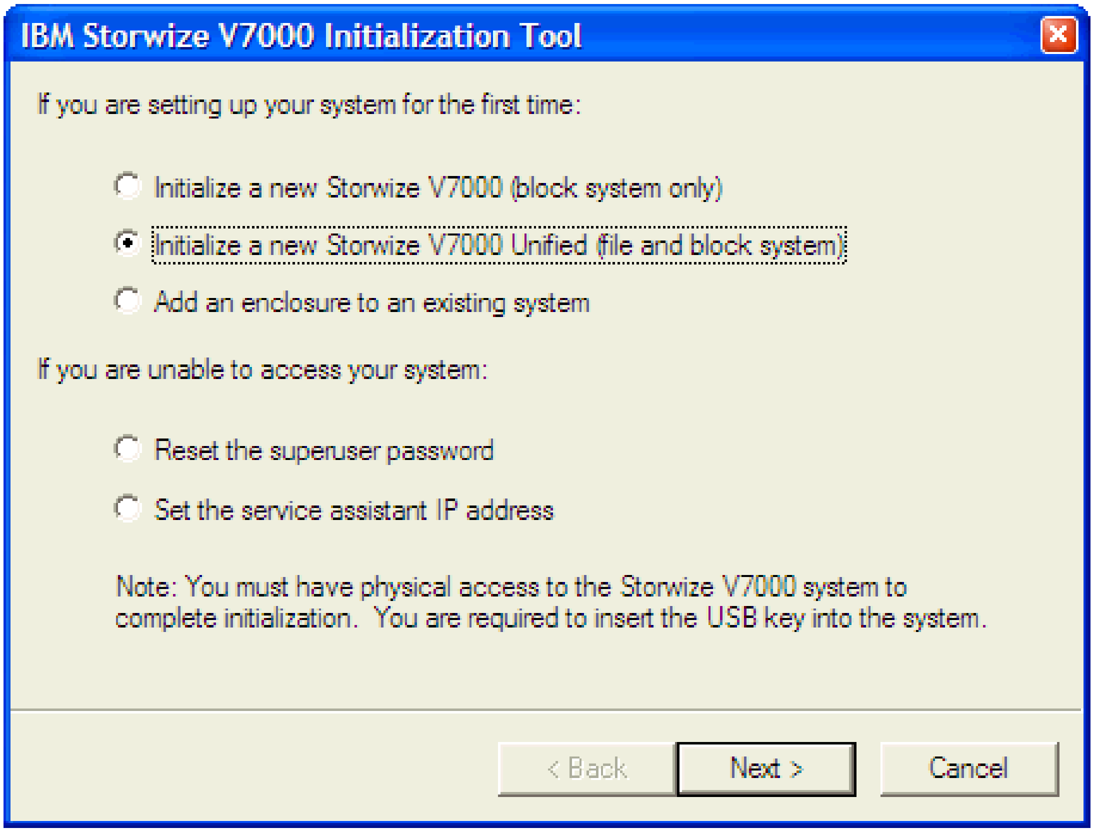
> **Note:** There is a key included with the Storwize V7000 controller unit and one with each file module. Use the key that is included with a file module, because it is likely to have the latest version of the Init Tool. If you download the latest tool, any included key works.
>
> Keep the keys stored and secure for later use in case of recovery, rebuild, or redeployment.

Follow these steps to configure the USB key:

1. Insert the key into a supported version of Microsoft Windows. If the tool does not automatically launch, open the USB key and run **`InitTool.exe`**. This launches the window shown in Figure 11-5 on page 158.

USB key files: The USB key contains the following files:

`autorun.inf`	Windows auto run file
`inittool.exe`	Storwize Initialization Tool (Init Tool)


IBM Storwize V7000 Initialization Tool

If you are setting up your system for the first time:

○ Initialize a new Storwize V7000 (block system only)

● Initialize a new Storwize V7000 Unified (file and block system)

○ Add an enclosure to an existing system

If you are unable to access your system:

○ Reset the superuser password

○ Set the service assistant IP address

Note: You must have physical access to the Storwize V7000 system to complete initialization. You are required to insert the USB key into the system.

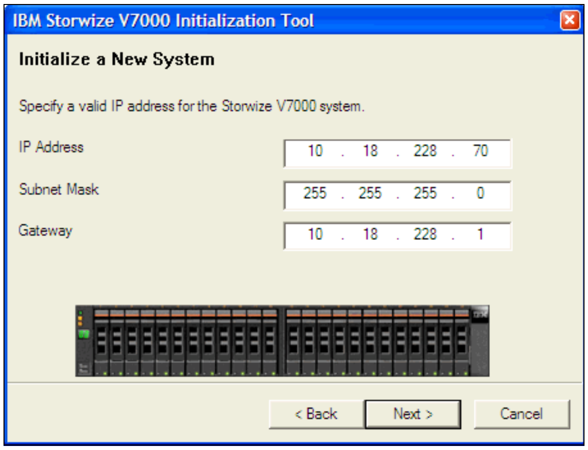
Figure 11-5 Init tool first window

2. Set the storage service IP addresses. The service IP addresses assigned to the Storwize V7000 storage enclosure nodes are not considered part of the cluster configuration. These addresses are the base addresses for the nodes and are active independently from the cluster software. They are not set or changed by the cluster configuration. It is important to set these addresses in case access is needed during recovery.

   Use the Set Service Assistant IP address option on the Init Tool to prepare the USB key. Insert the key in the enclosure nodes to set their address. The tool creates one address. Therefore, the procedure needs to be done twice, once for each node.

3. Select the **Unified** option (File and Block). This universal tool is used for all Storwize V7000 installations. The first option, *block system only*, sets up the initialization process for installing just a Storwize V7000 storage controller. The second option is required to initialize both the storage controller and the file modules for a unified configuration.

4. Click **Next** to get the first setup options, as shown in Figure 11-6.


IBM Storwize V7000 Initialization Tool

Initialize a New System

Specify a valid IP address for the Storwize V7000 system.

IP Address 10 . 18 . 228 . 70

Subnet Mask 255 . 255 . 255 . 0

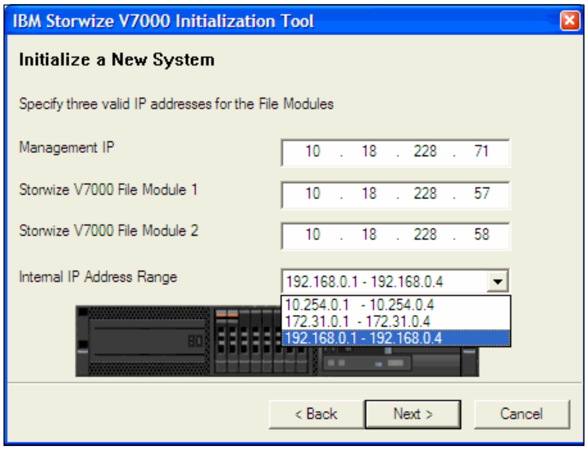
Gateway 10 . 18 . 228 . 1

< Back | Next > | Cancel

Figure 11-6 Init tool system settings

This IP address is the management address for the *System* component, which is the Storwize V7000 storage controller. As explained in the planning section, all of the management IP addresses must be in the same subnet.

5. Enter the IP address, mask, and optional gateway. Then, click **Next** to get the window shown in Figure 11-7.


IBM Storwize V7000 Initialization Tool

Initialize a New System

Specify three valid IP addresses for the File Modules

Management IP 10 . 18 . 228 . 71

Storwize V7000 File Module 1 10 . 18 . 228 . 57

Storwize V7000 File Module 2 10 . 18 . 228 . 58

Internal IP Address Range 192.168.0.1 - 192.168.0.4

10.254.0.1 - 10.254.0.4
172.31.0.1 - 172.31.0.4
192.168.0.1 - 192.168.0.4

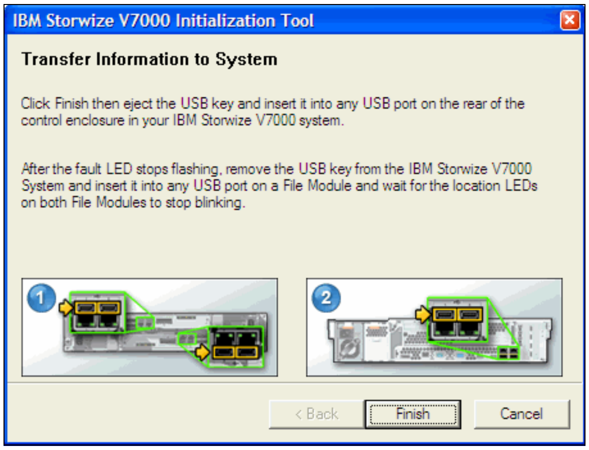
< Back | Next > | Cancel

Figure 11-7 Init tool file settings

6. On this panel, give the IP details for the unified management interface and the IP addresses for file modules 1 and 2.

   From the pull-down menu, select a range of addresses that are not being used in your network. These addresses are used internally between the modules and cannot be accessed from your network.

7. Click **Next** to see the instructions to use the key, as shown in Figure 11-8.


IBM Storwize V7000 Initialization Tool

Transfer Information to System

Click Finish then eject the USB key and insert it into any USB port on the rear of the control enclosure in your IBM Storwize V7000 system.

After the fault LED stops flashing, remove the USB key from the IBM Storwize V7000 System and insert it into any USB port on a File Module and wait for the location LEDs on both File Modules to stop blinking.

Figure 11-8 Init tool final window

8. Click **Finish**. Verify that the key now has two new files, `satask.txt` and `cfgtask.txt`.

> **Tip:** If you need to try the following steps again, it is wise to make a copy of the files on the key. Store the copies on your workstation so it is easier to resume from this point.

9. Eject and remove the USB key from your workstation.

> **USB key files:** The USB key contains the following files:
>
> ```
> autorun.inf ........ Unchanged
> inittool.exe ....... Unchanged
> satask.txt ......... Storage command file, contains the initialize command
> cfgtask.txt ........ File module command file, contains initialize command
> ```

11.6.2 Initialize the Storwize V7000 controller

Follow these steps to initialize the Storwise V7000 controller:

1. Ensure that both the node canisters are in *candidate* status before continuing. If you know the service IP address for either node, browse to that address and log on by using the superuser password (default: `passw0rd`). The status of the nodes is displayed on the home window. Or, verify that the status of the three LEDs on each canister is flashing on-off. Read from the right on the top canister and from the left on the bottom one.
2. If the status is incorrect, connect to the Service Assistant interface to resolve this. If it is not known, the IP address can be set by using the USB key. If the status shows as Service, perform the action to remove it from service. If the status is Active, an operational cluster is present and needs to be removed. Follow the Storwize V7000 procedures to remove the cluster or contact IBM for assistance.
3. When you have both nodes in candidate status, insert the key in any USB port on the rear of your Storwize V7000 controller enclosure. It is better to use a USB port on the top canister because the canister that is executing the initialize command becomes the first node, *node1*. This helps prevent confusion later, with node1 now being in Slot 1. The fault LED begins flashing. When the fault LED stops flashing, the task is complete and the key can be removed.

Tip: After the command is executed, a result file is written onto the key and a Secure Shell (SSH) key file. You can review the result file, `satask_result.html`, to verify that the creation of the NAS cluster was successful. Also, when successfully executed, the command file (`satask.txt`) is deleted to prevent it from being run again accidentally.

4. Verify that the three status LEDs are showing on-off on both canisters in the enclosure, indicating that the node canisters are active in a cluster.

Tip: In case the following steps need to be tried again, make a copy of the files on the key now, and store it on your workstation so it is easier to resume from this point.

USB key files: The USB key contains the following files:

```
autorun.inf ........ Unchanged
inittool.exe ....... Unchanged
.................... If successful, satask.txt has been deleted
cfgtask.txt ........ File module command file, contains initialize command
NAS.ppk ............ SSH key file, needed by the file module
satask_result.html.. Result output from the initialize command
```

11.6.3 Initialize the file modules

Initialize the file modules by performing the following tasks:

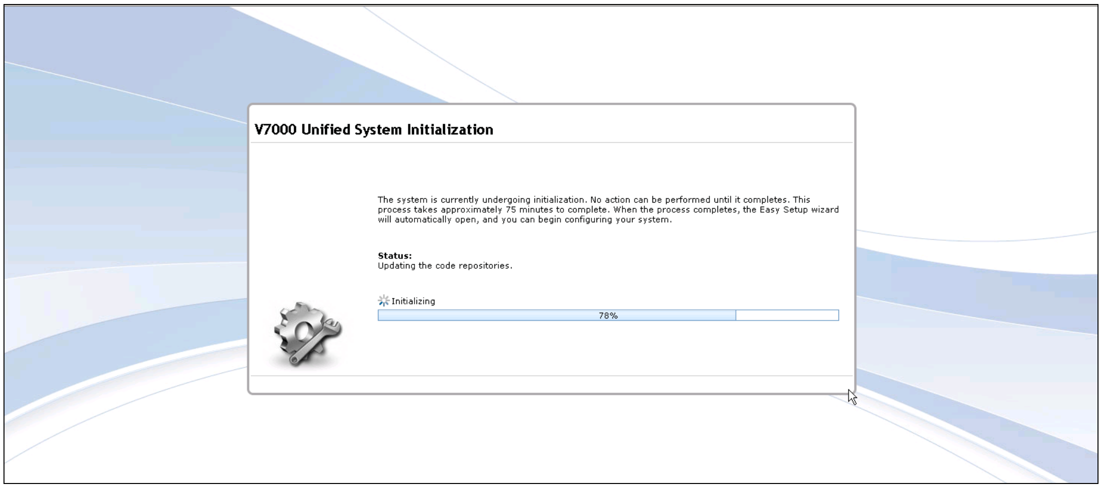
1. Verify that both the file modules are booted and ready to be initialized. This is indicated by the flashing blue attention indicator on each file module. Both must have this LED flashing before you continue.
2. Now insert the key into any USB port on one file module.
3. The blue LED comes on with a steady display and remains on until the process is finished. If the blue attention LED on the module that the key is inserted into begins flashing again, that indicates a failure. If this happens, remove the key and check the results file. See the information center for how to analyze the error and for recovery actions. Successful initialization is indicated by both blue LEDs turning off.
4. Wait for the initialization process to complete (both attention lights turn off). Normal initialization takes approximately 15 minutes. However, this time can be extended if the other file module needs to be upgraded to the same version (plus 1 hour) or if the control enclosure (Storwize V7000) requires a code upgrade (plus 2 hours).
5. Remove the USB key and insert the key into your workstation and perform either of these actions:
   a. Review the results file in an editor.
   b. Start the Init Tool program from the key (if it did not automatically run). This action inspects the results file and issues a message saying that the initialization was successful.

USB key files: The USB key contains the following files:

```
autorun.inf ........ unchanged
inittool.exe ....... unchanged
.................... If successful, cfgtask.txt has been deleted
.................... NAS.ppk is deleted when copied to File module
satask_result.html . unchanged
SONAS_result.txt ... Result output from the initialize command
```

6. Verify that you can access the cluster. Using the browser on a workstation that has IP connectivity to the management ports of the Storwize V7000 Unified system, select your management port IP file (`https://management_port_ip`). This is the management IP that you assigned to the Storwize V7000 Unified cluster when you initialized the USB key.

Initialization of the system begins and a window opens that is similar to Figure 11-9.



V7000 Unified System Initialization

The system is currently undergoing initialization. No action can be performed until it completes. This process takes approximately 75 minutes to complete. When the process completes, the Easy Setup wizard will automatically open, and you can begin configuring your system.

Status:
Updating the code repositories.

Figure 11-9 Storwize V7000 Unified system initialization

11.7 Base configuration

The Storwize V7000 Unified cluster is now created and consists of two main components:

- The storage consists of a control enclosure and optional expansion enclosures.
- The file server consists of two file modules.

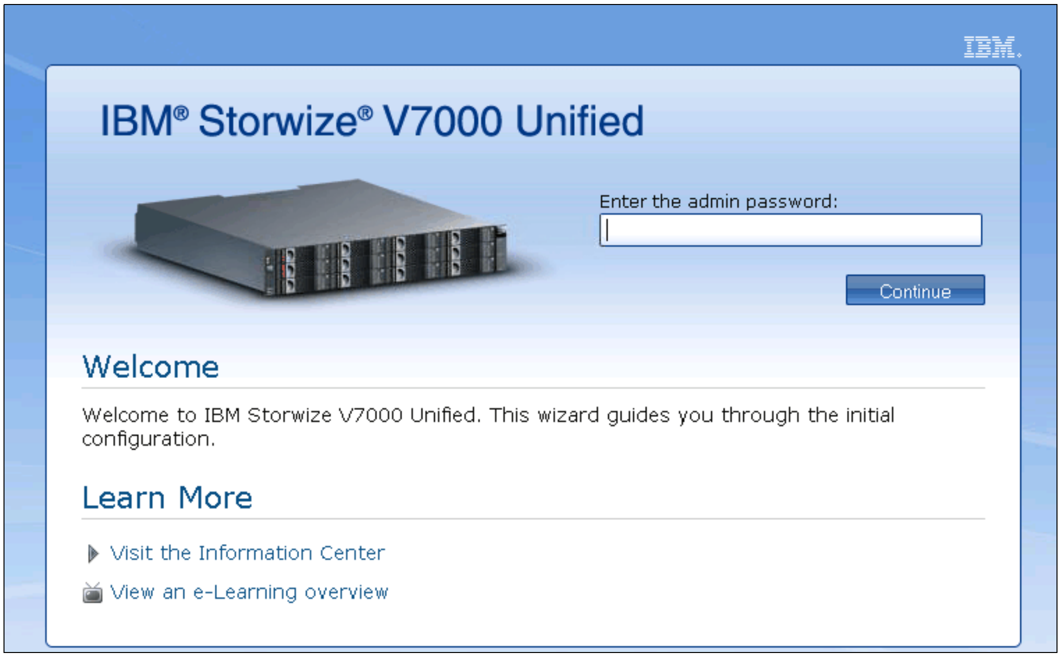
These can all be managed from the Storwize V7000 Unified graphical user interface (GUI) interface, which is presented from the primary file module's management IP address.

To use the cluster, it must first be configured. A setup tool called Easy Setup is provided. It runs only once, the first time that someone logs in to the cluster after initialization. The steps to configure the cluster by using Easy Setup are described next.

> **Note:** Easy Setup runs only once and cannot be started manually. Any configuration options that are skipped must be manually set up or changed from the appropriate configuration panels later.

11.7.1 Connect to the graphical user interface

Using the browser on a workstation that has IP connectivity to the management ports of the Storwize V7000 Unified system, select the IP file again (**https://*management_port_ip***). Next, the login window shown in Figure 11-10 opens. Ensure that you are connected directly to the Storwize V7000 Unified system, *not* the Storwize V7000 storage control enclosure.


IBM.
IBM® Storwize® V7000 Unified
Enter the admin password:
Continue

Welcome

Welcome to IBM Storwize V7000 Unified. This wizard guides you through the initial configuration.

Learn More
Visit the Information Center
View an e-Learning overview

Figure 11-10 Easy Setup login

This login window leads to the Easy Setup wizard, which guides you through the base configuration of the machine.

The default user name is `admin` and the default password is `admin`. For this initial setup login, only `admin` is available. Enter the password and click **Continue**.

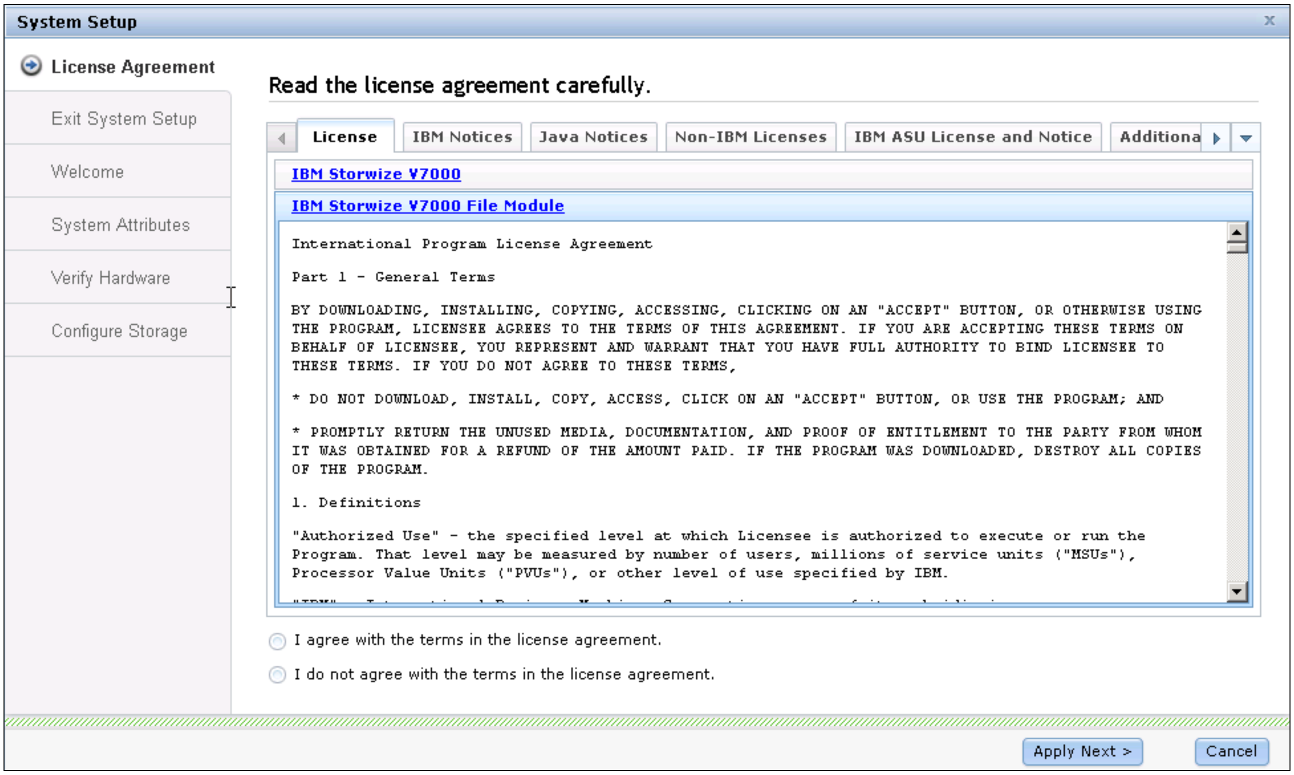
11.7.2 Easy Setup wizard

Important: It is important to have all of your parameters for configuring the cluster ready before you continue, preferably written on your planning sheet. Also, ensure that the various servers are operational and ready. Many of the configuration processes described in the following procedures test and confirm that the resources that you address are reachable and operating (for example, DNS servers, authentication servers, gateways, and so on). The setup steps might fail if you cannot contact them and connect to them.

License agreement

The first windows that Easy Setup presents are for the license agreement, as shown in Figure 11-11.

1. Read the license agreements on each tab and click the appropriate button.
2. Then, click **Next** to continue.



International Program License Agreement

Part 1 - General Terms

BY DOWNLOADING, INSTALLING, COPYING, ACCESSING, CLICKING ON AN "ACCEPT" BUTTON, OR OTHERWISE USING THE PROGRAM, LICENSEE AGREES TO THE TERMS OF THIS AGREEMENT. IF YOU ARE ACCEPTING THESE TERMS ON BEHALF OF LICENSEE, YOU REPRESENT AND WARRANT THAT YOU HAVE FULL AUTHORITY TO BIND LICENSEE TO THESE TERMS. IF YOU DO NOT AGREE TO THESE TERMS,

* DO NOT DOWNLOAD, INSTALL, COPY, ACCESS, CLICK ON AN "ACCEPT" BUTTON, OR USE THE PROGRAM; AND

* PROMPTLY RETURN THE UNUSED MEDIA, DOCUMENTATION, AND PROOF OF ENTITLEMENT TO THE PARTY FROM WHOM IT WAS OBTAINED FOR A REFUND OF THE AMOUNT PAID. IF THE PROGRAM WAS DOWNLOADED, DESTROY ALL COPIES OF THE PROGRAM.

1. Definitions

"Authorized Use" - the specified level at which Licensee is authorized to execute or run the Program. That level may be measured by number of users, millions of service units ("MSUs"), Processor Value Units ("PVUs"), or other level of use specified by IBM.

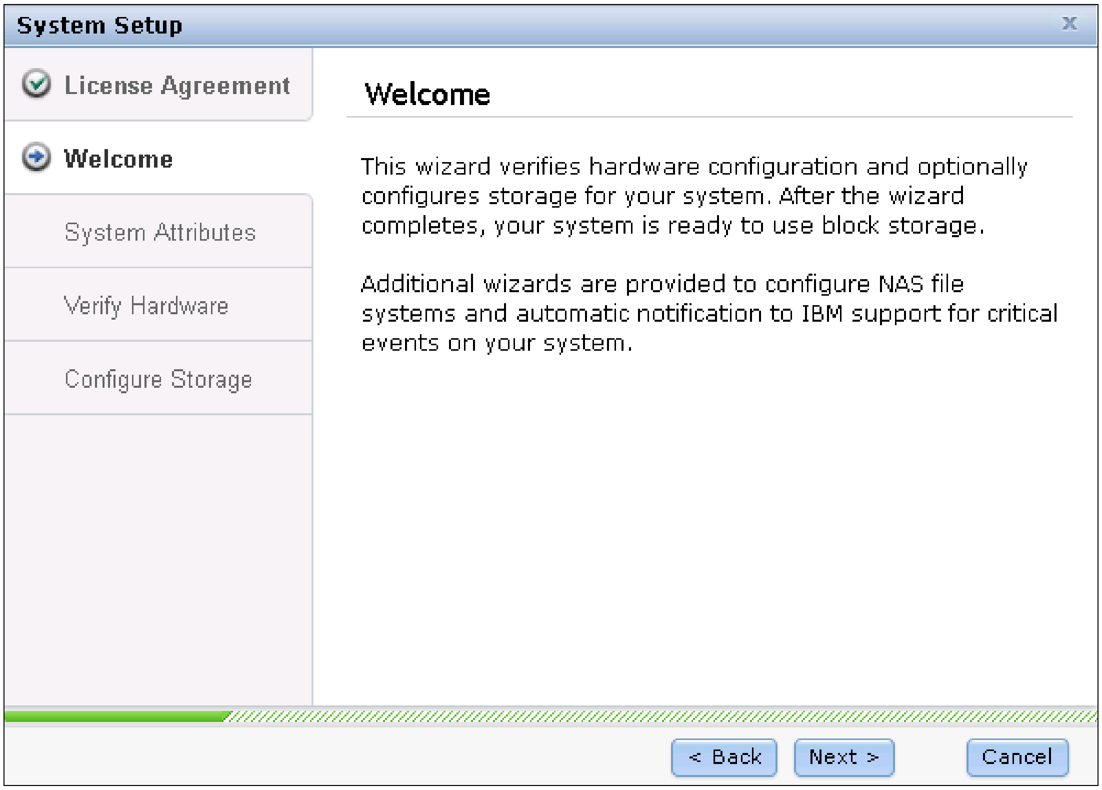
○ I agree with the terms in the license agreement.

○ I do not agree with the terms in the license agreement.

Figure 11-11 License Agreement window

Welcome window

After the license agreement has processed, you see the **Welcome** window that is shown in Figure 11-12 on page 165.



Welcome

This wizard verifies hardware configuration and optionally configures storage for your system. After the wizard completes, your system is ready to use block storage.

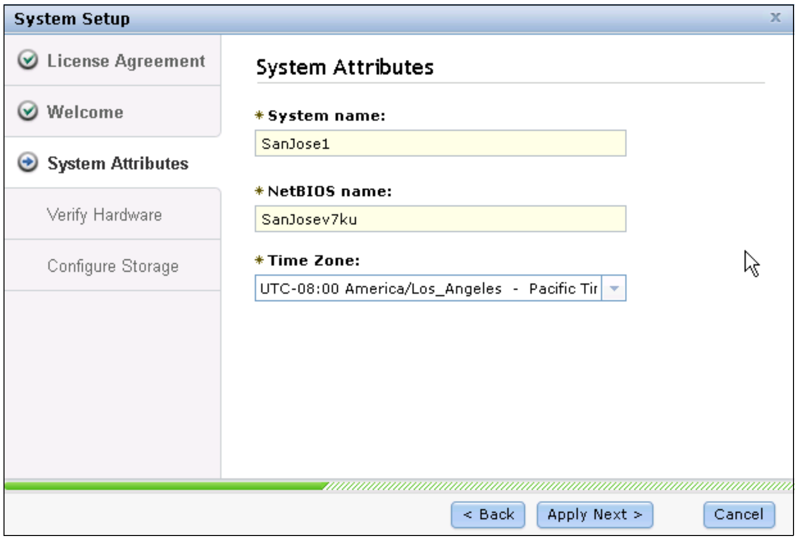
Additional wizards are provided to configure NAS file systems and automatic notification to IBM support for critical events on your system.

Figure 11-12 Welcome window

To proceed with the setup, click **Next.**

System attributes

1. Enter the required information in the fields in the System Attributes window, as shown in Figure 11-13.



System Attributes

*System name: SanJose1

*NetBIOS name: SanJosev7ku

*Time Zone: UTC-08:00 America/Los_Angeles - Pacific Tir

Figure 11-13 System attributes

The following field definitions describe the system attributes:

System name The name of this system: SanJose1. This name is displayed on the screens as the name for this cluster.

NetBIOS name The NETBIOS name for this cluster is what the network shows. This name is used in the Server Message Block (SMB) protocol.

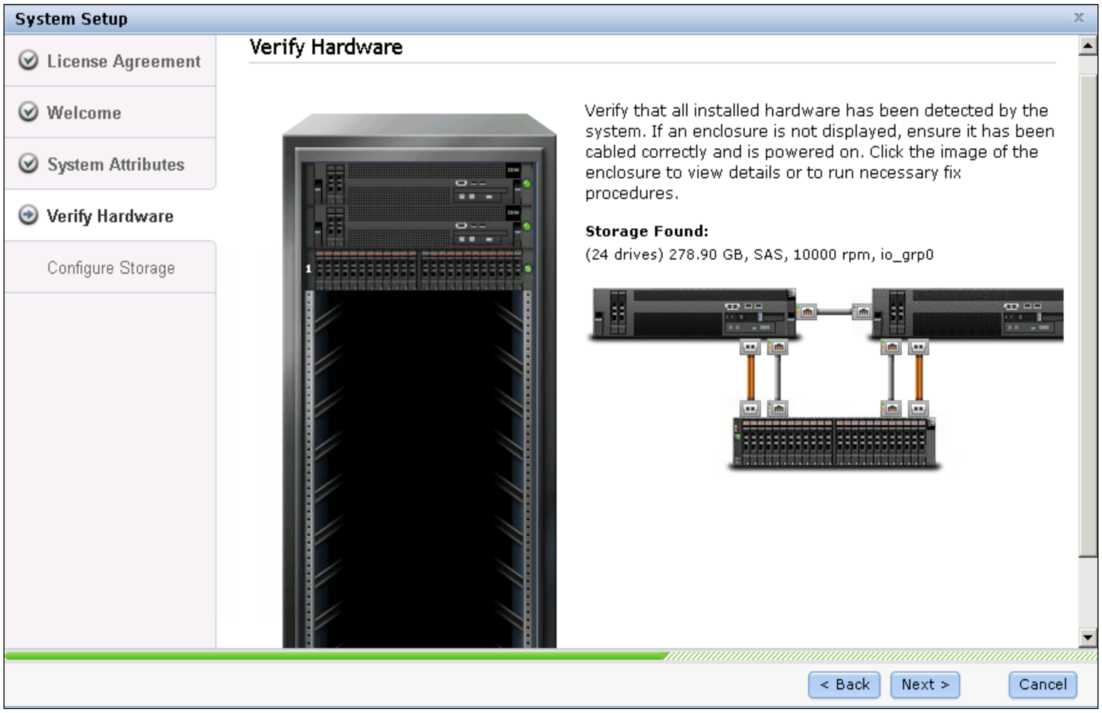
Time zone Choose the time zone from the selection that best represents the location of this machine.

2. Click **Next**. A progress window opens while configuration changes are made.
3. Wait for Task Completed, and then close.

Verify hardware

You are now presented with a graphical representation of the hardware.

1. Check that all modules and enclosures are correctly shown and that the cabling is correct. See Figure 11-14.



Verify that all installed hardware has been detected by the system. If an enclosure is not displayed, ensure it has been cabled correctly and is powered on. Click the image of the enclosure to view details or to run necessary fix procedures.

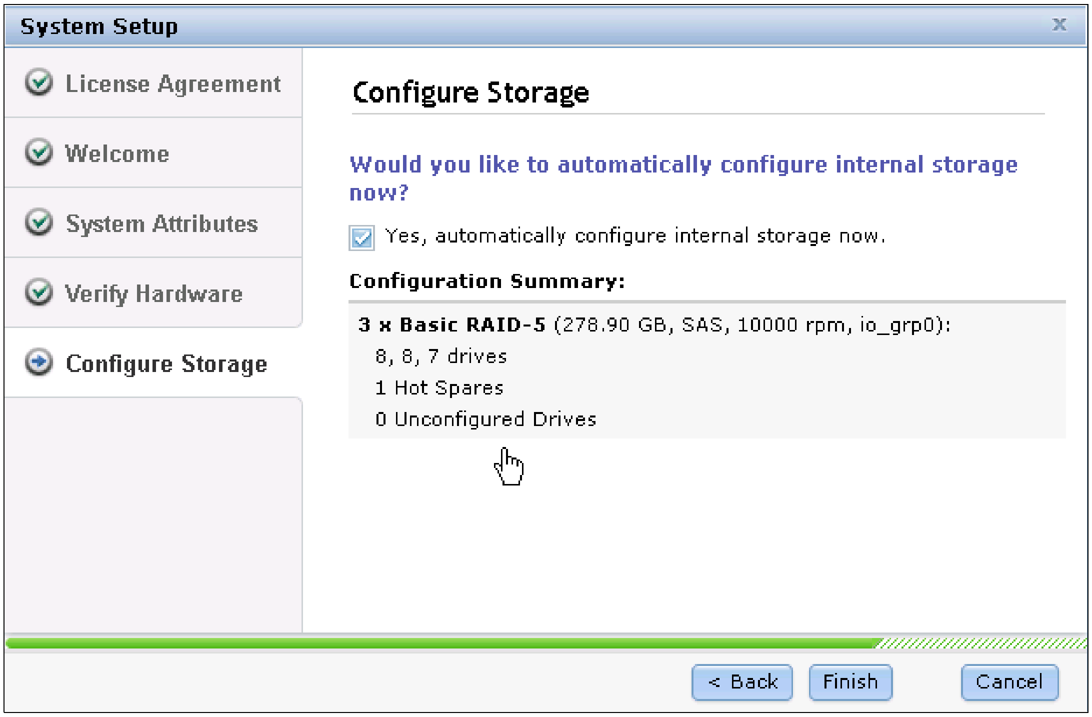
Storage Found:
(24 drives) 278.90 GB, SAS, 10000 rpm, io_grp0

Figure 11-14 Verify hardware

2. Proceed with setup by clicking **Next**.

Configure storage

The Configure Storage window is where you can configure your internal storage arrays. The system automatically detects the drives that are attached to it. These are configured in RAID arrays. You see a menu like the example in Figure 11-15 on page 167.



Configure Storage

Would you like to automatically configure internal storage now?

☑ Yes, automatically configure internal storage now.

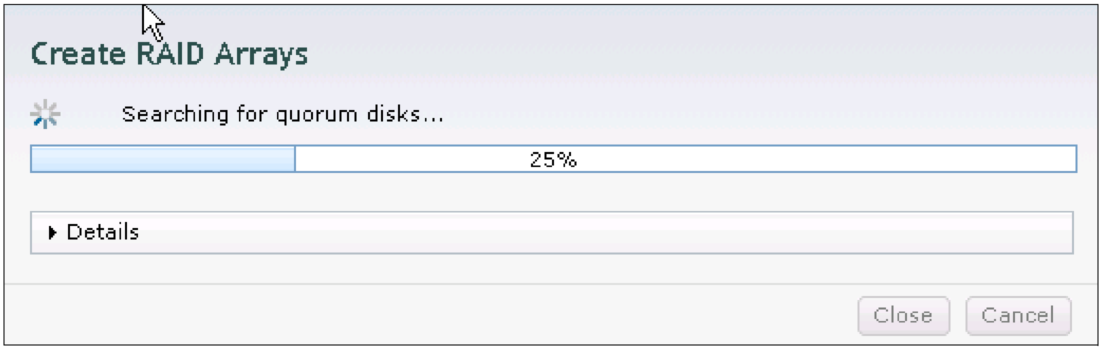
Configuration Summary:

3 x Basic RAID-5 (278.90 GB, SAS, 10000 rpm, io_grp0):
8, 8, 7 drives
1 Hot Spares
0 Unconfigured Drives

Figure 11-15 Configure storage

1. Click **Finish**.

   You can see which task is running by the display. Figure 11-16 shows an example.



Create RAID Arrays

Searching for quorum disks...

25%

Figure 11-16 Creating RAID arrays

Notification is given by an Information dialog window like the example in Figure 11-19 on page 169.

Use the mouse to hover over each component to display more detail, as shown in Figure 11-17 on page 168.

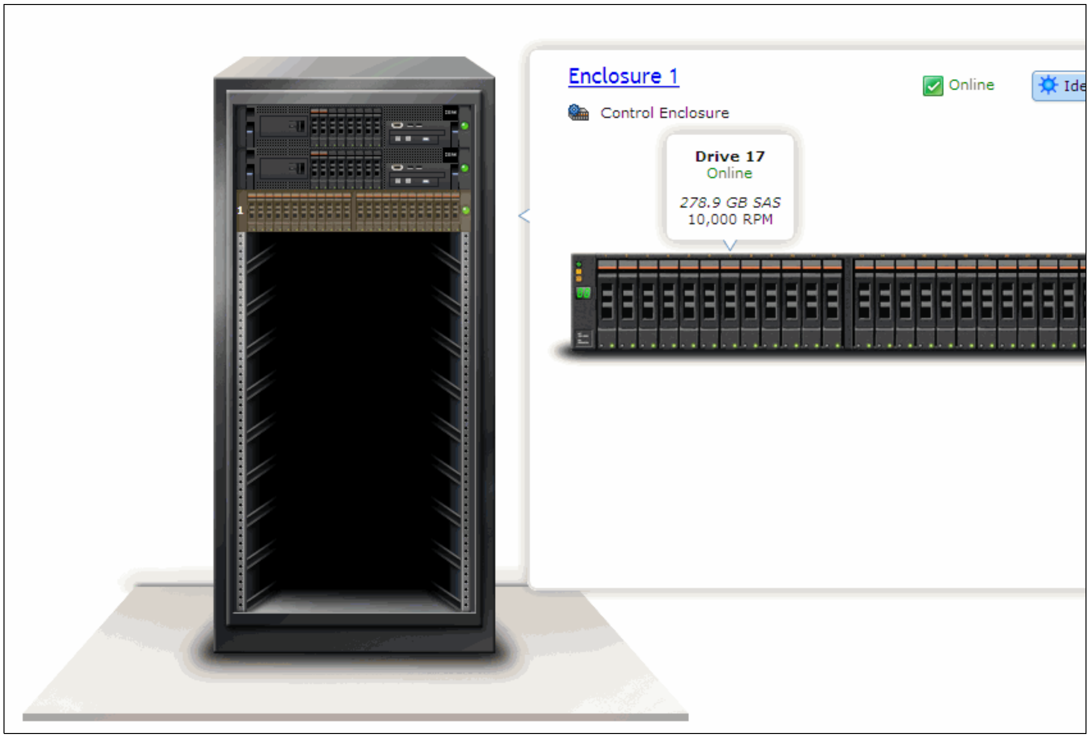


Figure 11-17   Hardware detail

If problems are found, attempt to resolve them now by using the options available on this window. Ensure that all expansion enclosures are included in the graphic. If components are missing, ensure that they are powered up and the cables are connected correctly.

2. Click **Next** to continue.

   A task status window is displayed.

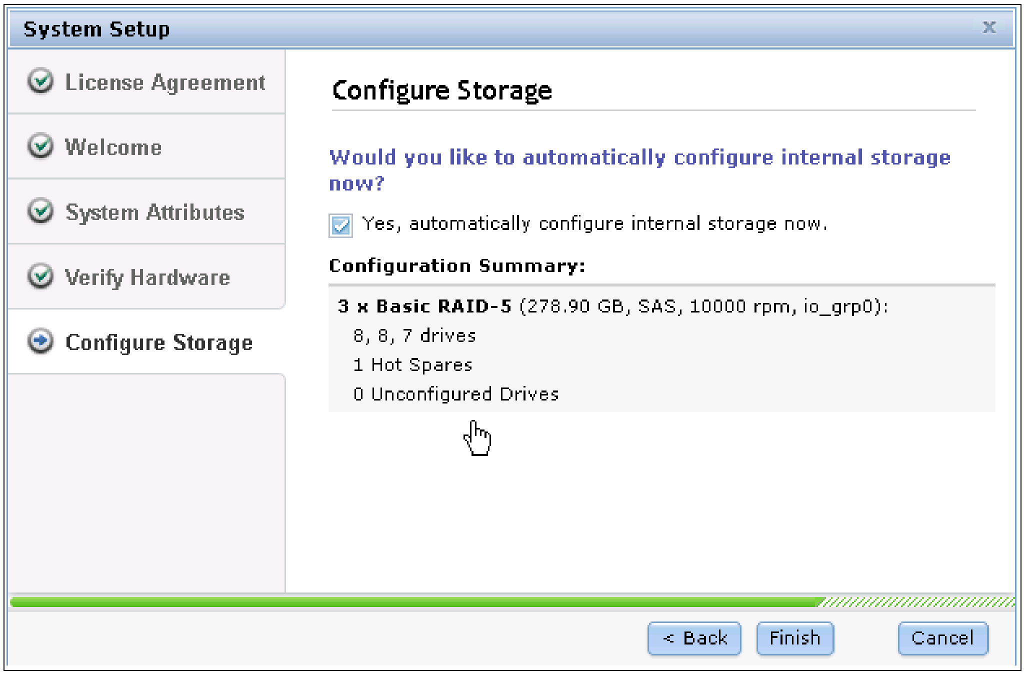
3. Wait for *Task Complete*, and then close.

Configure storage

After the storage is added, you are asked to configure it. The system will suggest a preferred configuration, which is shown on the next window.

1. If the default configuration is acceptable, click the **Yes** check box to automatically configure the storage, as shown in the example in Figure 11-18 on page 169. Otherwise, clear the box to skip auto-configuration. You must manually configure the storage later.

> **Note:** The Storwize V7000 Unified file module and IBM Scale Out Network Attached Storage (SONAS) give the best performance when you use arrays with eight data spindles because of the IBM General Parallel File System (GPFS) block size. Therefore, the automatic configuration might suggest 8+P arrays. Where possible, use 8+P or 4+P arrays. For more information, see the SONAS documentation.
>
> In a 24-drive enclosure, this conveniently gives an 8+P, 8+P, 4+P, S configuration.



Configure Storage

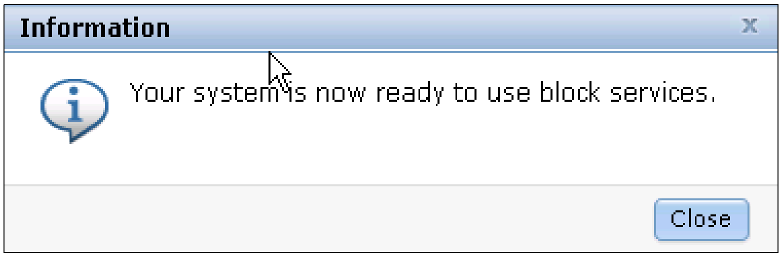
Would you like to automatically configure internal storage now?

☑ Yes, automatically configure internal storage now.

Configuration Summary:

3 x Basic RAID-5 (278.90 GB, SAS, 10000 rpm, io_grp0):
8, 8, 7 drives
1 Hot Spares
0 Unconfigured Drives

Figure 11-18 Automatically configure internal storage



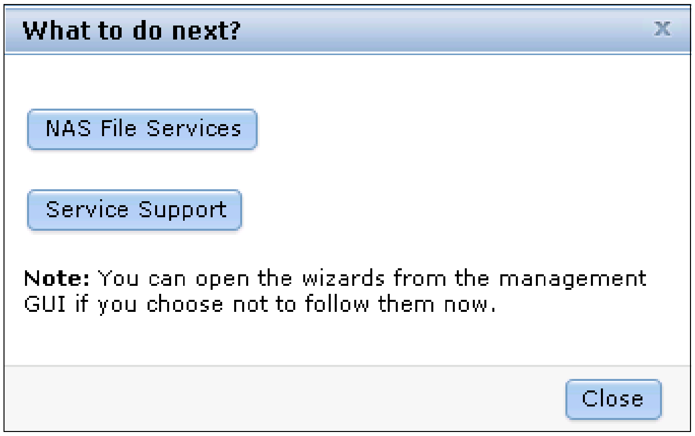
Your system is now ready to use block services.

Figure 11-19 Notification that Easy Setup is complete

2. After the System Setup wizard finishes and you click **Close**, you are prompted to complete the Support Services wizard.

What to do next

When you click Close upon completion of the Easy Setup wizard, you see pop-up window that asks what you want to do next. As Figure 11-20 on page 170 shows, the options are to configure NAS file services or configure service support.



Note: You can open the wizards from the management GUI if you choose not to follow them now.

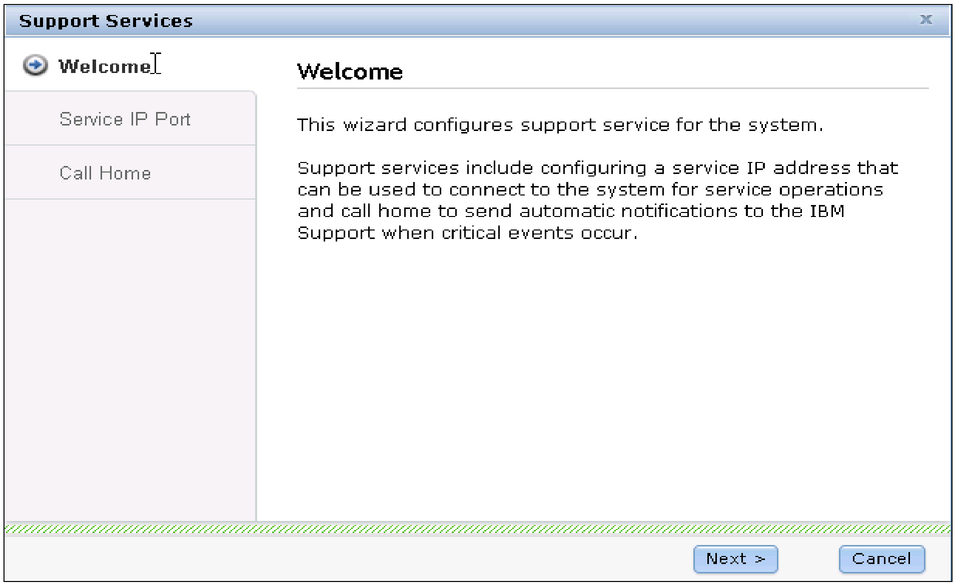
Figure 11-20 What to do next? window with options

If you click **Close**, you are given a message about the Call Home feature, and the Call Home Warning dialog window opens. You can bypass that now.

1. Start with Service Support setup.
2. To set values, click **Service Support**.

Support Services notifications

The next window, shown in Figure 11-21, gives the option to configure the support notifications.



Welcome

This wizard configures support service for the system.

Support services include configuring a service IP address that can be used to connect to the system for service operations and call home to send automatic notifications to the IBM Support when critical events occur.

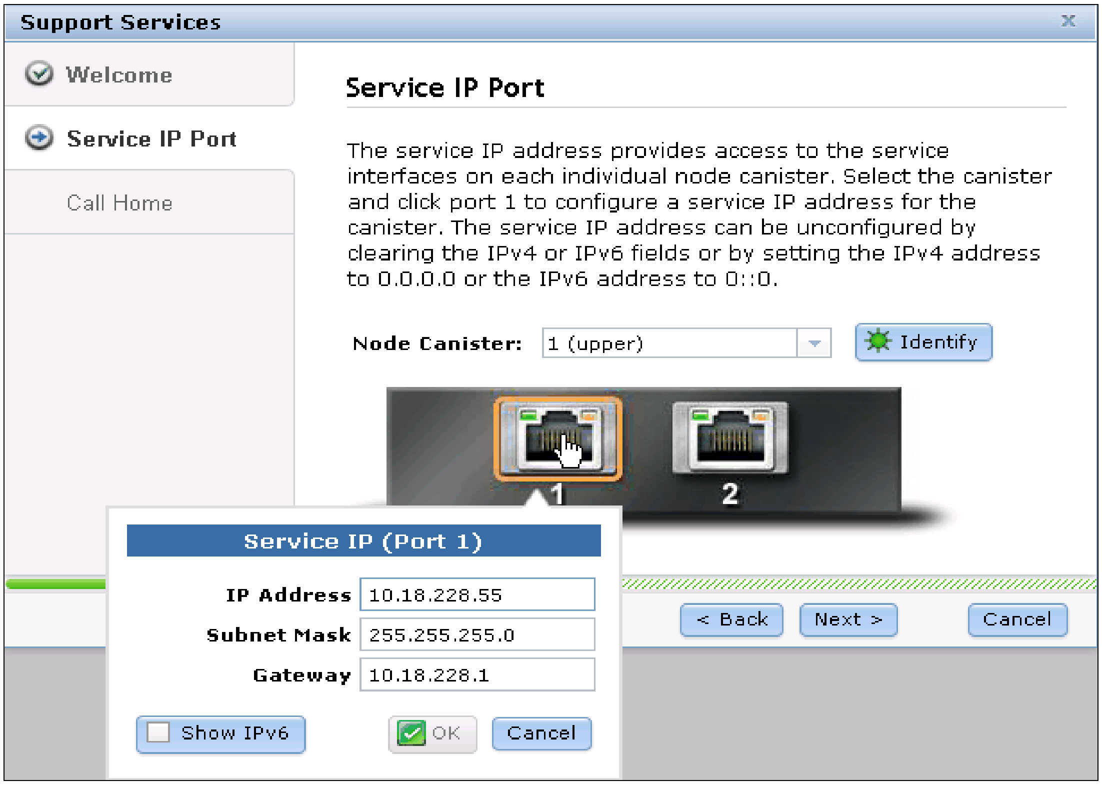
Figure 11-21 Support Services wizard Welcome window

This menu goes through setting up the Service IP Port and Call Home sections.

Click **Next** to start configuring these sections.

Service IP Port

The next window asks for your service IP address information for each canister, as shown in Figure 11-22 on page 171.



Support Services

Service IP Port

The service IP address provides access to the service interfaces on each individual node canister. Select the canister and click port 1 to configure a service IP address for the canister. The service IP address can be unconfigured by clearing the IPv4 or IPv6 fields or by setting the IPv4 address to 0.0.0.0 or the IPv6 address to 0::0.

Node Canister: 1 (upper)

Service IP (Port 1)

IP Address 10.18.228.55

Subnet Mask 255.255.255.0

Gateway 10.18.228.1

Figure 11-22 Service IP information

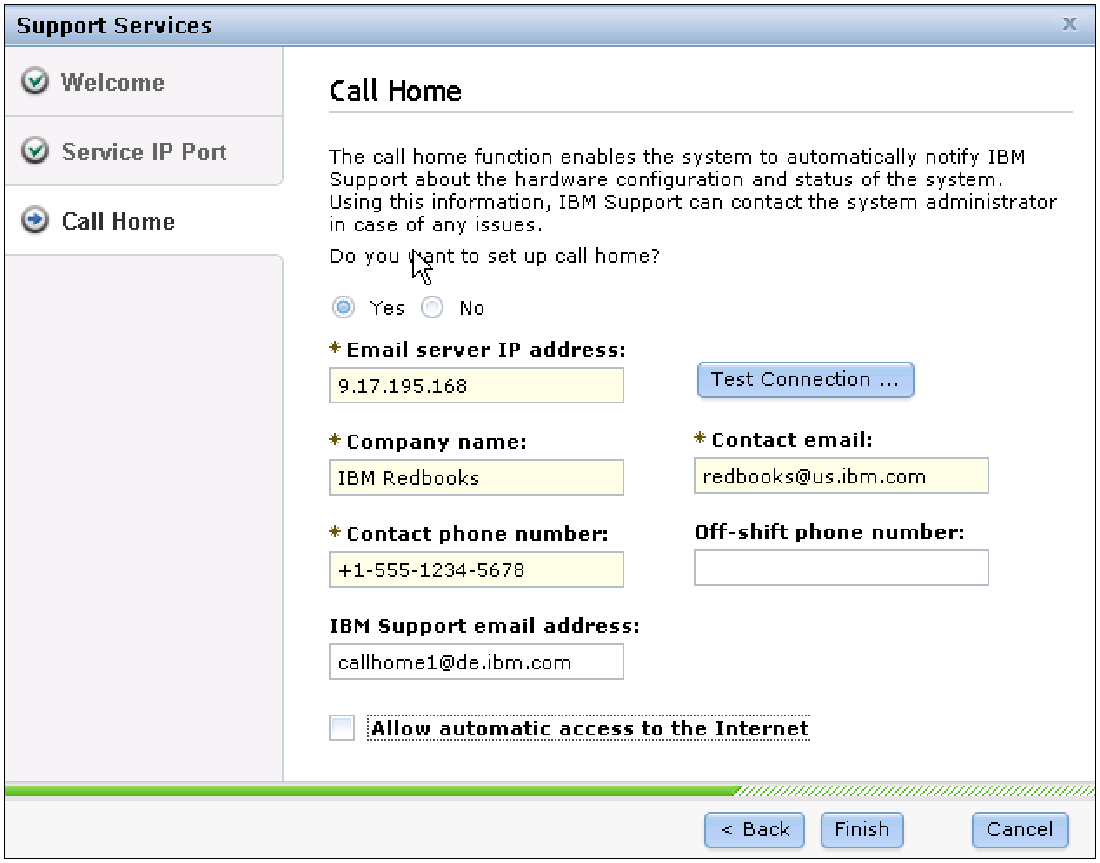
After this is complete for each node canister, you can start configuring the Call Home information.

Click **Next**.

Call Home feature

To configure the Call Home information, you see the window that is displayed in Figure 11-23 on page 172. It is necessary to configure the Call Home feature to allow IBM to be made aware of any hardware configuration issues and the overall health of the system

1. Fill in the required information.



Call Home

The call home function enables the system to automatically notify IBM Support about the hardware configuration and status of the system. Using this information, IBM Support can contact the system administrator in case of any issues.

Do you want to set up call home?

● Yes ○ No

*Email server IP address: 9.17.195.168

*Company name: IBM Redbooks

*Contact email: redbooks@us.ibm.com

*Contact phone number: +1-555-1234-5678

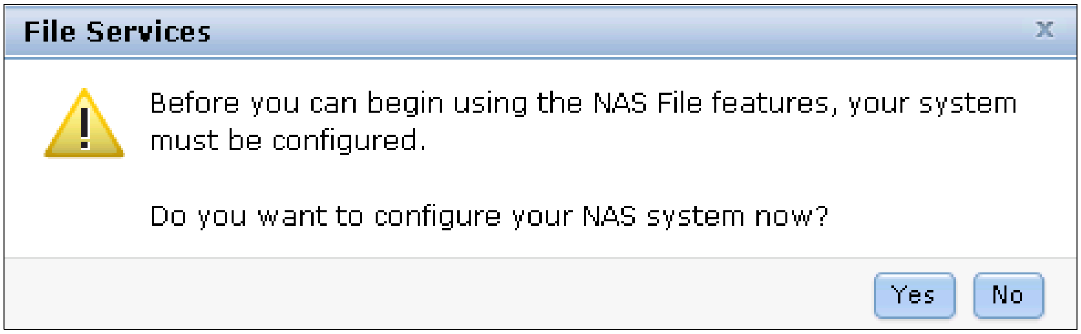
Off-shift phone number:

IBM Support email address: callhome1@de.ibm.com

□ Allow automatic access to the Internet

Figure 11-23 Call Home

2. After you have entered all of the information that will allow the machine to call home, click **Finish**.
3. The first field is the IP address of your email server, which needs to be accessible to the cluster and allows the cluster to send email.
4. Also, enter your company name, contact person's email address, and primary shift telephone number. The off-shift number is optional and required only if the primary number is not one that is answer 24 hours a day. These numbers must be phones that are answered at any time so that IBM Support personnel can contact you if a cluster calls home.
5. Leave the last field set as the default. This is the IBM address for Call Home alerts.
6. Click **Finish**.
7. To use the NAS part of the unified system, you must configure that side of it. When you choose any option under the NAS file folder, you are prompted to first configure with the dialog window that the example Figure 11-24 on page 173 shows.


File Services

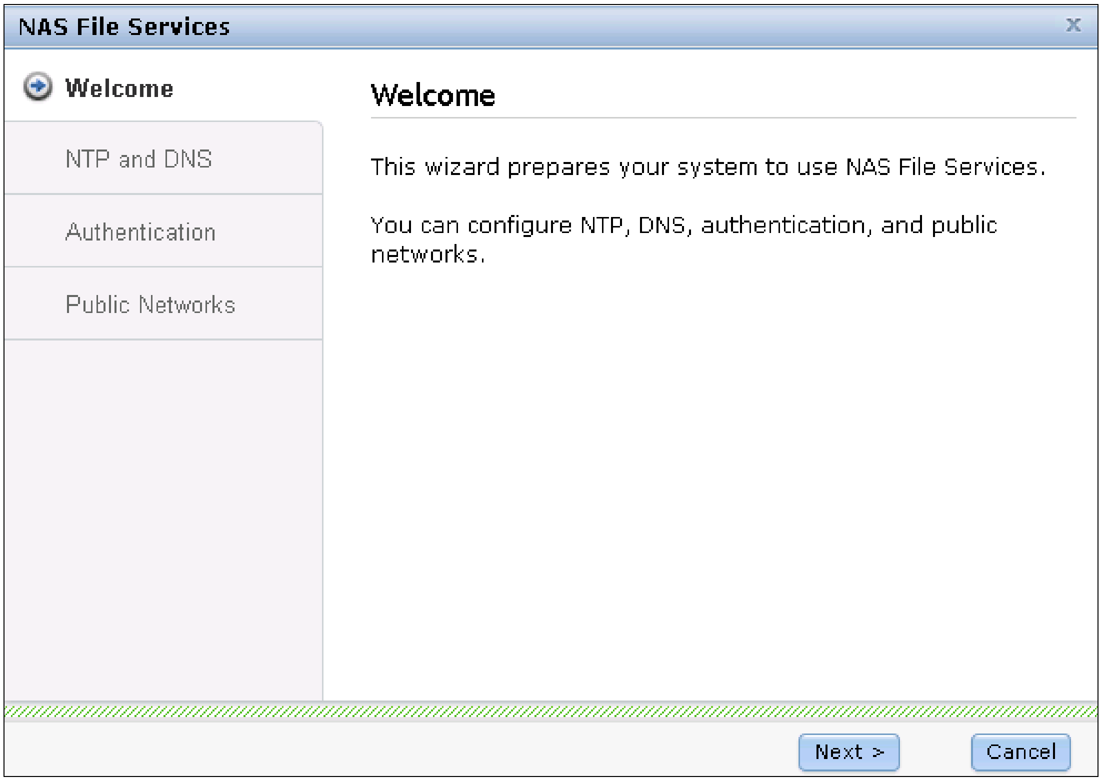
Before you can begin using the NAS File features, your system must be configured.

Do you want to configure your NAS system now?

Figure 11-24 Configure NAS

8. Click **Yes**.

   The Welcome screen for configuring the NAS File Service section then opens, as shown in example Figure 11-25.


NAS File Services

Welcome

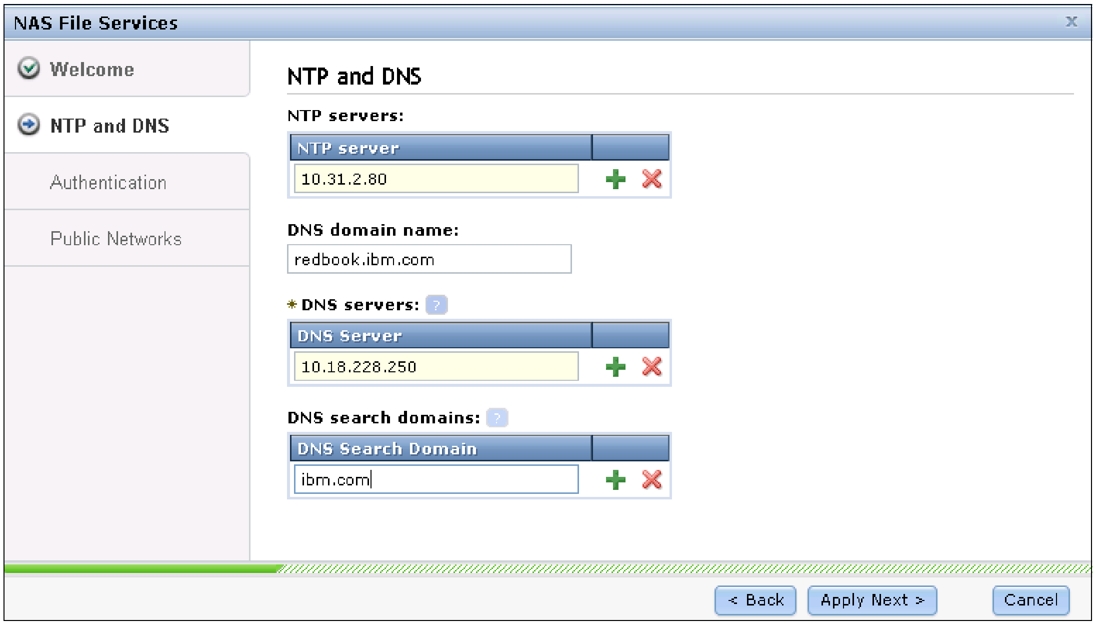
This wizard prepares your system to use NAS File Services.

You can configure NTP, DNS, authentication, and public networks.

Figure 11-25 Welcome NAS

   In the subsections that follow, we describe configuring the NTP, DNS, authentication, and public networks.

9. Click **Next** to continue.



NAS File Services

NTP and DNS

NTP servers:

NTP server
10.31.2.80

DNS domain name: redbook.ibm.com

* DNS servers:

DNS Server
10.18.228.250

DNS search domains:

DNS Search Domain
ibm.com

Figure 11-26

Network Time Protocol server

A Network Time Protocol (NTP) must be defined in the cluster to ensure consistent clocks in the file modules. This is needed for recovery and for resolving deadlocks.

Or, use the command-line interface, or CLI (`separate addresses with comma`):

```
setnwntp ip,xxx.xxx.xxx.xxx,xxx.xxx.xxx.xxx
lsnwntp
rmnwntp
```

Domain name service

1. Next, enter your domain name information. The process tests that the servers listed are present, and it fails if they cannot be contacted.

   Domain name — This is the public network domain, which is appended to your cluster name. This is typically common to your whole enterprise. For example, `customer.com`.

   DNS servers — IP address of your DNS server. To add a server IP to the list, click the **+** symbol, and use the X to delete an entry. Add as many DNS server entries as you want. At least one is required.

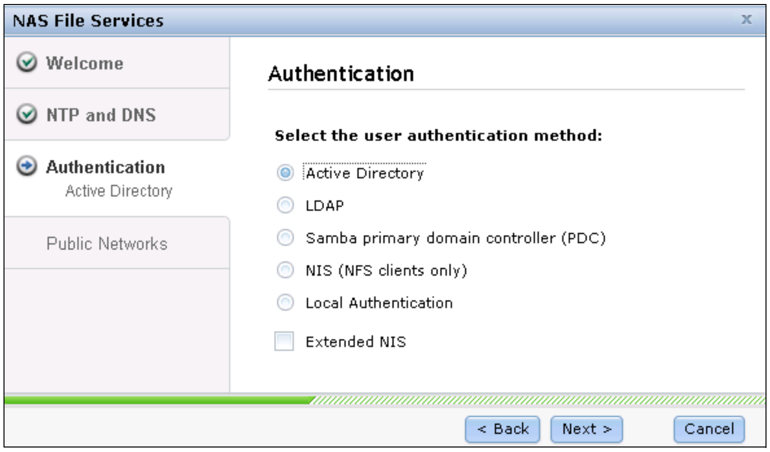
   DNS search domains — (Optional) Additional domain names to search.

2. When complete, click **Next**.
3. A progress panel displays. Wait for the Task Completed notice, and then close.

Authentication

In the next window (shown in Figure 11-27 on page 175), you can set the method that will be used for file access authentication and control. This is optional, so this step can be deferred or skipped. However, no file access is possible until this section is configured.

(Optional) Choose from the available authentication methods shown on Figure 11-27 on page 175, by clicking the appropriate selection, and then click **Next**.



Authentication

Select the user authentication method:

● Active Directory
○ LDAP
○ Samba primary domain controller (PDC)
○ NIS (NFS clients only)
○ Local Authentication

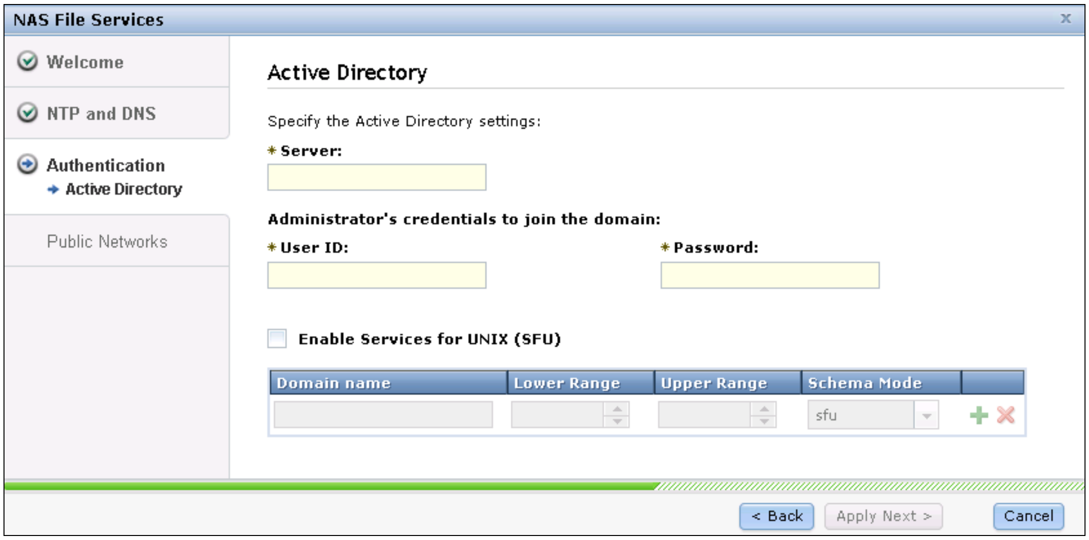
□ Extended NIS

Figure 11-27 Authentication window

Only one form of authentication can be active. Depending on your choice, you are taken through a series of setup panels as follows.

Active Directory

1. If you chose Active Directory, enter the information in the fields shown in Figure 11-28.



Active Directory

Specify the Active Directory settings:

* Server:

Administrator's credentials to join the domain:

* User ID:

* Password:

□ Enable Services for UNIX (SFU)

Domain name	Lower Range	Upper Range	Schema Mode
			sfu

Figure 11-28 Active Directory settings

Active Directory server This is the IP address of the Active Directory server.

User ID and password User ID and password combination that has sufficient authority to connect to the Active Directory server and access authentication and mapping information. Typically, this is the Administrator user ID, or an ID with equivalent authority.

Enable SFU If support for UNIX is required, enable the Services for UNIX (SFU) feature and complete the configuration that follows.

Domain name This is the domain that SFU belongs to.

UID and GID range Lower to upper limits of the range of user and group IDs that are used by the Active Directory server.

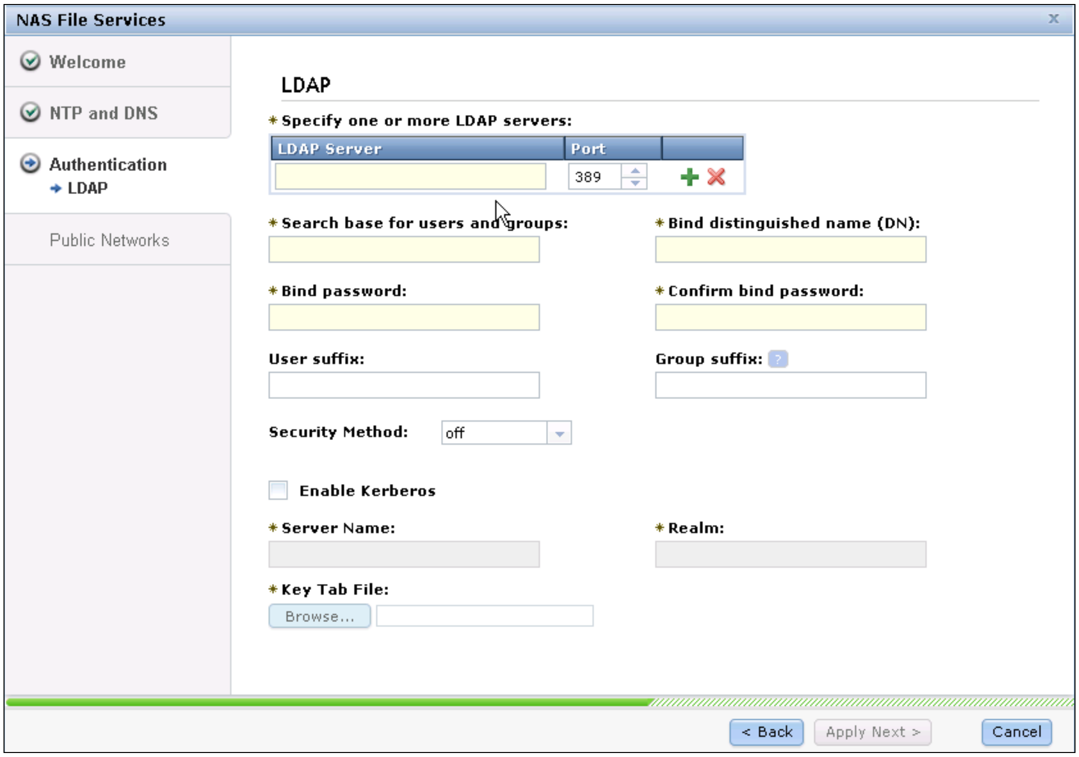
SFU schema The SFU schema mode that is being used.

Tip: Use the **+** symbol to add as many line items as required.

2. When you are ready, click **Finish**. A progress window opens while the configuration process runs.
3. Wait for Task Completed notice, and then close the window.

Lightweight Directory Access Protocol

For Lightweight Directory Access Protocol (LDAP) settings, you see the window shown in Figure 11-29.



NAS File Services

LDAP

* Specify one or more LDAP servers:

LDAP Server	Port
	389

* Search base for users and groups:

* Bind distinguished name (DN):

* Bind password:

* Confirm bind password:

User suffix:

Group suffix:

Security Method: off

□ Enable Kerberos

* Server Name:

* Realm:

* Key Tab File:

Figure 11-29 LDAP settings

Enter the required information:

LDAP server — This is the IP address of the LDAP server. Click the + symbol if you want to add extra servers.

Search base — The base domain suffix.

DN — The root distinguished name.

Bind password — User ID and password that is required to access the LDAP server.

User, Group suffix — User and Group suffix as defined by the LDAP server.

Security method — Select the Secure Sockets Layer (SSL) mode that is used. If SSL security is used, a certificate file is needed. When this option is selected, a new box is displayed for the Certificate. Click Browse to locate the certificate file on your workstation.

Enable Kerberos — If SSL is not used, then the option for Kerberos is displayed. Tick the box to enable.

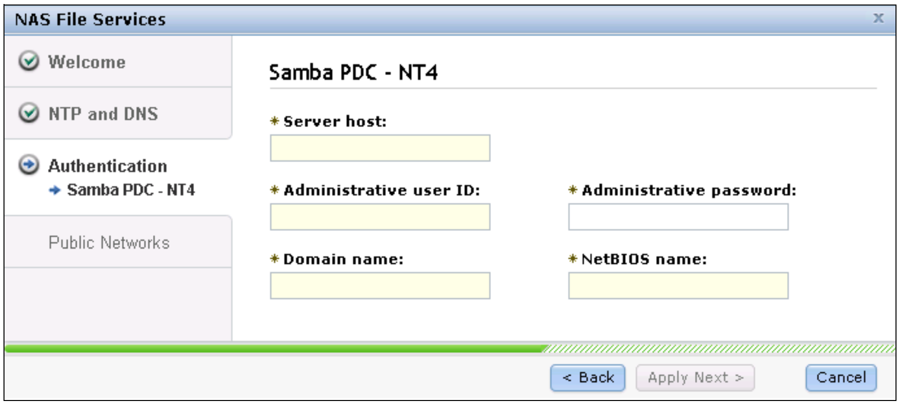
Kerberos name — Enter the name of the server.

Kerberos realm — Enter the Kerberos realm.

Key tab file Browse to the location of the Kerberos key tab file.

Samba primary domain controller

If Samba PDC was selected, the Samba PDC - NT4 configuration window opens, as shown in Figure 11-30.



Samba PDC - NT4

* Server host:

* Administrative user ID:

* Administrative password:

* Domain name:

* NetBIOS name:

Figure 11-30 PDC settings

Enter the required information as follows:

Server host This is the IP address of the NT4 PDC server.

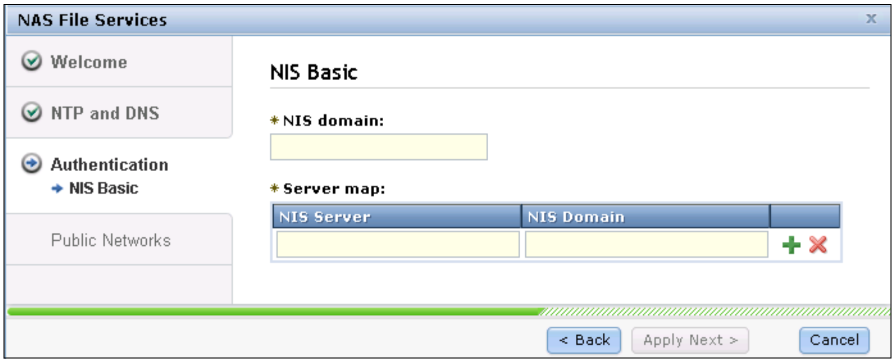
Admin ID, password The user ID and password that is used to access the NT4 server that has administrative authority.

Domain name The NT4 Domain name.

NetBios name NT4 NetBIOS name.

NIS Basic

If you are using basic NIS authentication, you see the setup window shown in Figure 11-31.



NIS Basic

* NIS domain:

* Server map:

NIS Server	NIS Domain	
		+ ✗

Figure 11-31 NIS Settings

The following definitions apply to NIS basic settings:

NIS domain The name of the primary NIS domain.

NIS Server/Domain The address of each server. For each server, map the supported domains. Domains are entered as a comma-separated list for each server. Use the **+** symbol to add extra servers.

Extended NIS

For Active Directory and PDC authentication, there is an option to include extended support for NIS. If this was selected, you are also presented with a configuration settings window, as shown in Figure 11-32.

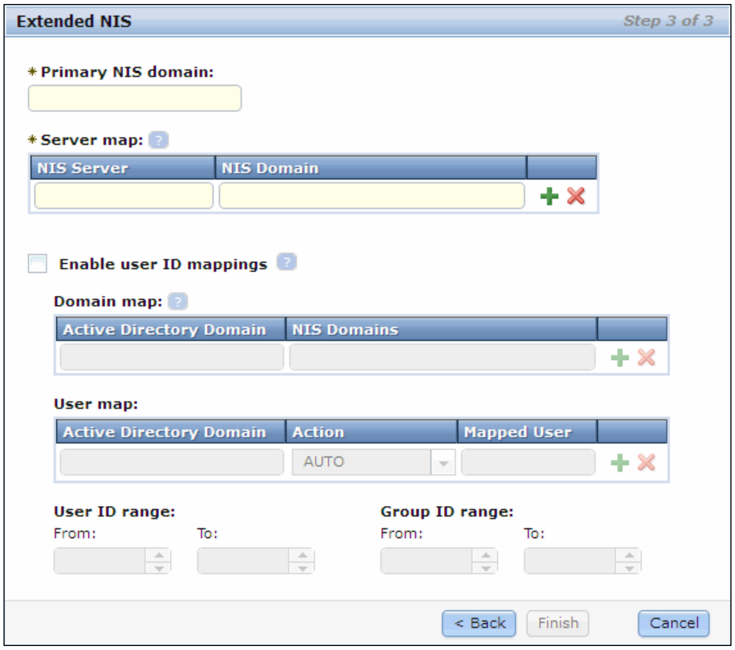


Figure 11-32 Extended NIS settings

The following definitions apply to extended NIS settings:

Primary domain The name of the primary NIS domain.

NIS Server/Domain The address of each server. For each server, map the supported domains. Domains are entered as a comma-separated list for each server. Use the **+** symbol to add more servers.

If NIS is used for user ID mapping, click the check box to enable it and complete the remaining fields. If not, this panel is complete.

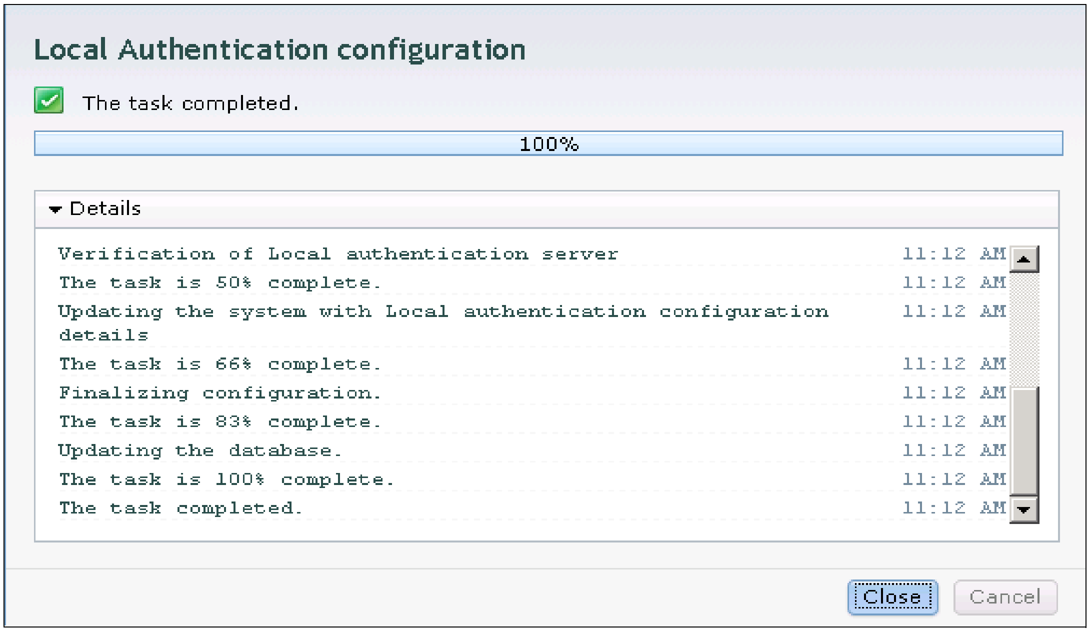
Domain map Add entries here to map the Active Directory domain to the NIS domains. Use the **+** symbol to add lines.

User map Add entries for user mapping exceptions as required.

Local Authentication

Select the Local Authentication option to configure the system with an internal authentication mechanism in which users and groups are defined locally on this system. If this was selected, you are also presented with the window as shown in Figure 11-33 on page 179.

> **Note:** With Local Authentication, you enter user and group information locally. That is covered in Section 11.13.3, “Create local users by using local authentication for NAS access” on page 200.



Local Authentication configuration

The task completed.

100%

Details

```
Verification of Local authentication server                    11:12 AM
The task is 50% complete.                                      11:12 AM
Updating the system with Local authentication configuration    11:12 AM
details
The task is 66% complete.                                      11:12 AM
Finalizing configuration.                                      11:12 AM
The task is 83% complete.                                      11:12 AM
Updating the database.                                         11:12 AM
The task is 100% complete.                                     11:12 AM
The task completed.                                            11:12 AM
```

Figure 11-33 Local Authentication configuration window

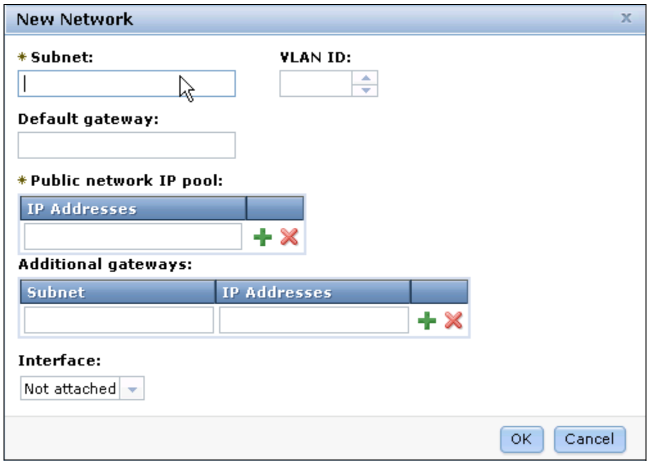
Completion

When the chosen method is configured and the processing is complete, as indicated by Task Completed, close the status window.

Public networks

You are now presented with a window to define the public networks, as shown in It is necessary to configure several networks because each network is tied to an Ethernet logical interface.

1. To create a network, click **New Network**, which opens the window shown in Figure 11-34.



New Network

* Subnet:

VLAN ID:

Default gateway:

* Public network IP pool:

IP Addresses	

Additional gateways:

Subnet	IP Addresses	

Interface: Not attached

Figure 11-34 New Network window

The following definitions apply to the New Network fields:

Subnet
The IP subnet this definition is for. The format of this value is using the CIDR syntax. *xxx.xxx.xxx.xxx/yy*, where yy is the decimal mask that is given as the number of left-aligned bits in the mask.

VLAN ID
If VLANs are being used, enter the VLAN number here; otherwise, leave the field blank. Valid values are 2-4095. VLAN 1 is not supported for security reasons.

Default gateway
The default gateway (or router) within this subnet for routing. This is not a required field if all devices that need to be connected to this interface are in the same subnet.

Interface pool
Using the + key, add IP addresses to the pool for use on this logical interface. These must be in the same subnet as entered above. A minimum of one address is required.

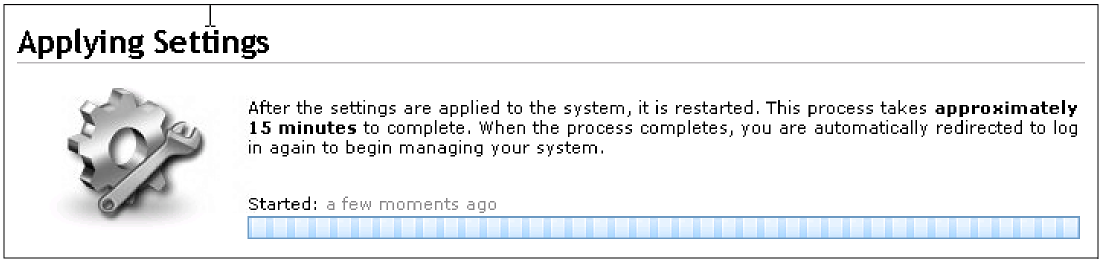
Additional gateways
If more than one gateway (router) exists in the subnet, add the IP addresses here.

Interface
Select the logical interface that this definition is assigned to.

2. A progress window displays. Wait for the completion message, and click **OK** when complete.
3. Repeat the process for each network by using the **New Network** selection until all required networks are defined. When you are done, click **Finish**.

   A reboot progress window now opens, indicating that the file modules are being restarted.
4. When that process is complete, close the window.

This is followed by the applying settings status window that is shown in Figure 11-35.



Applying Settings

After the settings are applied to the system, it is restarted. This process takes **approximately 15 minutes** to complete. When the process completes, you are automatically redirected to log in again to begin managing your system.

Started: a few moments ago

Figure 11-35 Applying settings

When that process finishes, Easy Setup configuration is complete. The home page for the Storwize V7000 Unified system is automatically displayed.

11.7.3 Set up periodic configuration backup

Management and configuration information is stored on the file modules in a trivial database (TDB). It is recommended that you set up a periodic backup of the TDB at a suitable time, and we recommend daily. This backup might be required by service personnel if the TDB becomes lost or corrupted or in the event of a recovery.

1. Start an SSH session with the cluster management address that you set when you initialized the cluster.
2. Log on with user `admin` and password `admin`**.**
3. Issue the command to perform the periodic configuration backup, shown in Example 11-1 on page 181.

Example 11-1

```
[7802378.ibm]$
[7802378.ibm]$ mktask BackupTDB --minute 0 --hour 2 --dayOfWeek "*"
EFSSG0019I The task BackupTDB has been successfully created.
EFSSG1000I The command completed successfully.
[7802378.ibm]$
```

If you receive an error that the management service is stopped, wait a few minutes for it to complete its startup from the recent reboot.

4. This command schedules a backup to run at 2:00 a.m. every day. You can change the time to suit your own environment.
5. Exit the session.

11.8 Manual setup and configuration changes

If you are redeploying an existing configuration, or if you skipped any steps in Easy Setup, or if you need to alter any values, use the following procedures to manually set or alter these configuration settings.

If Easy Setup was completed and all values are entered correctly, this section is merely for your reference and you can skip some of the topics. We suggest that you set up the optional support details to enable remote support functions for the cluster, as described in this section.

11.8.1 System names

If you need to change the cluster name, use the following CLI command:

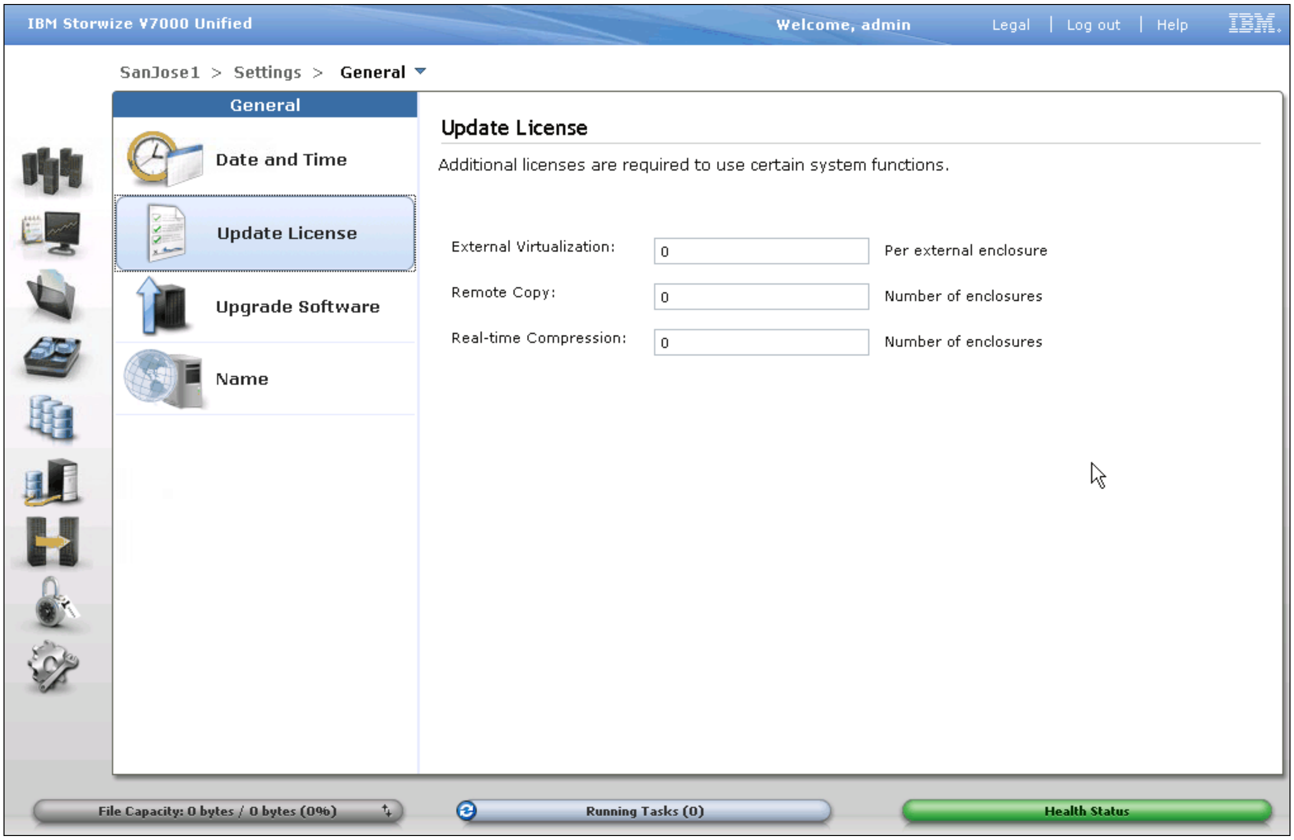
```
chsystem -name new_name
```

11.8.2 System licenses

If the system licenses need to be changed, select the window that is shown in Figure 11-36 on page 182. Select **Settings** → **General**, and then select **Update License**. Type over the values with the new ones to match the enclosure licenses that you are applying.

Although warning messages might be posted if the entered licenses are exceeded, IBM works by an honor system for licensing. Enter the value of only the license that you purchased. See Figure 11-36 on page 182.

> **Note:** To use the compression function, you must install the optional IBM Real-time Compression license.



Update License

Additional licenses are required to use certain system functions.

External Virtualization: 0 Per external enclosure

Remote Copy: 0 Number of enclosures

Real-time Compression: 0 Number of enclosures

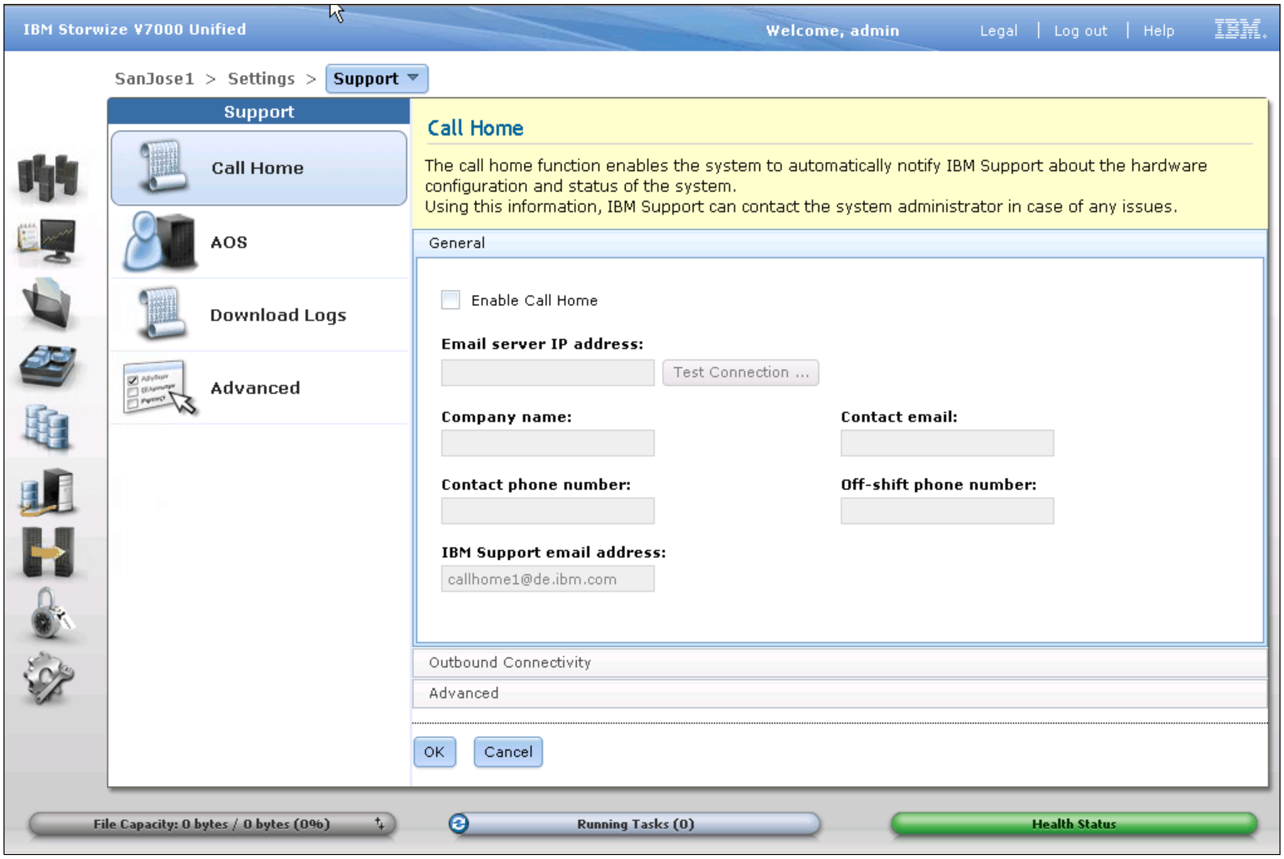
Figure 11-36 Update license

11.8.3 Support

Use the following panels to define or update details about remote support functions and to gather support information.

Call Home

To define Call Home details, go to the window that is shown in Figure 11-37 on page 183 by selecting **Settings** → **Support**. Then, click **Call Home**.



Call Home

The call home function enables the system to automatically notify IBM Support about the hardware configuration and status of the system.
Using this information, IBM Support can contact the system administrator in case of any issues.

□ Enable Call Home

Email server IP address:

Company name:

Contact email:

Contact phone number:

Off-shift phone number:

IBM Support email address: callhome1@de.ibm.com

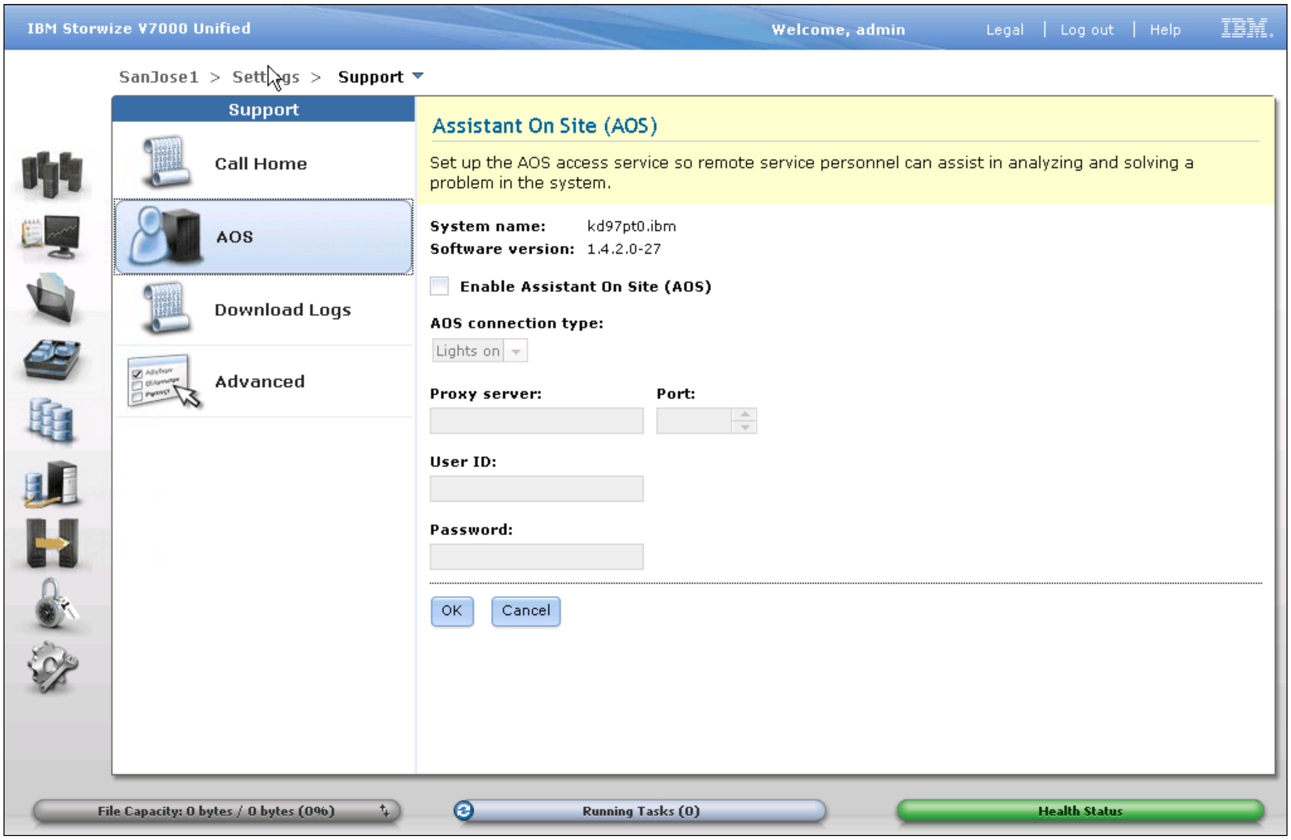
Figure 11-37 Call home definitions

Assist On-site (AOS)

Assist On-site is an IBM Tivoli software product that IBM Support staff members use to securely connect to the Storwize V7000 Unified system to determine the nature of problems and to perform actions, if necessary. This can be configured in two ways: lights on and lights out.

- *Lights on* requires manual authorization from the GUI onsite to allow a connection.
- *Lights out* connects immediately. This feature requires access to the Internet over HTTP and HTTPS protocols.

To configure, go to the setup window shown in Figure 11-38 on page 184, and select **Settings** → **Support**. Click **AOS**. To enable Assist On-site, click the check box and then select **Lights on** or **Lights off**. If your site uses a proxy for Internet access, enter the IP address and port details. Also, a user password is necessary. To disable the proxy, clear the Proxy server field.



Assistant On Site (AOS)

Set up the AOS access service so remote service personnel can assist in analyzing and solving a problem in the system.

System name: kd97pt0.ibm
Software version: 1.4.2.0-27

□ Enable Assistant On Site (AOS)

AOS connection type: Lights on

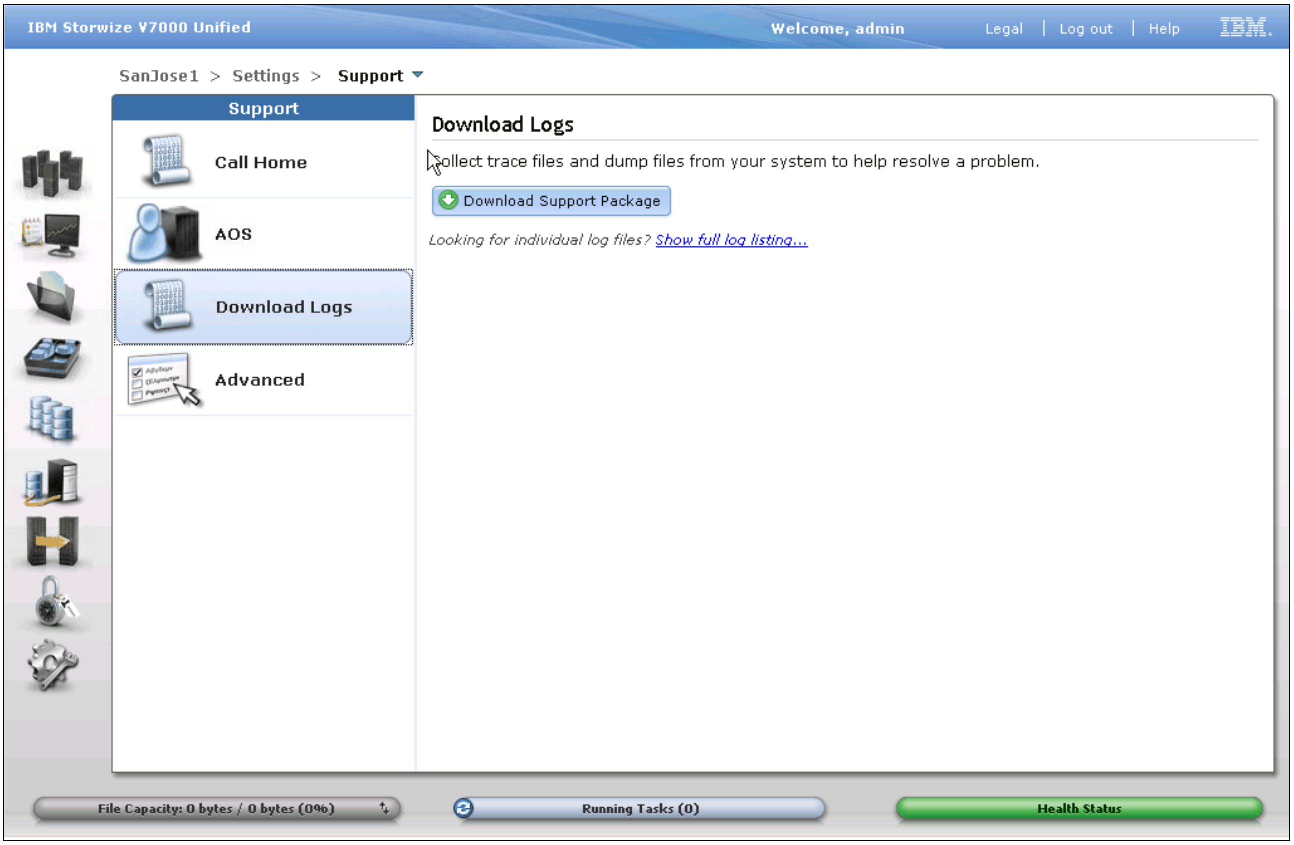
Proxy server:
Port:
User ID:
Password:

Figure 11-38 AOS configuration

Support logs

For most problems, IBM Support requests logs from the cluster. These are easy to produce in the **Download Logs** window that is shown in Figure 11-39 on page 185. Select **Settings** → **Support**. Then, click **Download Logs**. From this panel, you can view the support files by clicking **Show full log listing**. This displays the contents of the logs directory and includes previous data collects, dumps, and other support files. These files can be downloaded or deleted from this window.

To create and download a current support package, click **Download Support Package**. This opens a window with a message that asks for the type of package, as shown in Figure 11-39 on page 185. Select the log type that is requested by IBM Support. If there is any doubt, select all logs, but also consult IBM Support. Then, click **Download**. A progress window is displayed while the data is gathered. When complete, the window closes and a download option window is displayed. Use this window to download the file to the workstation that you are browsing from. The file can be retrieved any time later or by any user with access from the “Show full log listing” window.



Download Logs

Collect trace files and dump files from your system to help resolve a problem.

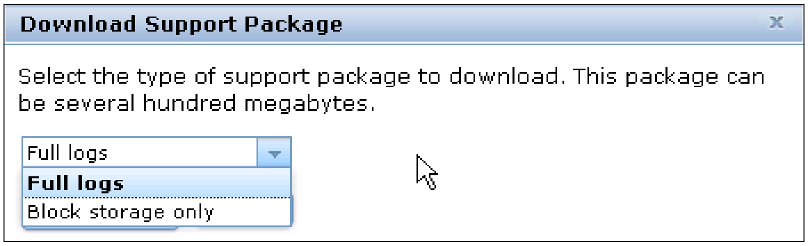
Download Support Package

Looking for individual log files? *Show full log listing...*

Figure 11-39 Support logs

To create and download a current support package, click **Download Support Package**. This opens a window that asks for the type of package, as shown in Figure 11-40. Select the log type requested by IBM Support. If there is any doubt, select all logs, but also consult IBM Support. Then, click **Download**. A progress window is displayed while the data is created and gathered. When complete, the window closes and a download option window is displayed. Use this window to download the file to the workstation you are browsing from.

Figure 11-40 shows the Download Support Package window.



Download Support Package

Select the type of support package to download. This package can be several hundred megabytes.

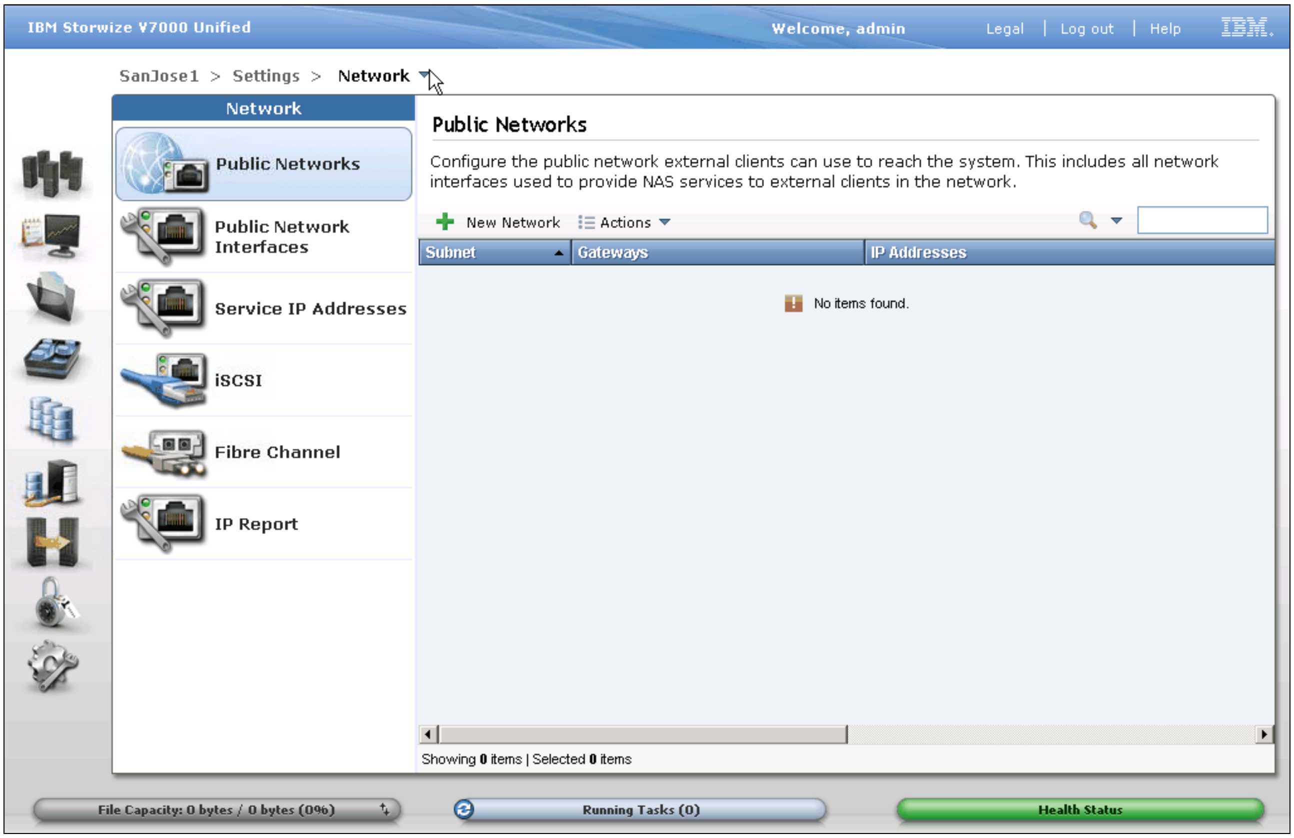
Figure 11-40 Download Support Package window

11.9 Network

There are several network adapters in the Storwize V7000 Unified system. Each is configured with IP addresses, and these can be changed as necessary. Here, we describe the various adapters and where to go to reconfigure them.

11.9.1 Public Networks

These addresses are on the 10 Gb Ethernet ports and the two client side 1 Gb ports. These are the addresses that the hosts and clients use to access the Storwize V7000 Unified system and to open file shares. Select **Settings** → **Network**. This option opens the window shown in Figure 11-41. Click **Public Networks**.



Public Networks

Configure the public network external clients can use to reach the system. This includes all network interfaces used to provide NAS services to external clients in the network.

Subnet	Gateways	IP Addresses

No items found.

Showing **0** items | Selected **0** items

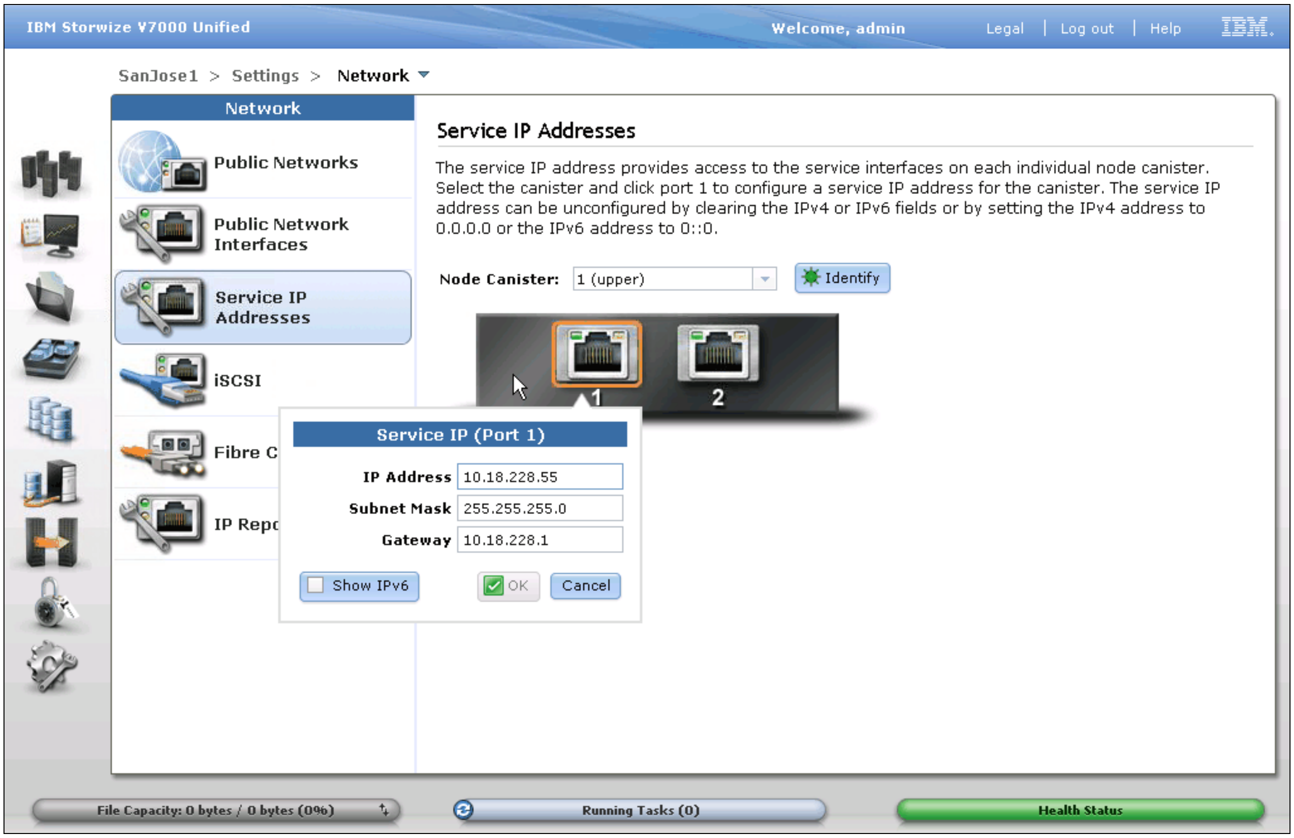
Figure 11-41 Public networks configuration

From here, you can add new network definitions and delete existing ones. Each definition is defined to one of the virtual adapters. See “Public networks” on page 179 for more information about configuring public networks.

11.9.2 Service ports

It is important that the service ports on the control enclosure are set to an IP address. These are seldom used in normal operation but are important when there is a problem, so it is better to have them set up beforehand. One IP address is needed for each node canister on its Port 1. This IP shares the port and coexists with the management IP that is presented on the same port, but only from the config node.

To set or alter these IP addresses, select **Settings** → **Network**. Click **Service IP Addresses** to display the configuration window, as shown in Figure 11-42 on page 187. Hover the mouse cursor over Port 1 to display the settings window. If you must change it, click in the fields and type the new values. Click **OK** to save. You must set both canisters. Therefore, use the pull-down menu to display the other canisters’ ports and configurations.



Service IP Addresses

The service IP address provides access to the service interfaces on each individual node canister. Select the canister and click port 1 to configure a service IP address for the canister. The service IP address can be unconfigured by clearing the IPv4 or IPv6 fields or by setting the IPv4 address to 0.0.0.0 or the IPv6 address to 0::0.

Node Canister: 1 (upper)

Service IP (Port 1)

IP Address 10.18.228.55
Subnet Mask 255.255.255.0
Gateway 10.18.228.1

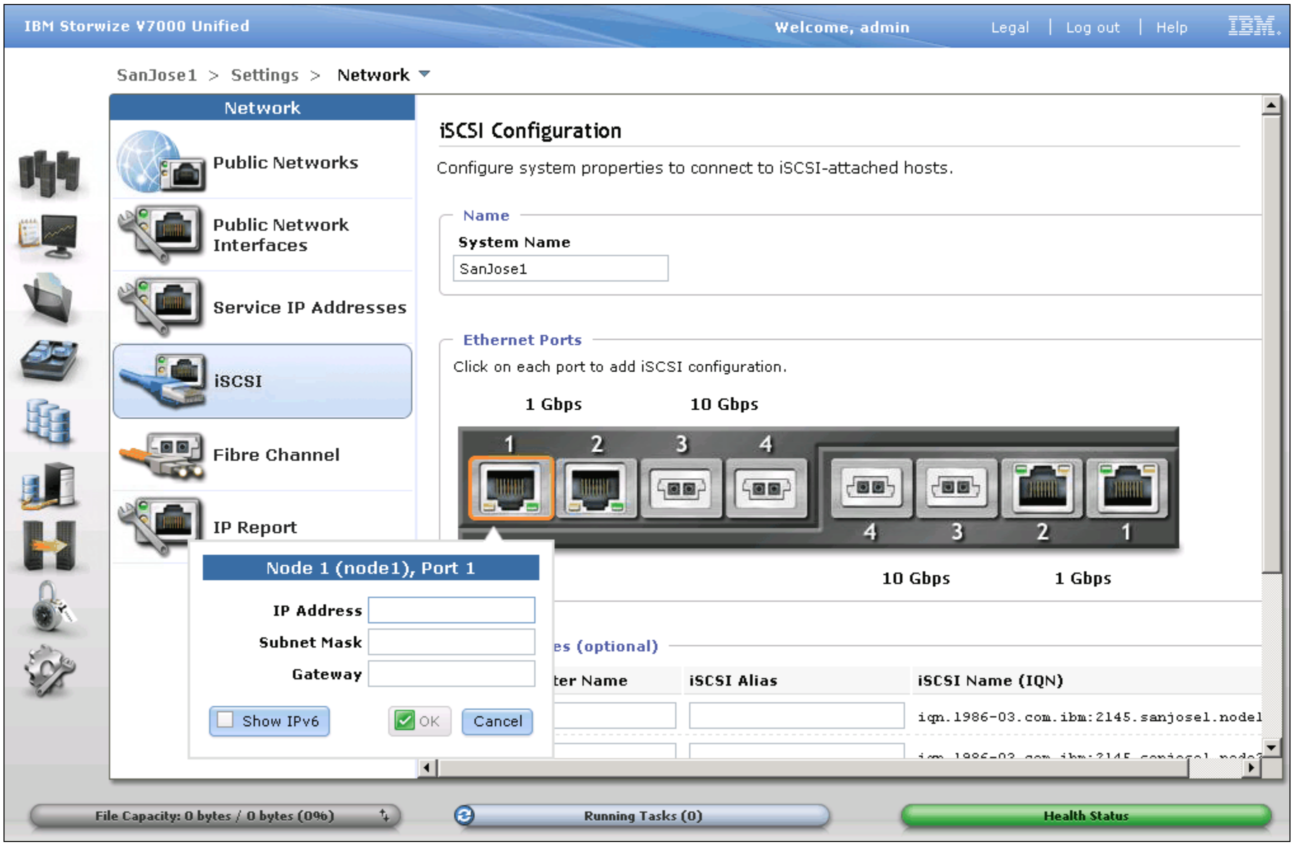
Figure 11-42 Service IP

11.9.3 Internet Small Computer System Interface

If you have the feature to add Ethernet ports to the control enclosure for Internet Small Computer System Interface (iSCSI) service, you must set the addresses for these ports. This is an optional feature and is used only for iSCSI connection to block storage.

Select **Settings** → **Network**, and then select **iSCSI** from the list. The iSCSI Configuration window that is shown in Figure 11-43 on page 188 opens.

Here, you see a graphical representation of the ports. There are two diagrams, one for each node canister. Hover over each port to display the configuration panel and enter the IP address details. Click **OK** to save.



iSCSI Configuration

Configure system properties to connect to iSCSI-attached hosts.

Name

System Name: SanJose1

Ethernet Ports

Click on each port to add iSCSI configuration.

Node 1 (node1), Port 1

IP Address:

Subnet Mask:

Gateway:

es (optional)

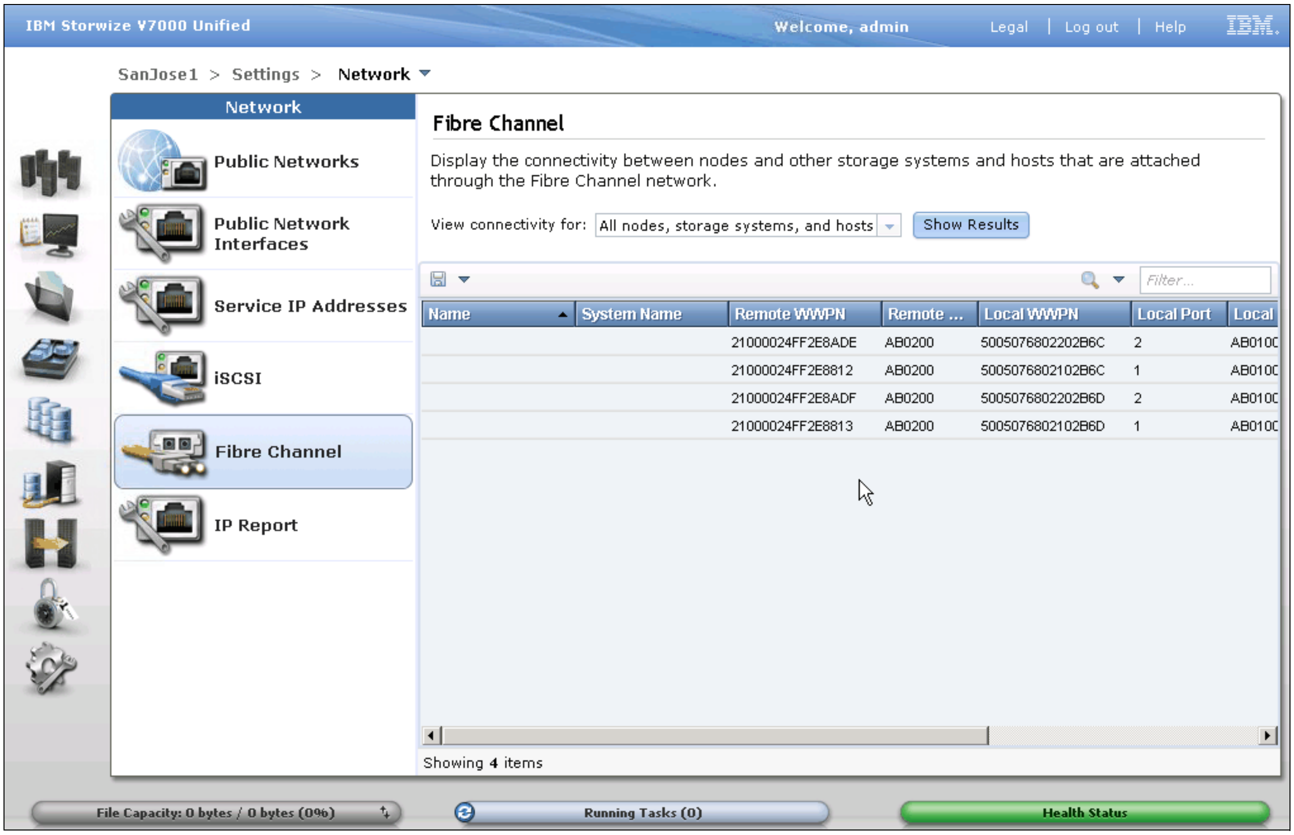
ter Name	iSCSI Alias	iSCSI Name (IQN)
		iqn.1986-03.com.ibm:2145.sanjose1.node1

Figure 11-43 iSCSI IP addresses

11.9.4 Fibre Channel ports

Use the Fibre Channel panel to display the Fibre Channel connectivity between nodes, storage systems, and hosts. The results for the ports for all nodes, storage systems, and hosts can be displayed. Go to **Settings** → **Network**. Select **Fibre Channel** from the list.

The display in Figure 11-44 on page 189 shows the results for all hosts that are attached to block volumes.



Fibre Channel

Display the connectivity between nodes and other storage systems and hosts that are attached through the Fibre Channel network.

View connectivity for: All nodes, storage systems, and hosts

Show Results

Name	System Name	Remote WWPN	Remote ...	Local WWPN	Local Port	Local
		21000024FF2E8ADE	AB0200	5005076802202B6C	2	AB010C
		21000024FF2E8812	AB0200	5005076802102B6C	1	AB010C
		21000024FF2E8ADF	AB0200	5005076802202B6D	2	AB010C
		21000024FF2E8813	AB0200	5005076802102B6D	1	AB010C

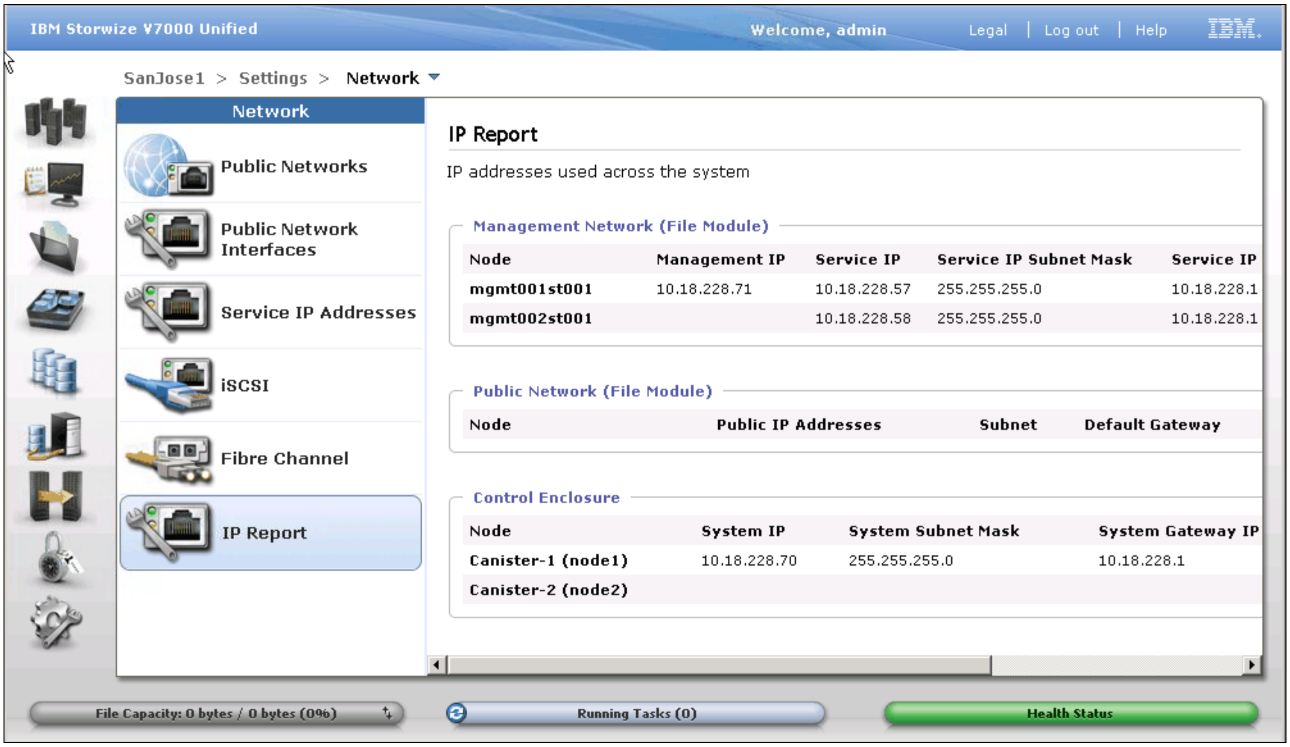
Showing 4 items

Figure 11-44 Shows Fibre Channel results for hosts

11.9.5 Fibre Channel ports

The IP Report panel displays all the IP addresses that are currently configured on the system. The File Module table provides details about all of the IP addresses that are currently configured to manage network-attached storage (NAS) services and public networks that clients use to connect to the system. File modules provide the services to access the file data from outside the system and provide the back-end storage and file system that store the file system data.

The control enclosure area shows details about all of the IP addresses related to managing block storage. The enclosure provides the services to access the block storage for block clients. Figure 11-45 on page 190 shows a sample display.



IP Report

IP addresses used across the system

Management Network (File Module)

Node	Management IP	Service IP	Service IP Subnet Mask	Service IP
mgmt001st001	10.18.228.71	10.18.228.57	255.255.255.0	10.18.228.1
mgmt002st001		10.18.228.58	255.255.255.0	10.18.228.1

Public Network (File Module)

Node	Public IP Addresses	Subnet	Default Gateway

Control Enclosure

Node	System IP	System Subnet Mask	System Gateway IP
Canister-1 (node1)	10.18.228.70	255.255.255.0	10.18.228.1
Canister-2 (node2)			

Figure 11-45 IP Report panel

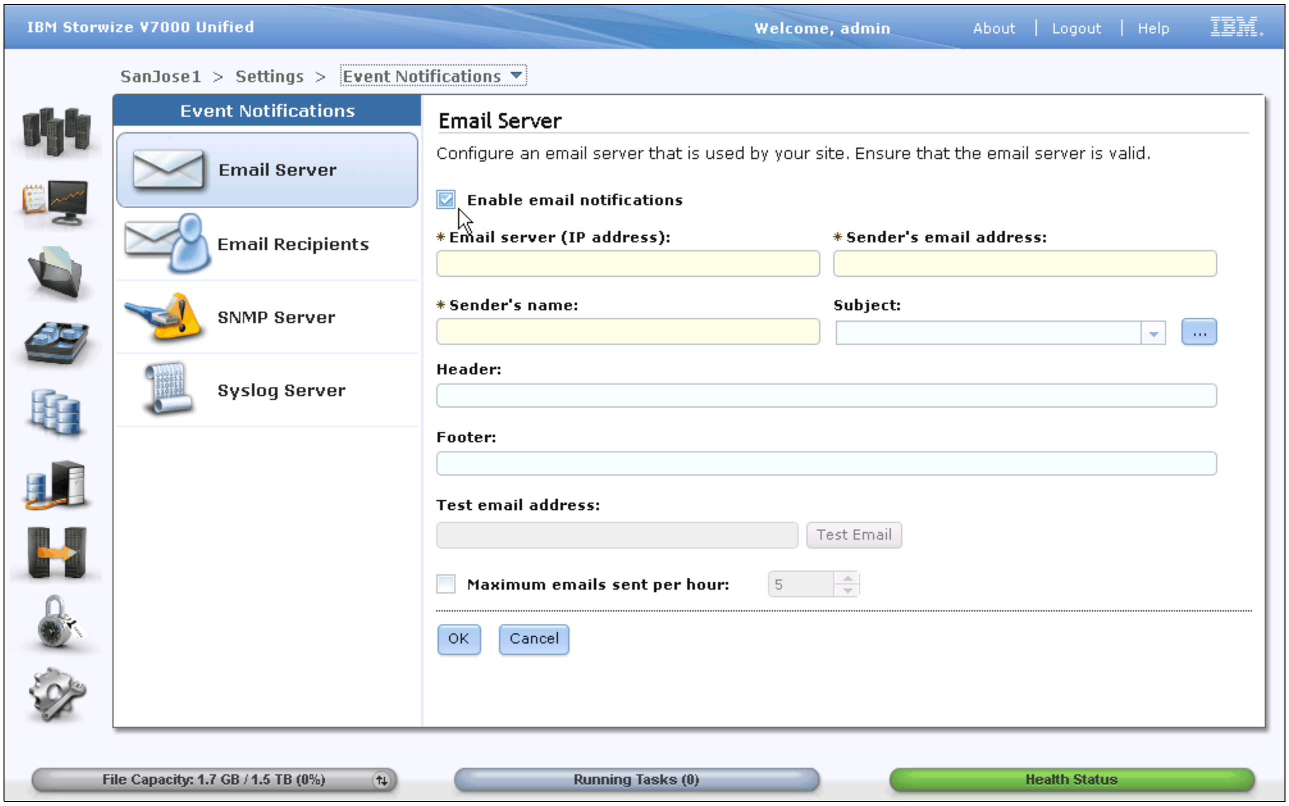
11.10 Alerts

The Storwize V7000 Unified system supports several methods of alerts about events and problems. The Call Home to IBM Support feature is described in 11.8.3, “Support” on page 182. Alerts can also be sent to an email address and to an SNMP server. To configure alerts, select **Settings** → **Event Notifications**.

11.10.1 Email

For email, first set up the details of your Simple Mail Transfer Protocol (SMTP) mail server on the panel, as shown in Figure 11-46 on page 191.

1. Click **Enable email notifications**, and then insert the IP address of your SMTP server. Also, you must enter the reply address for the email and a sender’s name. These identifiers on the email header show the recipient where the email is from. Use an easily recognized name to make alerts clear.
2. By using the ellipses button (**...**), define as many subject options as needed. Then, select the subject content from the pull-down menu. There is an option to add header and footer details to include about the event text.
3. A test email can be sent at any time to confirm the settings and prove the alerting path. Enter a valid email target and click **Test Email** (see Figure 11-46 on page 191). We recommend that you send a test email.



Email Server

Configure an email server that is used by your site. Ensure that the email server is valid.

☑ Enable email notifications

* Email server (IP address):

* Sender's email address:

* Sender's name:

Subject:

Header:

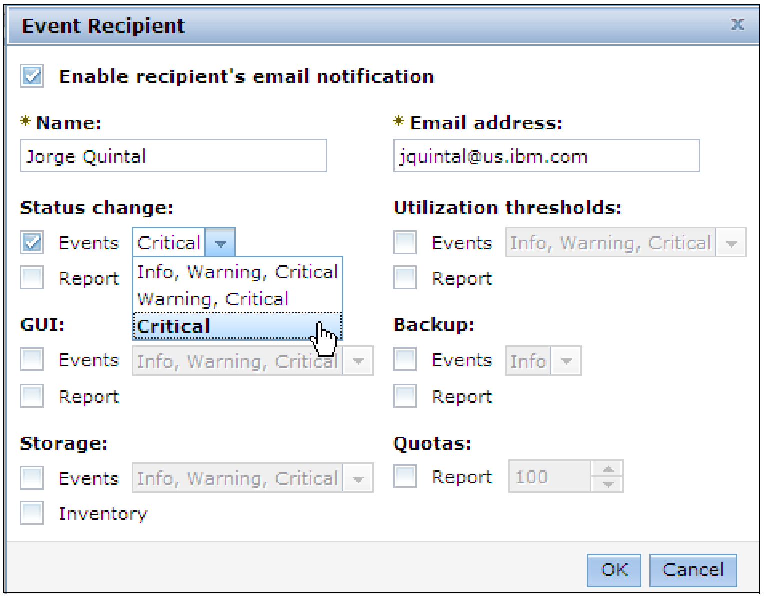
Footer:

Test email address:

□ Maximum emails sent per hour: 5

Figure 11-46 Email server configuration

4. Next, click email recipients to enter addresses. A list of current recipients is displayed. You can select any line to edit or delete from the action drop-down menu. To add an entry, click **New Recipient**. This opens the entry dialog window in Figure 11-47.



Event Recipient

☑ Enable recipient's email notification

* Name: Jorge Quintal

* Email address: jquintal@us.ibm.com

Status change:

☑ Events Critical

□ Report

GUI:

□ Events Info, Warning, Critical

□ Report

Storage:

□ Events Info, Warning, Critical

□ Inventory

Utilization thresholds:

□ Events Info, Warning, Critical

□ Report

Backup:

□ Events Info

□ Report

Quotas:

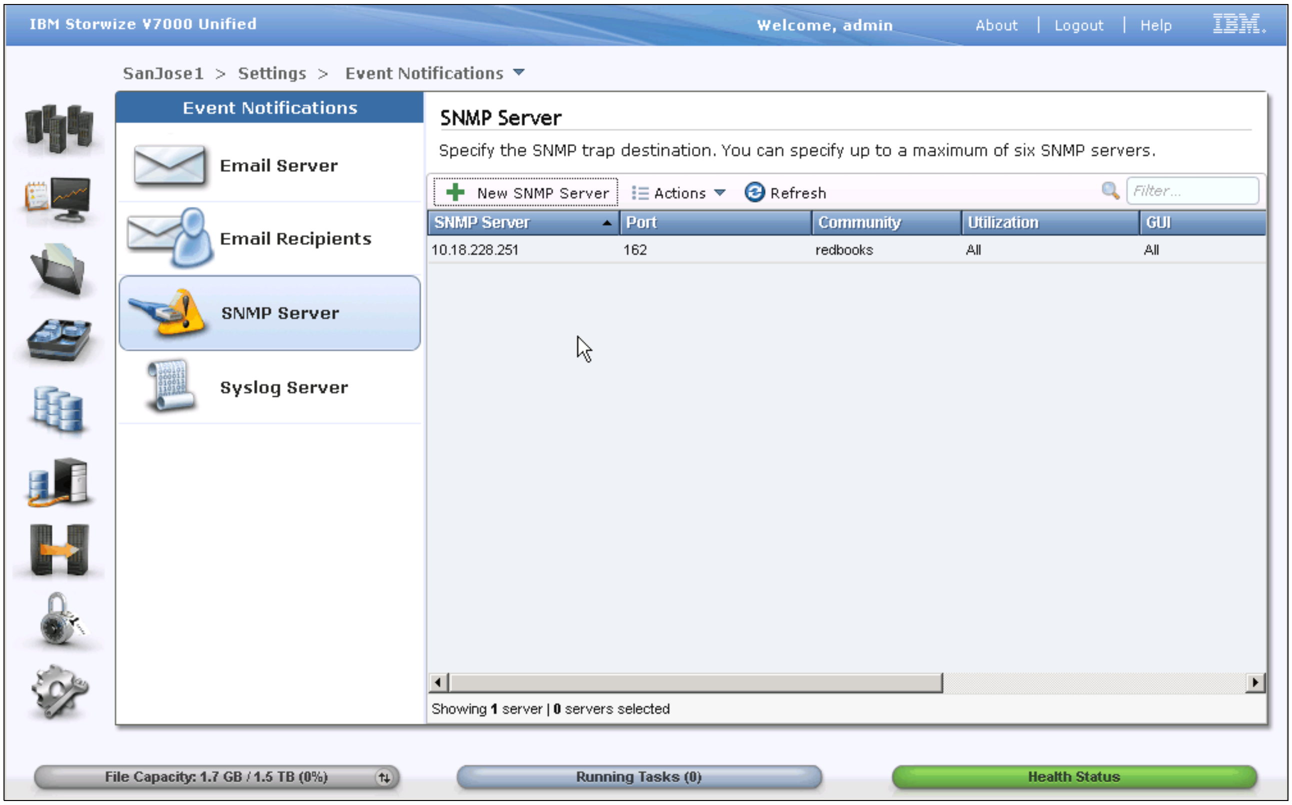
□ Report 100

Figure 11-47 Event Recipient window

5. Enter a name and email address for the recipient. Under Status change, click the **Events** check box for each type of event type that you want this recipient to be notified. Several events have a criticality level. Select the **Utilization threshold**. The Critical-only threshold is usually best to reduce the number of email messages that users receive.
6. Click the **Reports** box if reports are also required. For **Quotas**, choose the percentage of the hard limit for the threshold. Select others as appropriate to your environment.

11.10.2 Simple Network Management Protocol server setup

1. If you use a Simple Network Management Protocol (SNMP) server to monitor your environment and want to receive alerts from the Storwize V7000 Unified system, select **Settings** → **Event Notifications** and then click **SNMP**.
2. Complete the form fields that are shown in Figure 11-48.
3. Enter the server IP address and port.
4. Complete the SNMP community name.
5. Select the event types that are to be sent and for each event, and select the severity.



SNMP Server

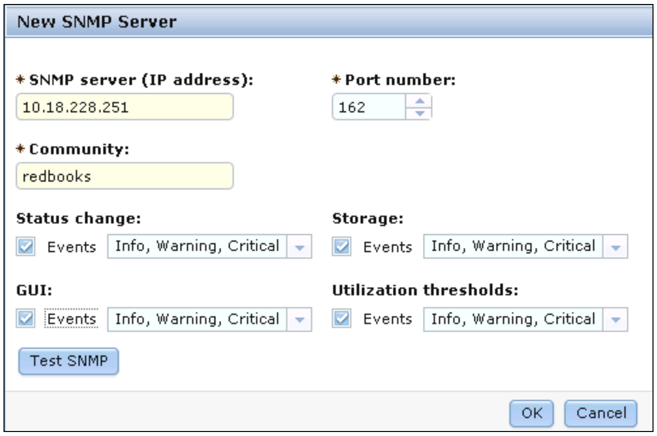
Specify the SNMP trap destination. You can specify up to a maximum of six SNMP servers.

SNMP Server	Port	Community	Utilization	GUI
10.18.228.251	162	redbooks	All	All

Showing **1** server | **0** servers selected

Figure 11-48 SNMP server settings

6. Click **New SNMP Server** to add a server, or highlight the line.
7. Use the actions menu to edit or delete in the pop-up window that is shown in Figure 11-49 on page 193.



New SNMP Server

* SNMP server (IP address): 10.18.228.251

* Port number: 162

* Community: redbooks

Status change: ☑ Events Info, Warning, Critical

Storage: ☑ Events Info, Warning, Critical

GUI: ☑ Events Info, Warning, Critical

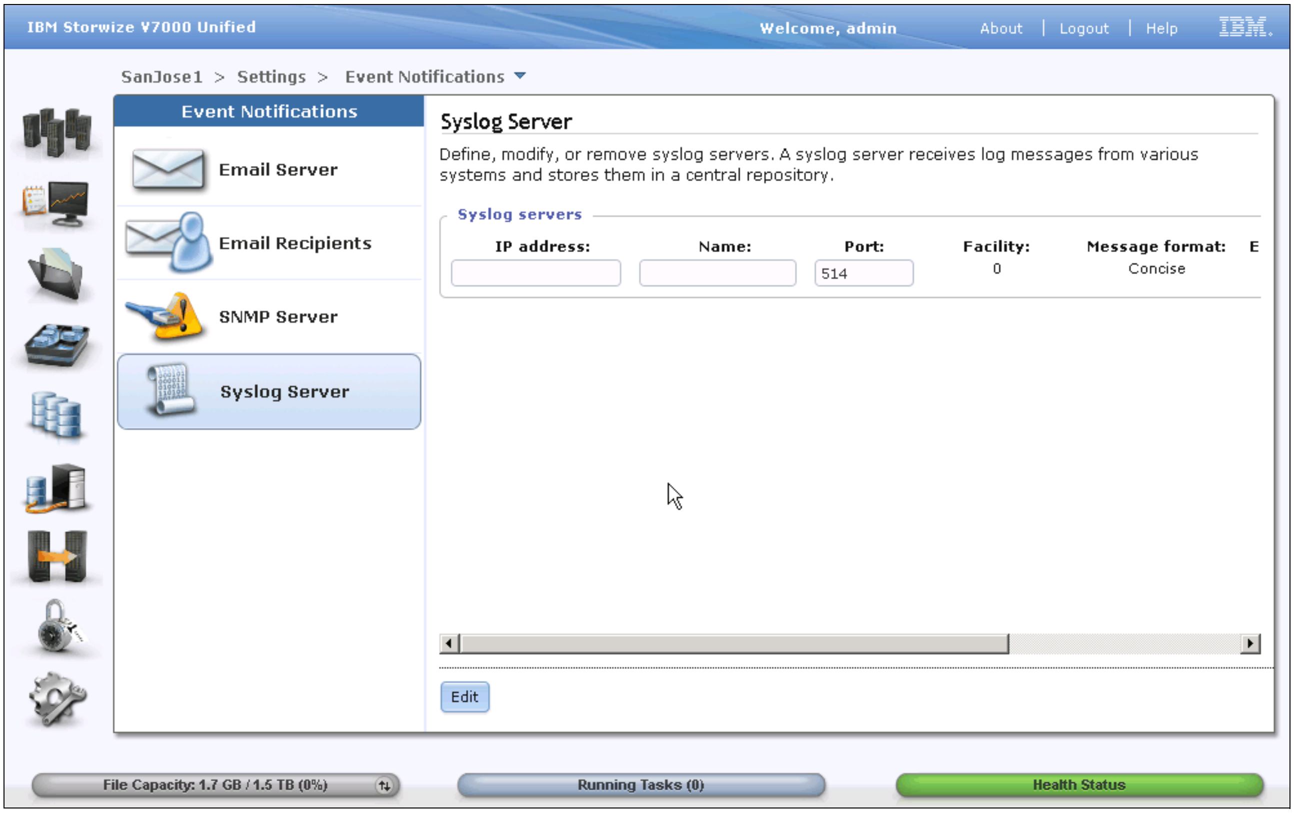
Utilization thresholds: ☑ Events Info, Warning, Critical

Figure 11-49 Edit SNMP server

11.10.3 Syslog server

If you want to offload cluster messages to a logging server, enter the details in the panel shown in Figure 11-50.

1. Select **Settings** → **Event Notifications** and then **Syslog Server**.
2. Enter the required information.



Syslog Server

Define, modify, or remove syslog servers. A syslog server receives log messages from various systems and stores them in a central repository.

Syslog servers

IP address:	Name:	Port:	Facility:	Message format:	E
		514	0	Concise	

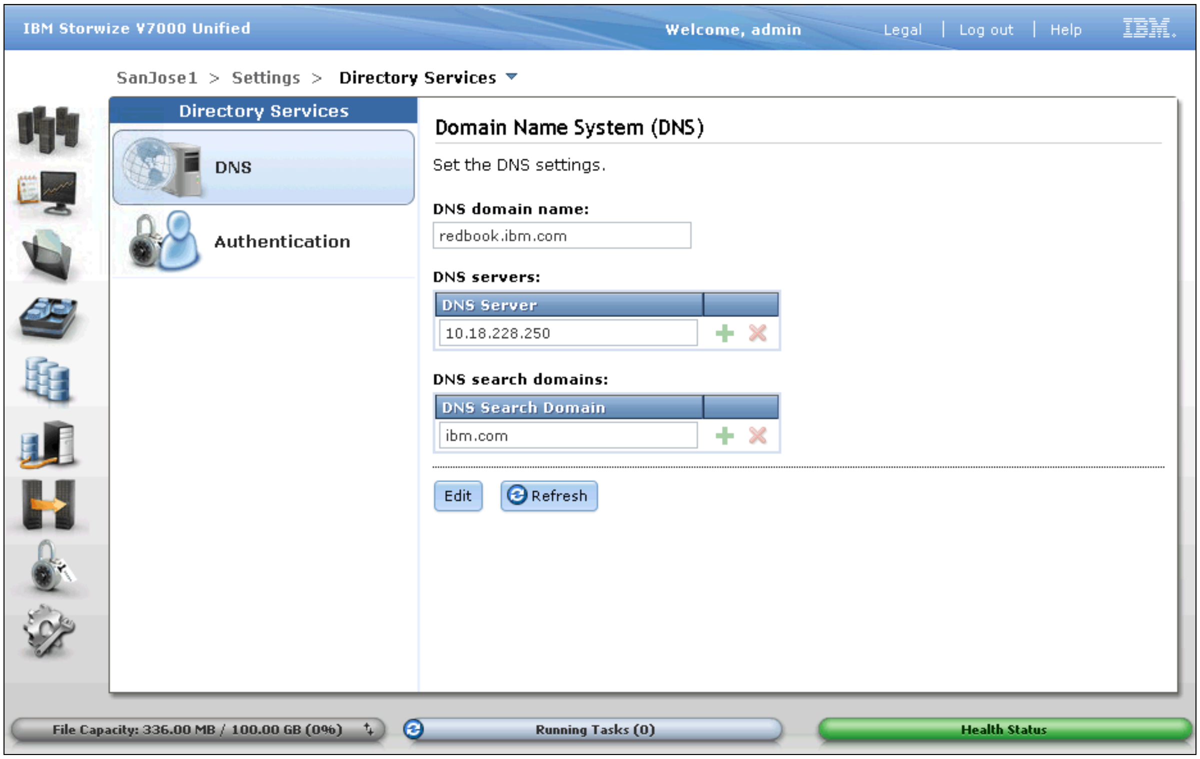
Figure 11-50 Syslog settings

11.11 Directory services and authentication

1. Select **Settings** → **Directory Services**.
2. Set up the DNS server settings by clicking **DNS** in the left panel.

11.11.1 Domain Name System

1. In the Domain Name System (DNS) settings panel shown in Figure 11-51, click **Edit** to make changes.
2. First, ensure that the DNS domain name is correct. This is important with Active Directory and is a key component with authentication.
3. Define your DNS server and any backup DNS servers by using the **+** icon to create more entries.
4. Next, add any search domains that are outside of the primary domain name that is involved in accessing this cluster.



Domain Name System (DNS)

Set the DNS settings.

DNS domain name: redbook.ibm.com

DNS servers:

DNS Server
10.18.228.250

DNS search domains:

DNS Search Domain
ibm.com

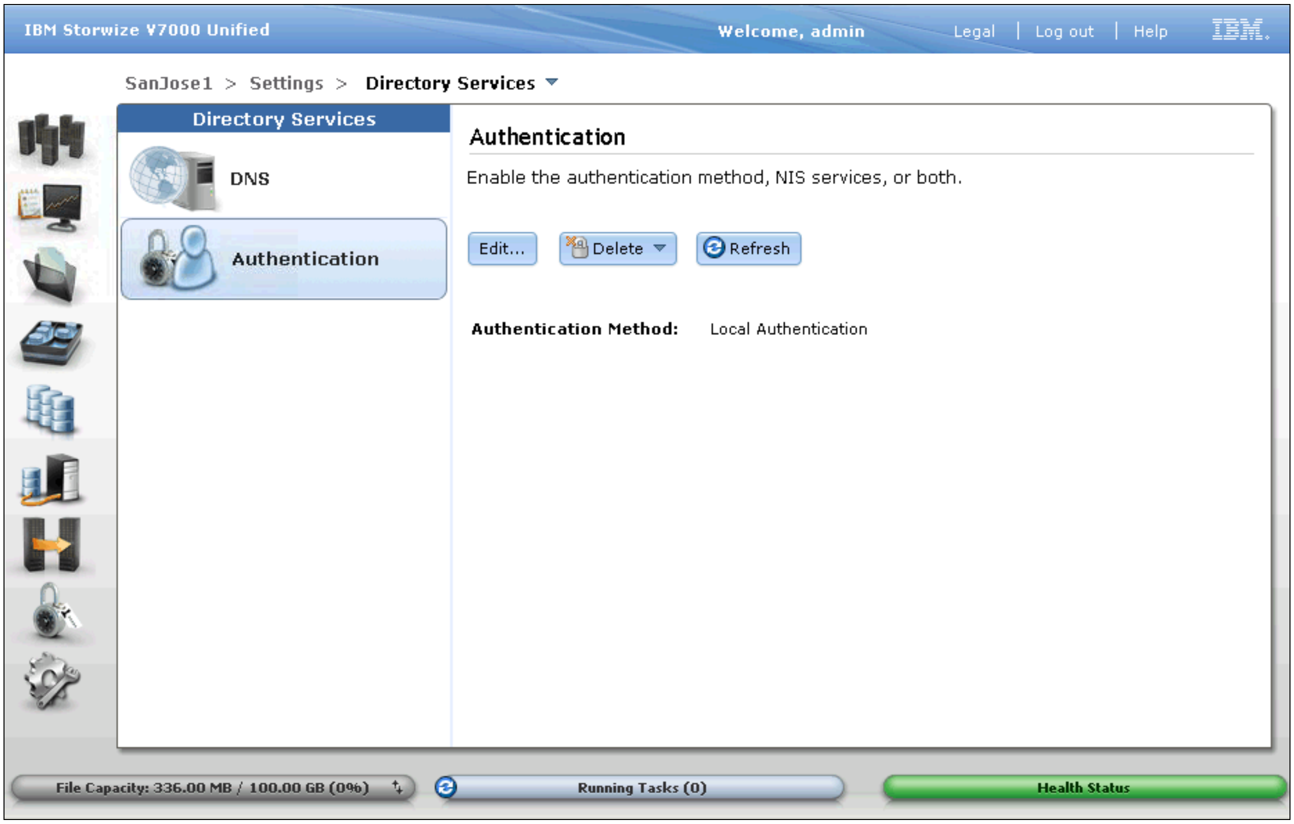
Figure 11-51 DNS settings

11.11.2 Authentication

1. Now, click the **Authentication** icon, as shown in Figure 11-52 on page 195.

   The system requires that you determine a method of authentication for users. It supports either a remote authentication service or a local authentication. Remote authentication is provided by an external server that is dedicated to authenticate users on the system.

   Before configuring remote authentication on your system, ensure that the remote authentication service is set up correctly. Local authentication is provided by an internal mechanism in the system that is dedicated to authentication.



Authentication

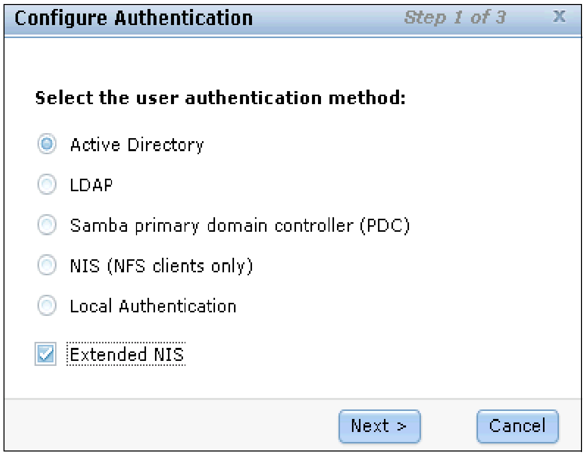
Enable the authentication method, NIS services, or both.

Authentication Method: Local Authentication

Figure 11-52 Authentication settings

> **Caution:** Use extreme care when editing the authentication settings on a running cluster, because that might cause loss of access to shares.

2. If you must make any changes, click **Edit**, which opens the authentication window shown in Figure 11-53.
3. A dialog window messages asks you to confirm that you want to edit these settings, as shown in Figure 11-54 on page 196.



Select the user authentication method:

● Active Directory

○ LDAP

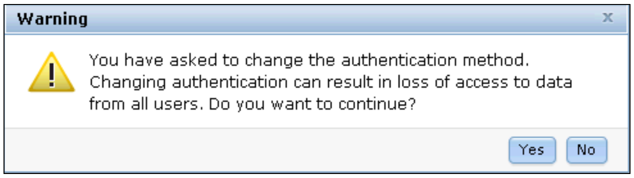
○ Samba primary domain controller (PDC)

○ NIS (NFS clients only)

○ Local Authentication

☑ Extended NIS

Figure 11-53 Editing window


Warning

You have asked to change the authentication method. Changing authentication can result in loss of access to data from all users. Do you want to continue?

Figure 11-54 Change authentication method?

The cluster GUI now takes you through the setup panels for the authentication setup. These are the same panels that the Easy Setup wizard guided you through. They are described in “Authentication” on page 174. If any configuration is already set, the panel entry fields are automatically filled with the existing values.

> **Note:** With local authentication, you enter user and group information locally. That is covered in Section 11.13.3, “Create local users by using local authentication for NAS access” on page 200.

11.12 Health check

To check the health of the system and correct any problems before continuing, check the status in the bottom of the panel.

1. First, check the color of the status indicator at the bottom of the GUI window. This is a good indicator of any serious issues. If it is not green, click **X** to get a summary of the areas that are unhealthy.
2. For more detail, use the CLI **`lshealth`** command to get a listing of each function in the cluster and its status. Use the **`-r`** parameter to force the task to refresh the information. You can also drill down into a particular function by using the **`-i`** parameter.
3. Next, review the event logs for the file and block storage, and review the logs for each file module in system details. Ensure that there are no unfixed events.

> **Important:** Make sure that storage pools are maintained in a green state, especially if compression is used. If the storage pool is allowed to run out of physical space, that causes compressed volumes to go offline.

11.13 User security

It is important to change the default passwords and define profiles for administrators of the cluster. Change the passwords for the following user IDs:

admin This is the default user ID for the Storwize V7000 Unified cluster.

superuser This is the default user ID for the Storwize V7000 storage enclosure.

You can also define more users as required.

At the time of writing, the root password for the file modules is known to IBM Support and implementation teams to assist in setup and recovery. In the future, when all needed functions are available through the GUI and CLI, this password will be changed. If this has occurred, you can ignore this step. If the password is still the widely known default one, ask your IBM team to assist you to change it. The CLI **`chrootpwd`** command changes it across both nodes.

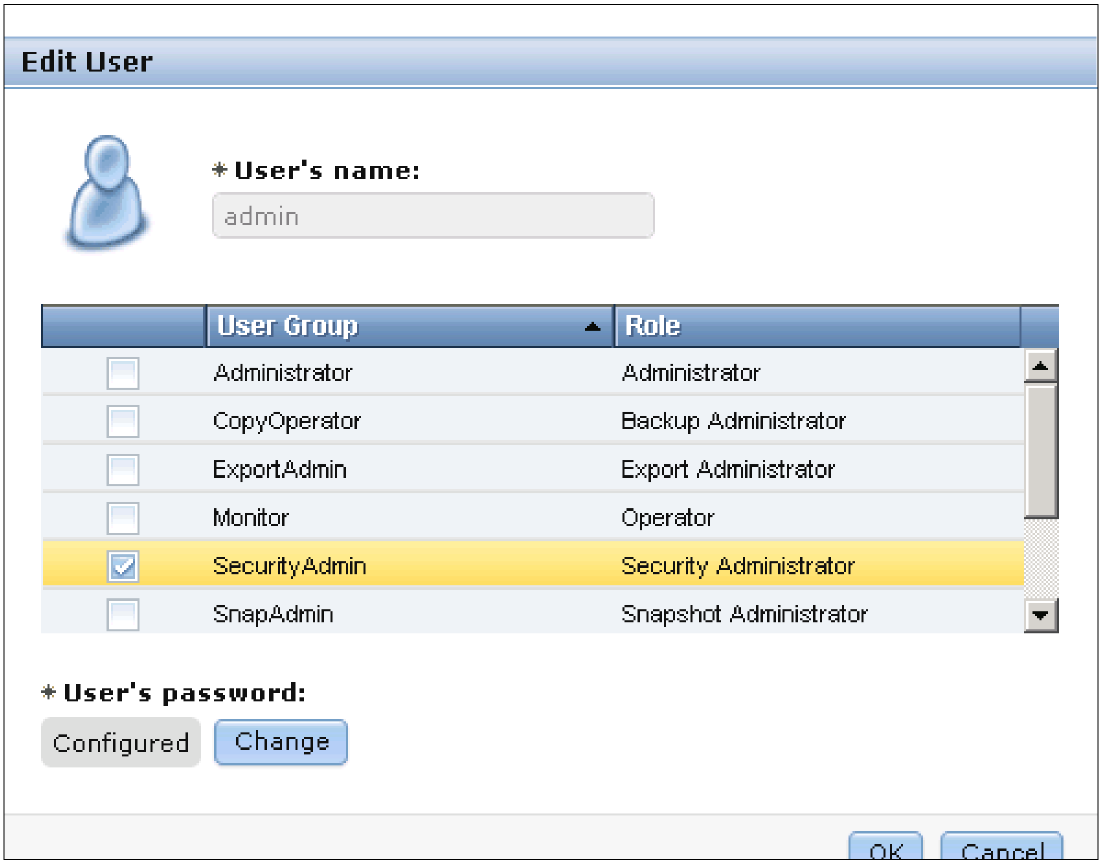
Warning: It is important to change the default passwords on both the cluster and the Storwize V7000 storage enclosure.

11.13.1 Change passwords

In the following sections, we show how to change the passwords.

Change the storage enclosure password

1. Log on to the Storwize V7000 storage GUI directly. Select **Access** → **Users**. The All Users view is displayed by default.
2. Highlight the **superuser** user definition and use the pull-down menu or right-click to select **Properties**. This action opens the Edit User properties window that is shown in Figure 11-55.
3. Click **Change** in the User's password section and enter a new password. Click **OK** to save the change.



Edit User

* User's name: admin

	User Group	Role
□	Administrator	Administrator
□	CopyOperator	Backup Administrator
□	ExportAdmin	Export Administrator
□	Monitor	Operator
☑	SecurityAdmin	Security Administrator
□	SnapAdmin	Snapshot Administrator

* User's password: Configured

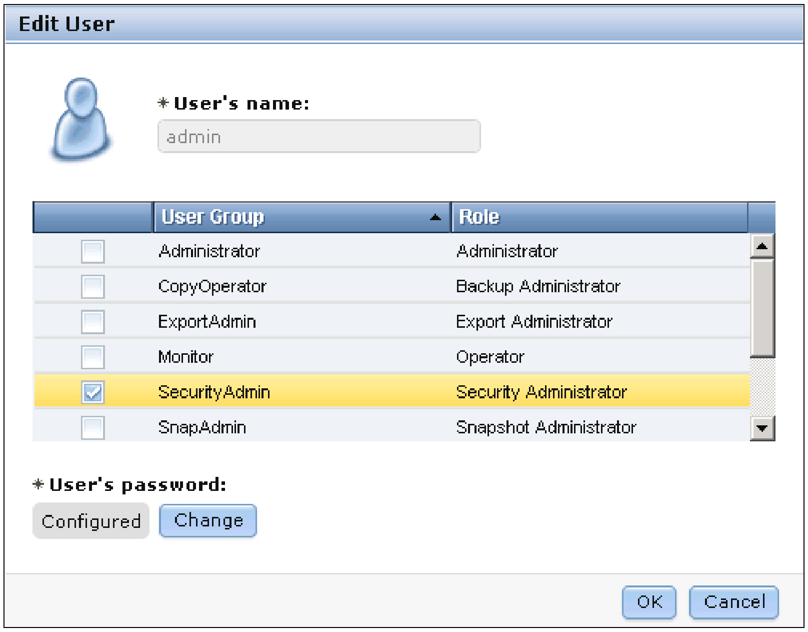
Figure 11-55 Edit user: block

Alternatively, you can use the following command:

```
svctask chuser -password xxxxxxxx superuser
```

Change the cluster admin password

1. On the Storwize V7000 Unified GUI, select **Access** → **Users**. The All Users view is displayed by default.
2. Highlight the **admin** user definition and use the **Actions** pull-down menu or right-click to select **Edit**. This selection starts the Edit User window, as shown in Figure 11-56.
3. Click **Change** in the User's password section and enter a new password.
4. Click **OK** to save.



Edit User

* User's name: admin

	User Group	Role
□	Administrator	Administrator
□	CopyOperator	Backup Administrator
□	ExportAdmin	Export Administrator
□	Monitor	Operator
☑	SecurityAdmin	Security Administrator
□	SnapAdmin	Snapshot Administrator

* User's password: Configured

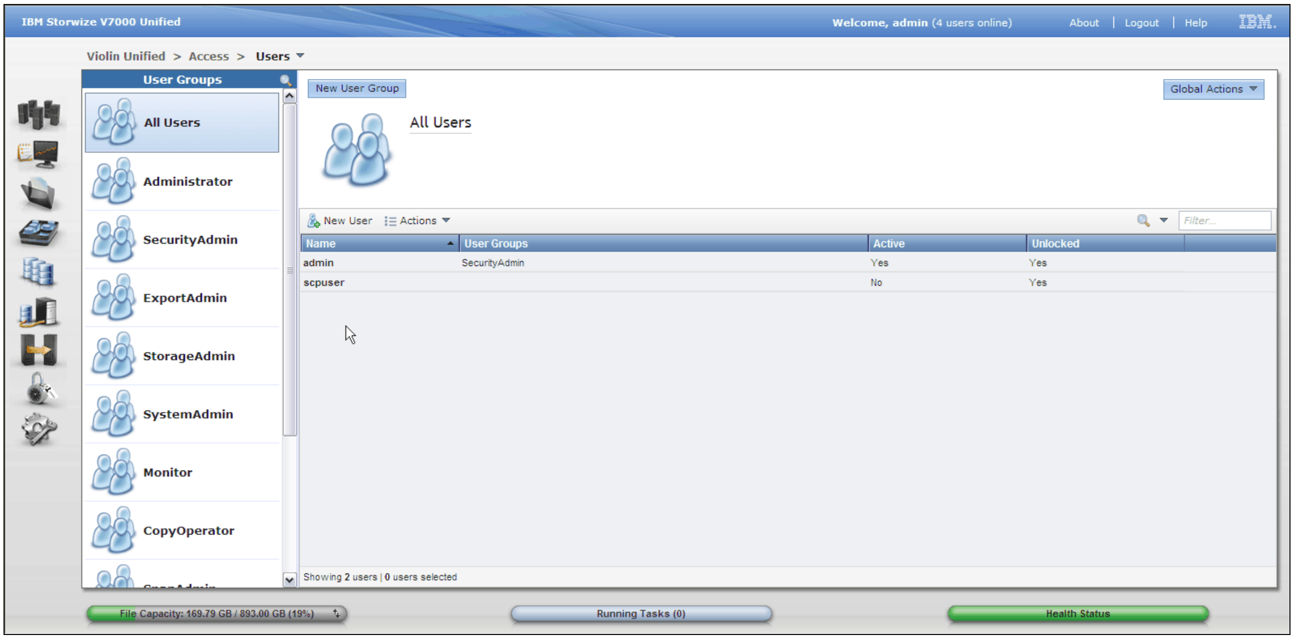
Figure 11-56 Edit user: Unified

Alternatively, you can use the following command:

```
chuser admin -p xxxxxxxx
```

11.13.2 Create cluster users

1. Select **Access** → **All Users**, and then click **New User**, as shown in Figure 11-57 on page 199.

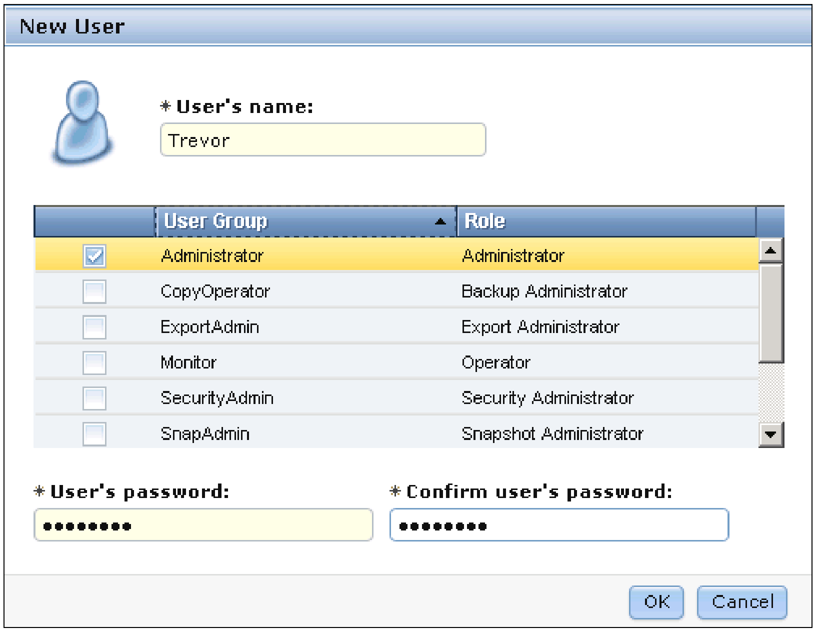


Name	User Groups	Active	Unlocked
admin	SecurityAdmin	Yes	Yes
scpuser		No	Yes

Showing 2 users | 0 users selected

Figure 11-57   Cluster users

This opens the New User window shown in Figure 11-58.



* User's name: Trevor

	User Group	Role
☑	Administrator	Administrator
□	CopyOperator	Backup Administrator
□	ExportAdmin	Export Administrator
□	Monitor	Operator
□	SecurityAdmin	Security Administrator
□	SnapAdmin	Snapshot Administrator

* User's password: ●●●●●●●●

* Confirm user's password: ●●●●●●●●

Figure 11-58   New User window

2. Type in the user’s name and select the level of authority from the list. Set a password and type it again to confirm.
3. Click **OK** to create that user account.

With local authentication, you enter user and group information by using the method that is described in Section 11.13.3, “Create local users by using local authentication for NAS access” on page 200.

Alternatively, you can use the following command:

```
mkuser Trevor -p passw0rd -g Administrator
```

11.13.3 Create local users by using local authentication for NAS access

Besides external authentication, such as Active Directory or LDAP, the system supports user authentication and ID mapping by using a local authentication server for NAS data access. After you configure local authentication on your system, define groups and users who are registered in the local authentication service as eligible to access data. Use the NAS protocols that are supported by this system.

Using local authentication eliminates the need for a remote authentication service, such as Active Directory or Samba primary domain controller (PDC), so it simplifies authentication configuration and management. Local authentication is best for environments where no external authentication service is present or if the number of users is relatively small. Local authentication supports up to 1000 users and 100 user groups. For configurations where all users are in a single group, the system supports 1000 users. A single user can belong to 16 groups. For larger numbers of users and groups, use remote authentication to minimize performance impact on the system.

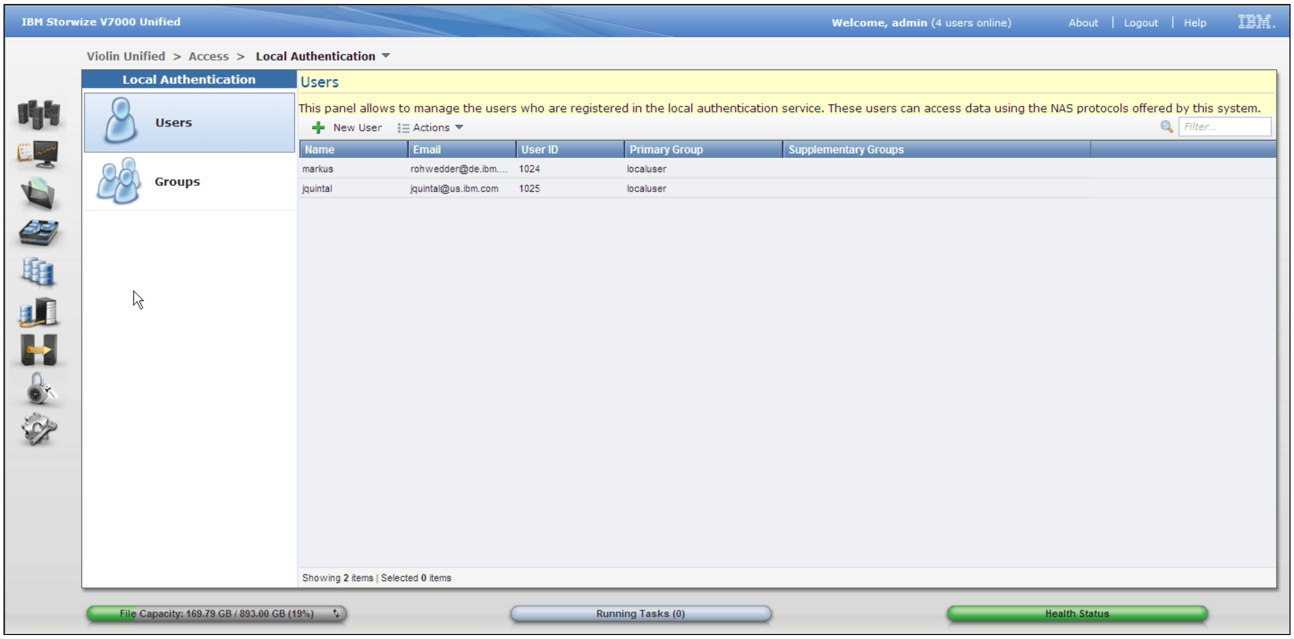
When creating users and groups for local authentication, ensure that user names and group names are consistent across all systems in your environment. If NAS users access data from two or more systems, ensure that those users have the same user names and primary group on each system. Consistent user and group attributes are required for using advanced functions, such as the IBM Advanced Cloud Engine and asynchronous replication. Consistency also provides flexibility in moving data between systems in your environment and simplifies migration to an external LDAP server.

When managing multiple systems, the administrator needs to designate a primary system that contains all users and groups. Create any new user and group on the primary system first, and then create them on any other systems by using the same user ID or group ID that was defined on the primary system. This practice helps ensure that a user ID or group ID is not overloaded. When specifying user IDs and group IDs, you can have the system automatically generate these values. However, to ensure control over these values and to minimize any authorization problems that might be introduced over time, you can assign these values manually, instead, if you prefer.

When creating users and groups for local authentication, the user and group names are not case-sensitive. For example, if a user named John exists on the system, a new user named john cannot be created. This is a limitation of some of the supported protocols. In addition, NAS user and group names cannot be the same as CLI users and system users.

1. Select **Access → Local Authentication**, as shown in Figure 11-59 on page 201.

   New local users and new local groups can be created by using this process. You can modify a current user by using the pull-down menu or by right-clicking a user or group name.


IBM Storwize V7000 Unified
Welcome, admin (4 users online)
About | Logout | Help
IBM.
Violin Unified > Access > Local Authentication
Local Authentication
Users
Groups
Users
New User
Actions
Filter...
File Capacity: 169.79 GB / 893.00 GB (19%)
Running Tasks (0)
Health Status

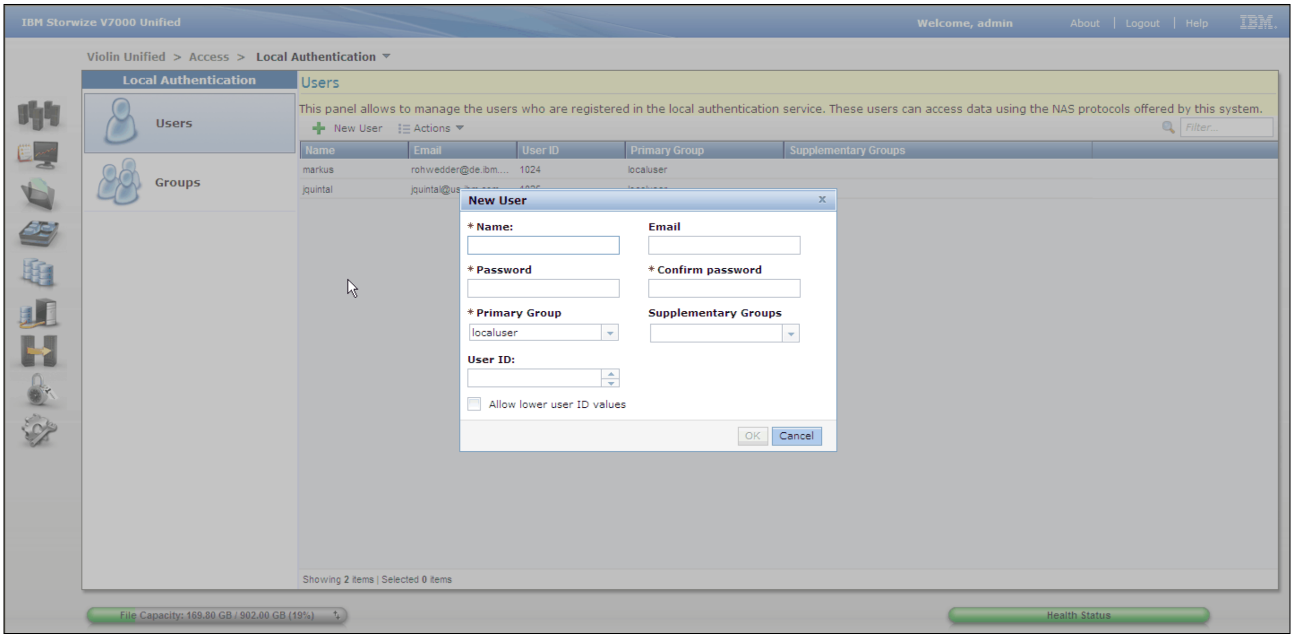
This panel allows to manage the users who are registered in the local authentication service. These users can access data using the NAS protocols offered by this system.

Name	Email	User ID	Primary Group	Supplementary Groups
markus	rohwedder@de.ibm....	1024	localuser	
jquintal	jquintal@us.ibm.com	1025	localuser	

Showing 2 items | Selected 0 items

Figure 11-59 Local authentication for a user

The pop-up window in Figure 11-60 shows how to add a new *user* for NAS access.



This panel allows to manage the users who are registered in the local authentication service. These users can access data using the NAS protocols offered by this system.

Name	Email	User ID	Primary Group	Supplementary Groups
markus	rohwedder@de.ibm....	1024	localuser	
jquintal	jquintal@us			

Showing 2 items | Selected 0 items

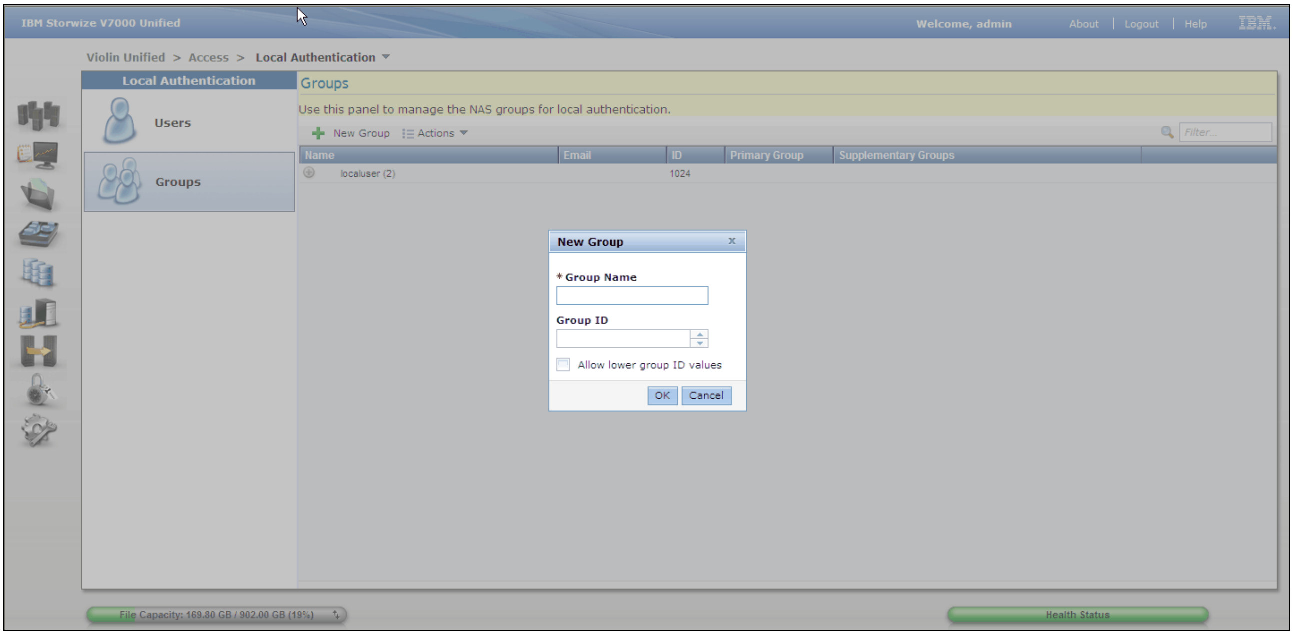
New User

* Name:	Email
* Password	* Confirm password
* Primary Group: localuser	Supplementary Groups
User ID:	

□ Allow lower user ID values

Figure 11-60 Add new user

The pop-up window in Figure 11-61 shows how to add a new *group* for NAS access.


IBM Storwize V7000 Unified
Welcome, admin
About | Logout | Help
IBM.
Violin Unified > Access > Local Authentication
Local Authentication
Users
Groups
New Group
Actions
Filter...
New Group
Group Name
Allow lower group ID values
OK
Cancel
File Capacity: 169.80 GB / 902.00 GB (19%)
Health Status

Groups

Use this panel to manage the NAS groups for local authentication.

Name	Email	ID	Primary Group	Supplementary Groups
localuser (2)		1024		

Figure 11-61   Add new group

11.14 Storage controller configuration

The Storwize V7000 storage component provides the storage arrays for the file systems. It is necessary to configure the controllers even if you are not using the Storwize V7000 subsystem for block devices. The basic configuration and initialization must have completed successfully. If the storage was automatically configured during the Easy Setup process and no block volumes are required, the configuration is complete and you can skip this step. Otherwise, continue.

Consult the IBM Storwize V7000 Unified Information Center for your version:

http://ibm.co/1lzcUBs

11.14.1 External SAN requirements

Block volumes are accessed over the Fibre Channel storage area network (SAN). The Storwize V7000 system has four Fibre Channel ports on each node. When it is part of the unified cluster, two ports from each node are dedicated to the file module connections. The remaining two are for connection to the SAN. As is normal practice, two SAN fabrics are needed in parallel for redundancy and manageability. Connect one port from each node to each fabric.

Zone these ports to each other in each fabric to create more internode links.

Host bus adapters (HBAs) on the SAN can now be zoned to the Storwize V7000 ports as required for access to the storage.

11.14.2 Configure storage

Presenting block volumes to hosts is the same for the Storwize V7000 Unified system as it is for the standard Storwize V7000 system. The only difference is that the Storwize V7000 Unified GUI is used for both file and block protocols, but the interface and functions are still the same.

For this reason, we recommend reading *Implementing the IBM Storwize V7000 V6.3*, SG24-7938 for guidance on configuration and management of the block storage. We provide only a summary of the steps that are involved to assist experienced users.

11.15 Block configuration

The following information summarizes the steps to configure block storage volumes for access by hosts over the SAN:

External storage
If external SAN-attached storage is being managed by the Storwize V7000 system, this must be added first. When installed and connected to the SAN, the storage needs to be configured to present volumes (logical unit numbers, or LUNs) to the cluster. The storage must map the volume as though the Storwize V7000 cluster is a host. The cluster then discovers these as MDisks.

Arrays
The physical drives that are fitted to the Storwize V7000 must be built into RAID arrays. These arrays become MDisks and are visible to the cluster when built.

MDisks
The array MDisks and the external MDisks can now be defined. They can be given descriptive names and are ready for use.

Pools
Extent pools, previously known as *MDisk groups*, which are used to build the volumes, must now be defined. At least one pool is required. There are various reasons why multiple pools might be used. The main reason is to separate disks into groups of the same performance. Another reason could be to separate uncompressed from compressed volumes.

Volumes
From the pools, the volumes that the hosts are using can be defined. A volume must get all of its extents from one pool. The volume can be generic, thin-provisioned, mirrored, or compressed.

Hosts
Each host that is accessing the cluster needs to be defined. This requires describing the host and giving it a name, defining the access protocols that it uses based on its operating system type, and defining its FC ports. The port definition includes the port’s worldwide name (WWN).

Mapping
The last step is to create mappings from the volumes to the hosts. A volume can be mapped to multiple hosts, provided that the hosts support disk sharing, are aware that the disk is shared, and can see multiple volumes.

11.15.1 Copy services

Several methods of copying volumes are available to suit your business requirements. These are covered in Chapter 8, “Copy services overview” on page 75.

Because copy services are covered in detail in several manuals and IBM Redbooks publications, we merely summarize the setup steps here and recommend that you consult those books. For the Storwize V7000 system, we recommend that you read *Implementing the IBM Storwize V7000 V6.3*, SG24-7938.

Partnership

To work with block remote copy services to another cluster, you must have defined a relationship with at least one other cluster. Partnerships can be used to create a disaster recovery environment or to migrate data between clusters that are in different locations. Partnerships define an association between a local clustered system and a remote system.

SAN connection

The clusters must be visible to each other over the SAN network. Create zoning so that at least one port from each node can see at least one port on all of the remote nodes. If it is dual fabric, ensure that this is true on both fabrics, for redundancy. It is not necessary, nor recommended, to zone all ports.

Partnership (local)

After the zoning is in place, select **Copy Services** → **Partnerships** to display the partnership list. Click **New Partnership**. This begins a discovery process. If there is at least one working path to a remote cluster, it is a candidate. Select the partner cluster from the list.

You must also specify the bandwidth setting. You can fine-tune this at any time, but it is important to set it correctly. The bandwidth defines to the cluster the maximum aggregate speed that is available to the other cluster. This is the speed for sending data. Do not overrate this setting, because that can flood the link. Do not underrate it, because that can affect throughput (underrating caps, or limits, the data rate). This setting applies only to data sent from this cluster. Data sent from another cluster is not affected by this parameter, but it will be affected by this same setting on the remote cluster.

The cluster shows as partially configured. It must now be defined from the remote site to be complete.

Partnership (remote)

Go to the remote cluster and perform the same operation as you did in Partnership (local). The partnership must be set up from both clusters to be operational.

The bandwidth setting on the remote cluster is an independent setting and controls the data flow rate from the remote back to the local. Normally, these are the same, but they can be set differently if you prefer.

When established, the partnership is fully configured.

Remote copy

1. Select **Copy Services** → **Remote Copy** to display the list of remote copy relationships.
2. To create a relationship, click **New Relationship**. You can use the IBM Metro Mirror and IBM Global Mirror Copy Services features to set up a relationship between two volumes so that updates that are made by an application to one volume are mirrored on the other volume. The volumes can be in the same system or on two different systems.
3. Now, select the type of mirror relationship from the Metro, Global, and Global with Change Volumes.

4. On the next panel, you have the choice of where the auxiliary volume is located. It is possible to set it on the same cluster, which you might find useful for some applications.
5. The next panel allows you to choose the volumes and lists all eligible volumes on the specified cluster. Use the pull-down menus to select the volume you want in this relationship.
6. You are then asked whether the relationship is already synchronized. Normally this is “no,” but there are situations during recovery and build when a relationship is being re-established between volumes and the data has not changed on either of them.
7. The last question is whether to begin starting the copy. This is the initial synchronization process that runs in the background until it is complete.

IBM FlashCopy

Because FlashCopy is available only on the same cluster, there is no requirement for a relationship.

1. To create a copy, select **Copy Services** → **FlashCopy Mappings**. Click **New FlashCopy Mapping**. This opens the window to create a mapping.
2. Eligible volumes are listed. Select the source and target volumes from the pull-down menus.
3. On the next window, select the preset type for the copy:

   Snapshot — This gives a point in time copy of the volume, thin-provisioned. No background copy. Intended for temporary use to freeze the volume.

   Clone — A one-time use, full copy of the volume.

   Backup — A full, point-in-time copy of the volume that can be repeatedly refreshed.
4. By using the **Advanced Settings** tab, you can adjust the background copy rate, set the mapping to incremental (to write only new changes), delete after completion, and set the cleaning rate.
5. Next, you are asked whether you want to add the mapping to a consistency group.

Consistency groups

There are two main reasons for using consistency groups. By grouping several copies together, management actions such as start and stop can be applied to the group to reduce workload. But more importantly, the cluster ensures that all members perform the action at the same time. For a stop, this means that the input/output (I/O) operations that are completed for the remote volumes or copy targets are stopped in sync, across all members of the group. That is important for system and application recovery.

1. To create a FlashCopy group, select **Copy Services** → **Consistency Groups**, and then click **New Consistency Group**. In the window that opens, give the group a name, and you are finished. To add a mapping to this group, either choose the group while you are creating the mapping or highlight the mapping in the list and move it to a consistency group.
2. To create a remote copy group, select **Copy Services** → **Remote Copy** and click **New Consistency Group**. Give the group a name and define whether the relationships are local-to-local or remote. You can create new relationships to be members now or add them later. To add a relationship, highlight the entry and use the action menus to add to a consistency group.

FlashCopy mappings

A FlashCopy mapping defines the relationship between a source volume and a target volume.

The FlashCopy feature makes an instant copy of a volume. To create an instant copy of a volume, you must map the source volume (the disk that is copied) to the target volume (the disk that receives the copy). The source and target volumes must be of equal size.

A mapping can be created between any two volumes in a system. The volumes do not have to be in the same I/O group or storage pool. When a FlashCopy operation starts, a checkpoint is made of the source volume. No data is copied when a start operation occurs. Instead, the checkpoint creates a bitmap file. It indicates that no part of the source volume has been copied. Each bit in the bitmap represents one region of the source volume. Each region is called a *grain*.

After a FlashCopy operation starts, read operations to the source volume continue. If new data is written to the source or target volume, the existing data on the source is copied to the target volume before the new data is written to the source or target volume. The bitmap file is updated to mark that the grain of the source volume has been copied so that, later, write operations to the same grain do not recopy the data.

File copy services

Use this function to select different methods to replicate data to and from different file systems. The system supports two types of file system replication: File system replication and remote caching:

- *File system replication* provides asynchronous replication of all file system data on one system to a remote file system, over an IP network. Separate the two systems geographically to provide data recovery and high availability.

  Asynchronous replication allows one or more file systems in a Storwize V7000 Unified system file namespace to be defined for replication to another V7000 Unified system over the client network infrastructure. Files that are created, modified, or deleted at the primary location are carried forward to the remote system at each invocation of the asynchronous replication.

  Asynchronous replication is configured in a single-direction, one-to-one relationship, so that one site is considered the source of the data and the other is the target. The replica of the file system at the target remote location is intended to be used in read-only mode until a system or network failure or other source file system downtime occurs. During a file system failure recovery operation, failback is accomplished by defining the replication relationship from the original target back to the original source.

- *Remote caching* provides transparent data distribution among data centers and multiple remote locations over a wide area network (WAN). Remote caching provides local access to centrally stored files and allows users in remote locations to work with files without creating inconsistencies. Data created, maintained, updated, and changed on the home system can be viewed and used on a cache system that is located anywhere in the WAN.

11.16 File services configuration

To create a share (or export), several building blocks need to be in place. We describe these in this section and give examples of how to create each one.

11.16.1 File service components

The file services function is built in layers. You already completed the storage hardware. Depending on the choices during Easy Setup, the storage pools might already be built, too.

Managed disks

The bottom layer is the physical storage, which is based on the Storwize V7000 storage controller. Its configuration might include extra expansion enclosures or external storage systems. The physical storage is made up of RAID arrays. Managed disks (MDisks) are created from these arrays. For external storage, LUNs are created from the arrays. These are presented to the Storwize V7000 and managed as MDisks.

If the Automatically Configure Storage check box was selected during the Easy Setup procedure, the storage was configured with a preferred practice approach by using all of the available disks. They are grouped into arrays as a single MDisk and added to the same pool.

Pools

These MDisks are grouped into pools, and the logical blocks of storage are merged together to create a single, sequential pool of blocks (or *extents*). The system stripes the sequence across the MDisks to increase performance. From these pools, volumes can be created and presented to hosts as Fibre Channel-attached SCSI LUNs or to the file modules to build a file system on. This is the *block storage layer*.

Volumes

Block volumes that are to be presented as FC-attached SCSI volumes are described in 11.15, “Block configuration” on page 203. Volumes are defined for file systems by the file system build process, so there is no need to create them. The file system build process creates several volumes that are created, depending on its size and internal requirements. Also, because they are used only by the file system, the volumes are not visible in the GUI. Volumes were previously known as *vdisks*. There are two types of volumes that the build process can use: uncompressed and compressed volumes.

File systems

A minimum of one file system is required. This provides the base file system and global namespace from which shares and exports can be created. Therefore, each file system comprises a structure of directories or paths that are based on a single root. Any of these directories can be used to create a share or export. The directory name that is being shared is known as a *junction point*. This file system is the IBM GPFS. A minimum of one storage pool is required to build a file system, although a file system can use multiple pools, and multiple file systems can be created from a pool. A special type of file system can also be created that uses two storage pools, with a default system pool for uncompressed data, such as metadata, and noncompressible data and a compressed pool for compressible data.

File sets

To improve the manageability of the file system, GPFS uses a method that creates subsets of the file system for control of file system management at a more granular level. These are called *file sets*, and they behave similarly to a file system. The file sets root can be anywhere in the directory structure of the parent file system. The sets include all files and directories above that junction point.

When creating a file set, the base directory path must exist. However, the directory (or junction point) that is being defined must not exist already, because it is to be created as part of the file set creation process.

You must also define the file set as either dependent or independent. A dependent file set shares the same file system and inode definitions as the parent independent file set that contains it. If set to independent, the file set has its own inode space. That allows for independent management, such as quotas.

When a file system is created, an initial file set is also created automatically in the root directory.

Shares and exports

These are basically the same thing. The terms are simply used differently for Common Internet File System (CIFS), Microsoft Windows (share), and UNIX (export). Each export picks up a junction point in the file system. It then presents these files and subdirectories that are in the directory as allowed by the authentication process. These directory structures and files are shared by using the selected network sharing protocols that the hosts use to access the export.

For each export, you need the following components:

File system
This is the base file system as configured on the file module.

Junction point
This is the directory in the file system that contains the data and subdirectories that are being shared. If files already exist in the directory or subdirectory, for security reasons, access controls must be obeyed to create the share. If the junction does not exist, it is created.

Network protocol
Choose from the available supported protocols. For example: CIFS, NFS, and File Transfer Protocol (FTP).

Depending on the protocol, some of the following parameters are also needed:

User
This is the owner of this file set. When there are files in the file set, this cannot be changed from the Storwize V7000 Unified system. This user must be known to the authentication server.

Access control
If this is a new and empty file set, you can define the access control list (ACL) controls at export creation time.

Access control

When complete, access control presents this share to the network over the defined protocol. Before this share can be accessed and files can be read or written, the access control server must be updated. All users who need access to the export need to exist already. Access control of the files and directories is defined at the authentication server. Although you can allow access to be defined while no files exist during export creation, after files are written, the Storwize V7000 Unified system administrator cannot change access control and cannot define a new access to avoid this restriction.

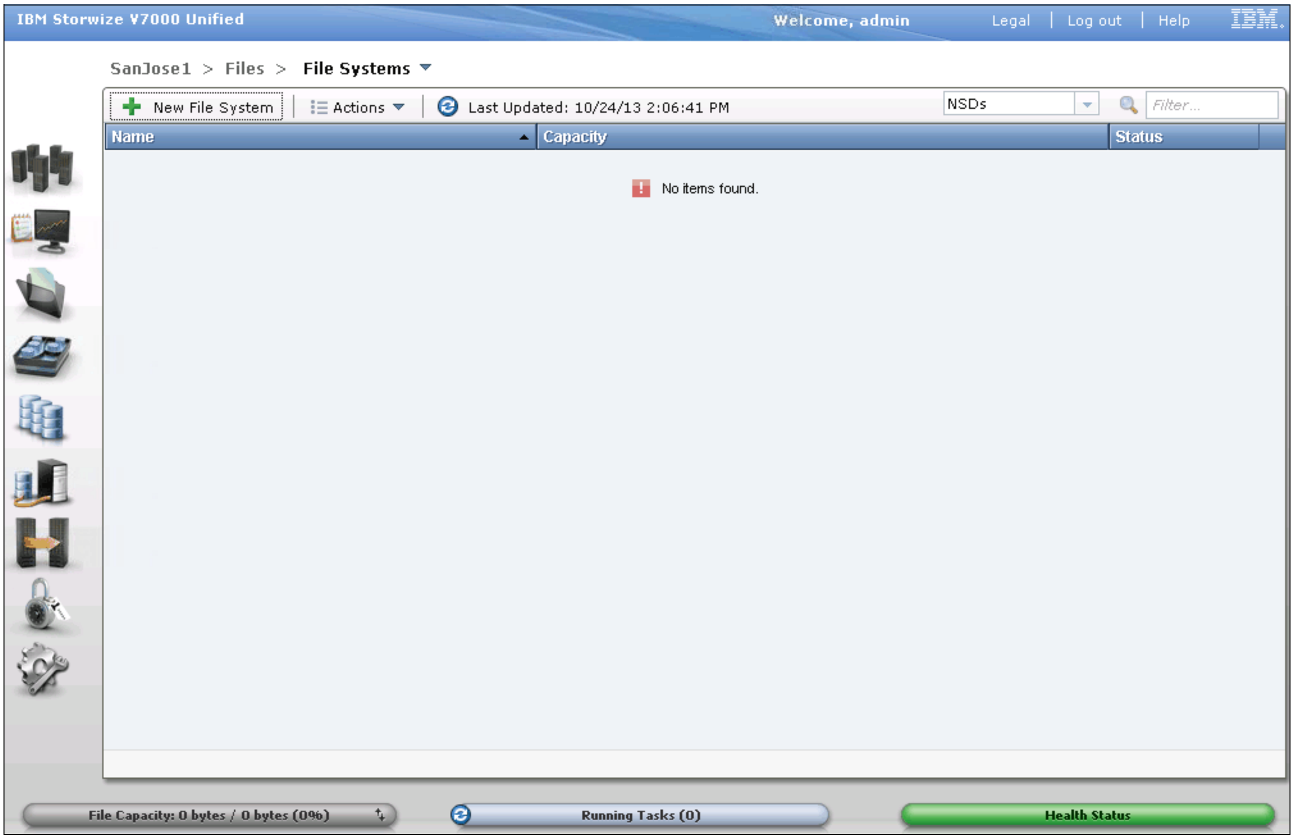
Operation and administration of the access control server is beyond the scope of this book.

11.16.2 File system examples

The following examples demonstrate the processes, show the steps that are involved, and show the GUI windows to expect. These are limited examples, so they do not cover all possible configurations. See Chapter 16, “IBM Real-time Compression in the Storwize V7000 Unified system” on page 289 for information about compressed file systems.

Create a standard uncompressed file system

1. Select **Files** → **File Systems** to display the File Systems panel.
2. Click **New File System** to create a file system, as shown in Figure 11-62.



IBM Storwize V7000 Unified
Welcome, admin
Legal | Log out | Help
IBM.
SanJose1 > Files > File Systems
New File System
Actions
Last Updated: 10/24/13 2:06:41 PM
NSDs
Filter...
Name
Capacity
Status
File Capacity: 0 bytes / 0 bytes (0%)
Running Tasks (0)
Health Status

No items found.

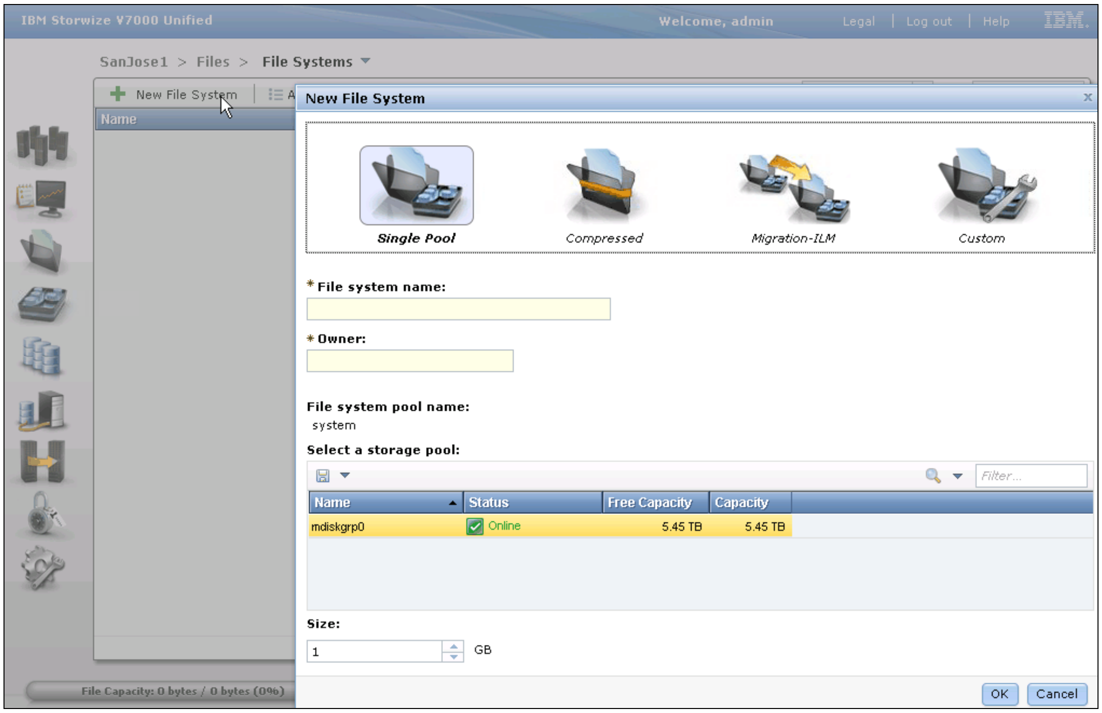
Figure 11-62 File systems

This opens the New File System window that is shown in Figure 11-63 on page 210. There are four options to build a file system. They are selected by clicking the appropriate icon at the top of the page.

The *Single Pool* option is the most commonly used. From the list of defined storage pools, you select a single pool to provide the extents for this file system. You are limited to the free space of that pool. After entering a name for the file system and selecting the pool, you can choose how much of the available free space to use for this file system. Do this by adjusting the slider at the bottom of the window.

3. When you are finished, click **OK** to build the file system.

   A pop-up window shows the progress of the build. The process first builds the Network Storage Devices (NSD) and then the volumes. Typically, five NSDs are created, and their size is set to make up the specified file system size.



* File system name:

* Owner:

File system pool name:
system

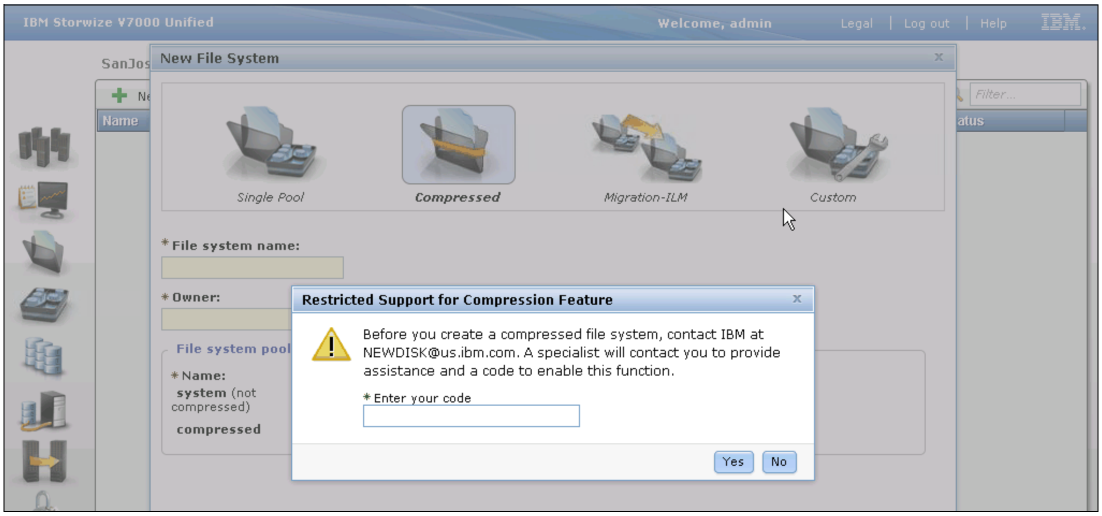
Select a storage pool:

Name	Status	Free Capacity	Capacity
mdiskgrp0	✓ Online	5.45 TB	5.45 TB

Size: 1 GB

Figure 11-63 New File System panel

The second option is to create a compressed file system. To compress data selectively, you will be prompted to contact IBM at `NEWDISK@us.ibm.com`. An IBM specialist will contact you to provide assistance and a code to enable this function. Figure 11-64 shows an example of the dialog window where you enter in the code given to you by the IBM Support representative.



Restricted Support for Compression Feature

Before you create a compressed file system, contact IBM at NEWDISK@us.ibm.com. A specialist will contact you to provide assistance and a code to enable this function.

* Enter your code

Figure 11-64

Another option is to create a file system as a *Migration ILM*. IBM SONAS software has a feature that allows the file system to use multiple pools with different performance capabilities. This is used to build a file system that automatically migrates data to the slower storage, based on predefined thresholds.

Explaining the use of this feature is beyond the intended scope of this book. If you want to use ILM for migration, contact your IBM representative for assistance and check the following IBM Redbooks publications about SONAS software:

- *SONAS Implementation and Best Practices Guide*, SG24-7962
- *SONAS Concepts, Architecture, and Planning Guide*, SG24-7963

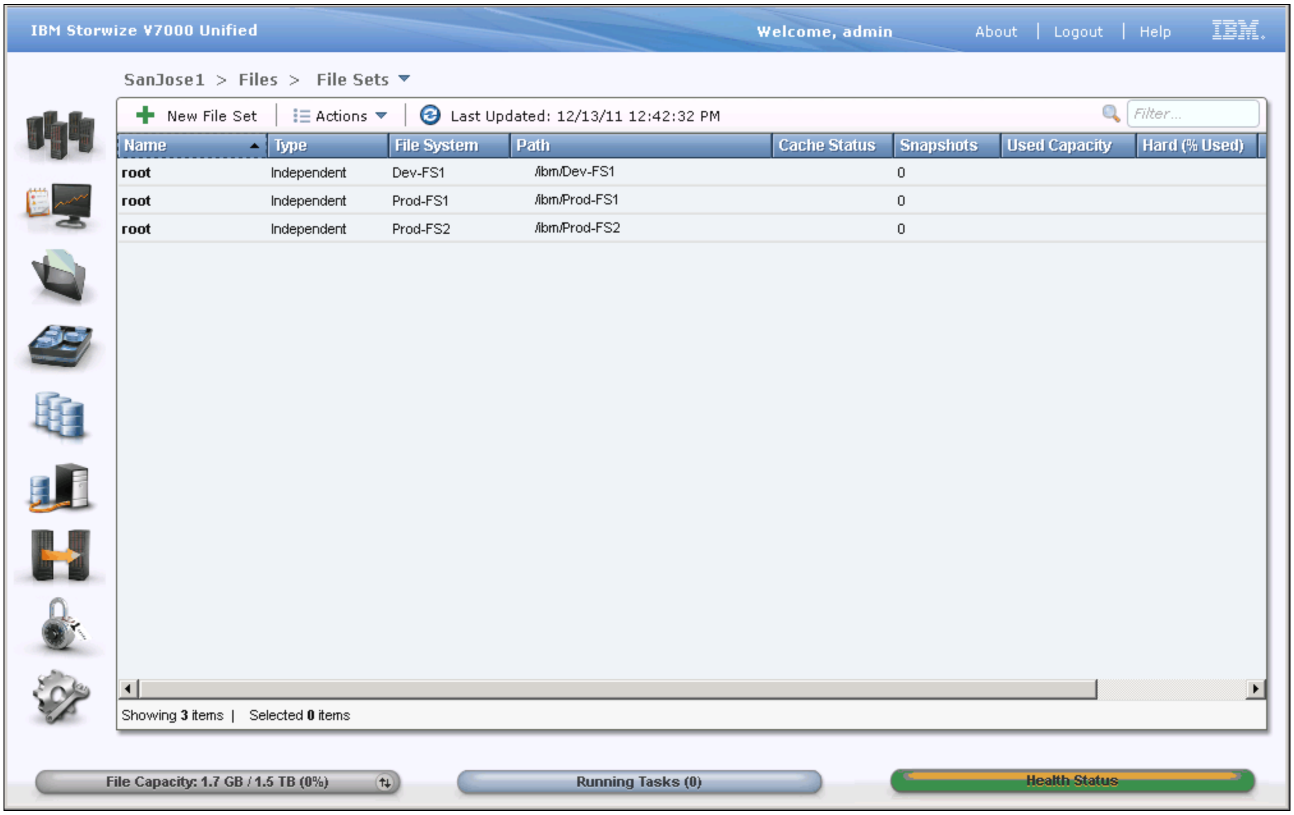
The fourth option is *Custom*. This gives you greater flexibility in creating the file system. The storage can be provisioned from multiple pools, and you can manually define the file system policies. This feature requires in-depth knowledge of GPFS, which you can use to set policies to automate management of data in a file system. You can set and run policies from the command line. If you want to use policies, seek assistance from IBM.

Create a file set

There is no need to create file sets, but they are useful in defining and managing the data shares.

1. To work with file sets, select **Files** → **File Sets**.

   The cluster automatically defined a file set on the base of each file system, called *root*, as shown in Figure 11-65.


IBM Storwize V7000 Unified
Welcome, admin
About | Logout | Help
IBM.
SanJose1 > Files > File Sets
New File Set
Actions
Last Updated: 12/13/11 12:42:32 PM
Filter...
File Capacity: 1.7 GB / 1.5 TB (0%)
Running Tasks (0)
Health Status

Name	Type	File System	Path	Cache Status	Snapshots	Used Capacity	Hard (% Used)
root	Independent	Dev-FS1	/ibm/Dev-FS1		0		
root	Independent	Prod-FS1	/ibm/Prod-FS1		0		
root	Independent	Prod-FS2	/ibm/Prod-FS2		0		

Showing 3 items | Selected 0 items

Figure 11-65 File Sets panel

   More file sets can be created at any junction point in a file system.

2. To create a file set, click **New File Set** at the top of the window.

   This opens the New File Set window. You can use the basic option to define the file set and its path, or you can use the custom option to complete the quota and snapshot details. These can be added later if you want them.

Tip: When browsing for the path, right-click a directory to add a subdirectory under it.

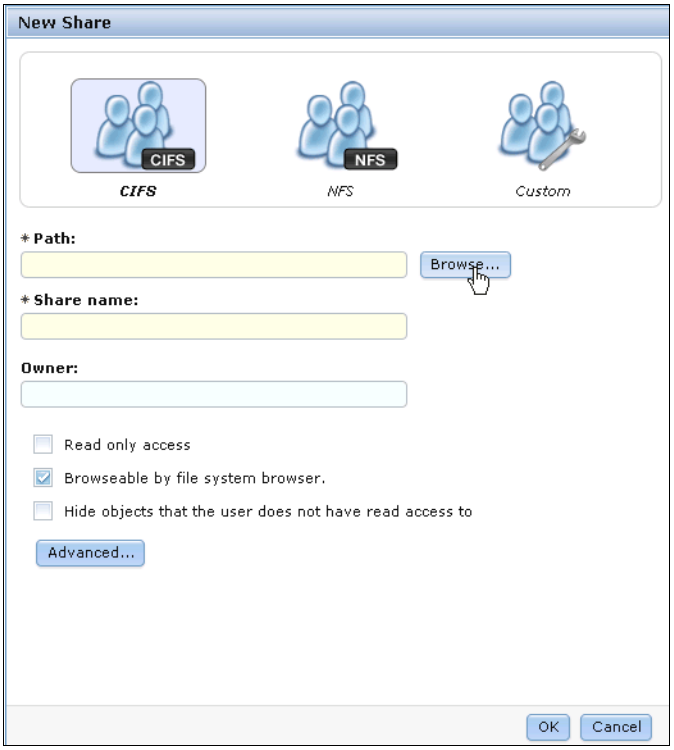
Shares

To manage shares, select **Files** → **Shares**. This opens the main shares window and lists all shares on the cluster. To add a share, click **New Share** at the top of the window to open the New Share window, which is shown in Figure 11-66.

A share can be accessed by using any of the supported access methods or any combination of them. Use buttons at the top of the window to simplify the configuration entry. Use the custom function to use multiple access methods.

All methods have the same common information:

Share name	The name by which the share is known and accessed in the network.
Path	The directory path to the shared directory, which includes the file system.
Owner	The user ID of the owner of this share.



☐ Read only access

☑ Browseable by file system browser.

☐ Hide objects that the user does not have read access to

Figure 11-66 New Share window

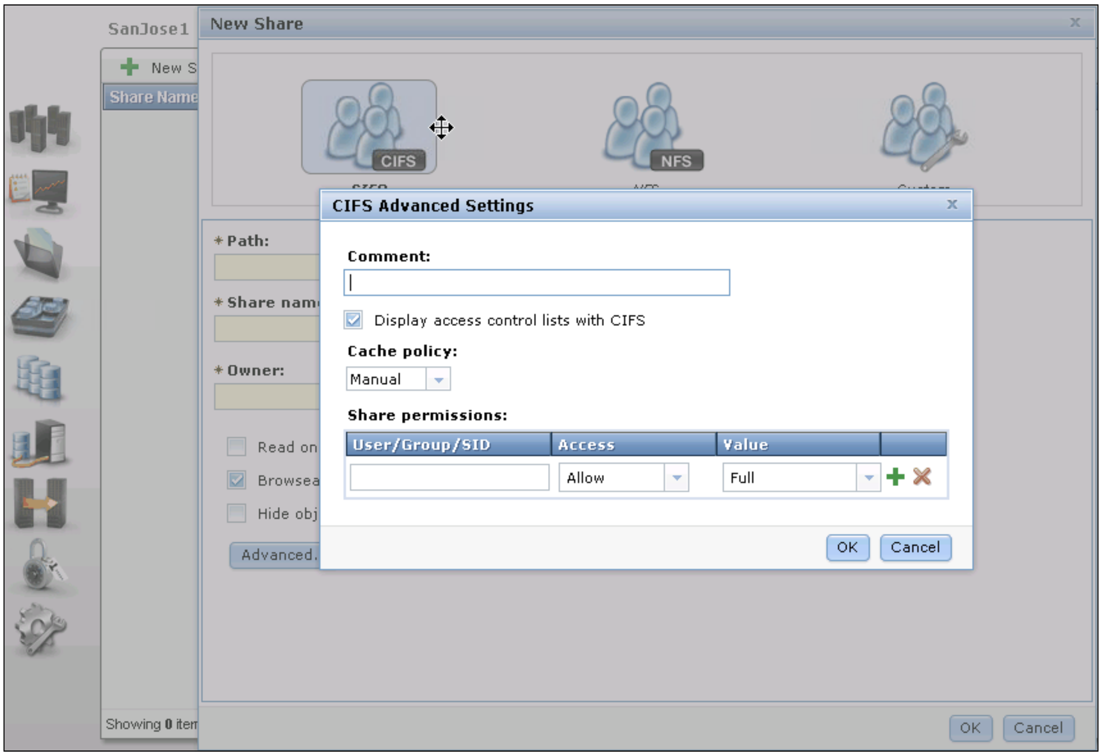
Create a Common Internet File System share

When you are creating a Common Internet File System (CIFS) share, enter the common information first.

1. If you are using the Custom option, click the **CIFS** tab and then click **Enable the CIFS access control list (ACL)**. If you are creating CIFS only, this information is on the main panel.
2. Select the **Read-only**, **Browsable**, and **Hide Objects** options, and then click **Advanced**.

This action opens the window that is shown in Figure 11-67, where there is a field to enter a comment about the share.

In this panel, you can add the CIFS users or groups that have access to this share and specify what access they have. You can add as many entries as you want. Security policy allows this access to be defined for a new share. However, when data is already written to the directory (or subdirectories), access control can be altered only by using the Active Directory service and its associated security processes.



CIFS Advanced Settings

Comment:

☑ Display access control lists with CIFS

Cache policy: Manual

Share permissions:

User/Group/SID	Access	Value
	Allow	Full

Figure 11-67 CIFS Advanced Settings panel

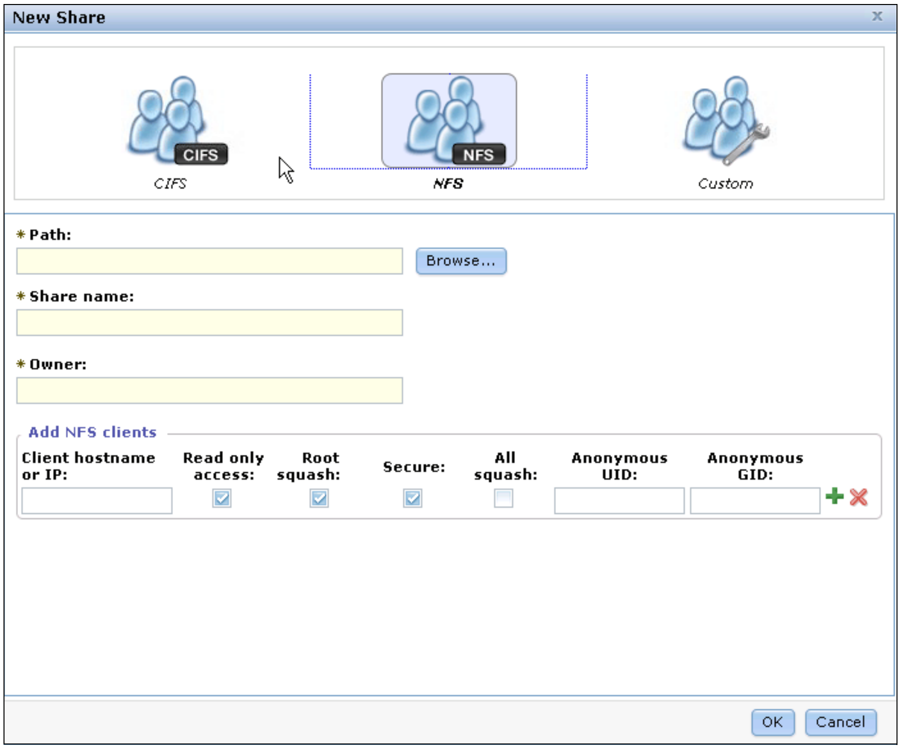
Create an NFS export

1. To configure a share for NFS access, click the **NFS** icon at the top of the New Share window. Or, if you are using Custom for multiprotocol, click the **NFS** tab, as shown in Figure 11-68 on page 214.

2. Use the **Add NFS clients** section to add an entry for each client ID that has NFS access to the share. Use the **+** icon to add each new entry. Set the access to read-only or not, as you prefer.

3. Unless you have a good reason not to, leave the **Root squash** option checked (see Figure 11-68 on page 214).

   If it is not selected, that allows the root user on any system to receive root-privileged access to the share, so full authority is granted. Using root squash removes all access from a remote connection from root, unless specifically allowed.



New Share

* Path:

* Share name:

* Owner:

Add NFS clients

Client hostname or IP:	Read only access:	Root squash:	Secure:	All squash:	Anonymous UID:	Anonymous GID:
	☑	☑	☑	□		

Figure 11-68   NFS share

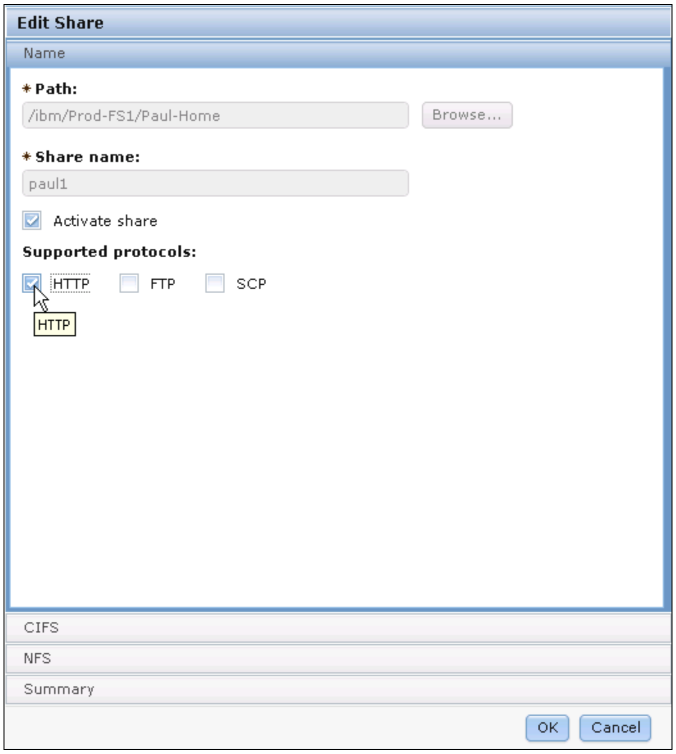
Caution: Always enable *root squash* to prevent unauthorized access.

4. Click the **Secure** check box if you are using an NFS secure connection to access this share.

Create an HTTP, FTP, or SCP export

If this share is already defined, use the Edit function to add any of these protocols. If this is a new share, create a share and enter the common information as previously described.

1. Click the **Custom** icon to show the options, as shown in Figure 11-69 on page 215.


Edit Share

Name

*Path: /ibm/Prod-FS1/Paul-Home

*Share name: paul1

☑ Activate share

Supported protocols:

☑ HTTP □ FTP □ SCP

Figure 11-69 HTTP share

2. Now click the HTTP, FTP, and Secure Copy Protocol (SCP) options as you prefer.

User access and authentication are the same as for the other protocols when you use the configured authentication method.

This completes the basic implementation.

12

Antivirus feature

This chapter describes the antivirus feature that is built into the IBM Storwize V7000 Unified Disk System. We also describe how the system interacts with the optional external scan engines for virus protection, what configuration options exist, and how to set up that feature.

12.1 Overview

In general, the antivirus protection is intended for Microsoft Windows and Common Internet File System (CIFS) users who require an extra level of data protection, such as against malicious, virus-type software. Because UNIX environments are much less exposed, there is no antivirus protection required, typically. Consequently, this antivirus feature of the Storwize V7000 Unified system is not supported for Network File System (NFS).

The Storwize V7000 Unified system includes an antivirus connector interface to communicate with external scan engines. It offers these options:

- Antivirus protection for data in CIFS file shares only
- McAfee and Symantec antivirus products are supported
- Scalability and high availability of the antivirus scans can be achieved by defining multiple scan nodes
- Scan on individual files: on file open command
- Scan on individual files: on file close command (after creation or modification)
- Scheduled scans on all files with these configurable options:
  - Manually (for example, after update of antivirus software or the known virus signatures)
  - On a defined schedule

The co-operation between the Storwize V7000 Unified system and the external antivirus scan engines is shown in Figure 12-1.

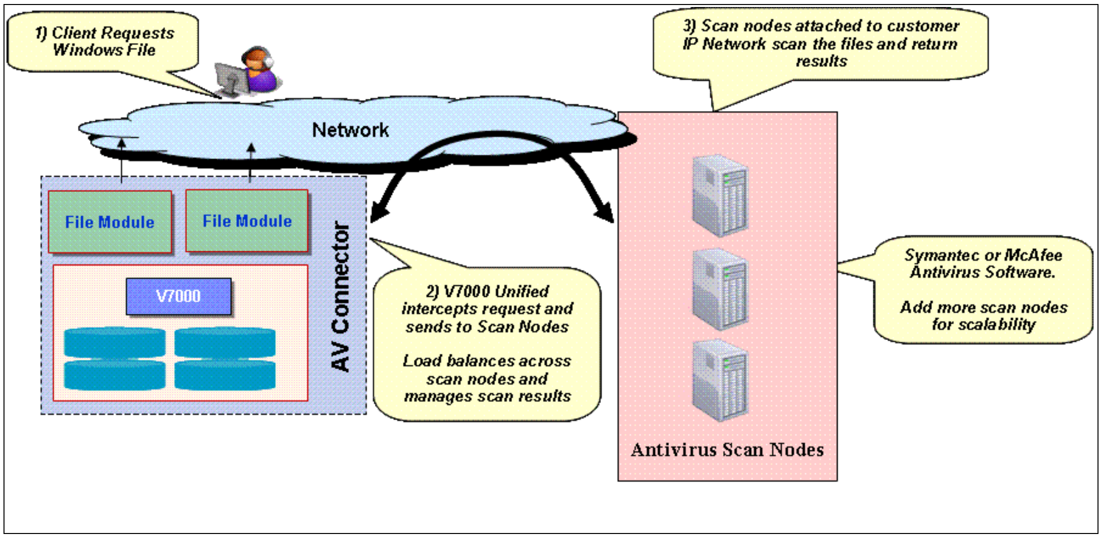

1) Client Requests Windows File
3) Scan nodes attached to customer IP Network scan the files and return results
Network
File Module
File Module
V7000
AV Connector
2) V7000 Unified intercepts request and sends to Scan Nodes
Load balances across scan nodes and manages scan results
Symantec or McAfee Antivirus Software.
Add more scan nodes for scalability
Antivirus Scan Nodes

Figure 12-1 The V7000 Unified system intercepts I/O operations to enable an antivirus scan first

If the antivirus option is configured, the antivirus connector in the Storwize V7000 Unified system intercepts the defined I/O operations on individual files (on open for read, if configured or on close after write). It then sends a scan request to the defined external scan nodes. These use a stored file signature to verify whether the file needs to be scanned. This is required if the signature of a file changed, either because the file changed (which can be a sign of an infection) or an antivirus software update invalidated the associated signatures of all files. In this case, the scan engine scans the file to ensure that there is no infection and update the file’s signature after completing the scan successfully.

12.2 Scanning individual files

For individual files, there is a mandatory scan on file open and a configurable option to scan the file on close, after write actions. The antivirus settings use an inheritance model, which means that the settings are also applied to all subdirectories of the specified path. For individual file scans, the scope of files to be scanned can be configured by using an inclusion list or exclusion list. It is also possible to specify whether access to a file is to be denied if it cannot be scanned. For example, access to a file can be denied if no scan nodes are currently available or if there is no further bandwidth to be able to scan this file.

The antivirus connector intercepts the file open (for read) and close (after write, if configured) in Samba software:

- Scan on file open for read action:
  - Only if the file changed or the virus signature changed since the last scan
  - Result of last scan (signature) is stored in extended attributes of a file and includes this information:
    - Time of last scan
    - Virus signature definition used
- Optional: Scan on file close after write action:

  This is a proactive scan operation. It might improve performance of the next file open for read if no other changes occur in the meantime.
- Optional: Deny access if file scanning is not available (increases security).

The performance requirements for the file scans determine the required scalability:

The ability to scan the files needs to be fast enough on open for the read action to determine the bandwidth and that number of scan nodes that are required.

Note: After scanning a file when it is opened, the next scan can occur only when the file is closed, if it is configured accordingly. If there are simultaneous read and write accesses to the same file with byte-range locking while it continues to stay open, the safety of the individual write updates cannot be guaranteed by the antivirus scan. A write process might write suspicious data into the file, which a subsequent read might pick up unscanned and unprotected while the file is still open. For example, this is the case for a VMware VMDK file while the virtual system is running.

12.3 Scheduled scan

In addition to the scan operations on individual files, a scheduled scan (bulk scan) of all files in a specified path can be configured. This can be an entire file export or an export of just a subdirectory tree from within the file. The scan then includes all files and subdirectories of the path specified. All of the files and subdirectories in the path will then be rescanned after the antivirus vendor updates its software or the virus signatures.

This proactive, scheduled scan eliminates the need for all files to be scanned on first access, when users request them and are waiting for the files. A scheduled scan provides the following benefits:

- Helps to mitigate impact on performance on file access after an antivirus virus signature update, for example, when everyone logs in the next morning
- Updates the interval of virus signatures, and the antivirus software determines the required scan interval, which might be required every night
- Stores the result of a scheduled scan in the extended attributes (EAs) of the files, which means that there is no need to scan the file on first access if its signature did not change.
- Verifies the file signatures and EAs on every file that is open to provide assurance that there is no change to the file since the last scan

Important: When using hierarchical storage management (HSM) and antivirus bulk scans, a bulk scan does not rescan files that are migrated off of the file system by using HSM. This means that no file recall is required, which preserves the results of the HSM policies defined. Scanning a file updates its *last access time* property.

12.4 Set up and configure antivirus

In this section, we describe the steps and options available to set up and configure the antivirus feature according to specific needs. The options are described by using the graphical user interface (GUI), but the setup can also be done by using the command-line interface (CLI).

The infrastructure for the antivirus scanning process is provided outside of the Storwize V7000 Unified system. This infrastructure depends on the following prerequisites being handled by the IBM client's staff:

- The client supplies and maintains scan nodes.
- The client installations that support the antivirus vendor product on scan nodes:
  - Symantec
  - McAfee
- The scan nodes are attached to the client's IP network and can communicate with the V7000 Unified system.

There are other important considerations. For instance, availability and scalability are achieved by provisioning multiple scan nodes:

- During antivirus setup, a pool of scan nodes is configured to be used in the V7000 Unified system.
- Each antivirus connector instance randomly chooses a scan node from this pool.
- If a scan node fails, it is temporarily removed from the pool, and these checks follow:
  - The antivirus connector checks regularly to see whether the node is back online.
  - If it is online again, it is automatically re-added to the pool of available scan nodes.

12.4.1 Antivirus setup steps

Navigate to the antivirus configuration panels by selecting **Files** → **Services** → **Antivirus** panels, as shown in Figure 12-2.

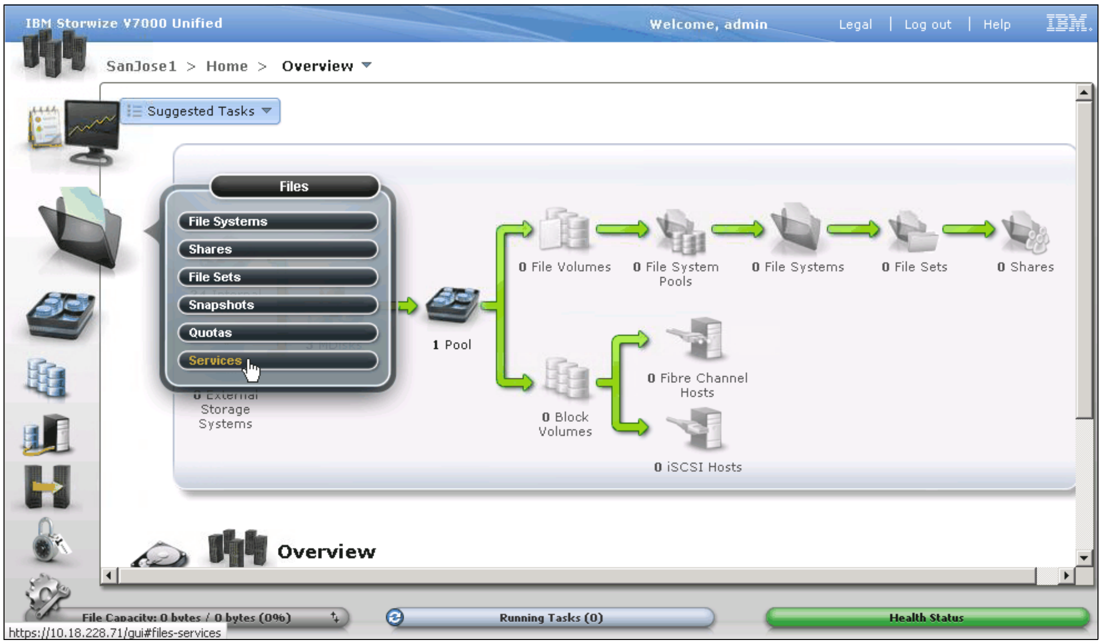
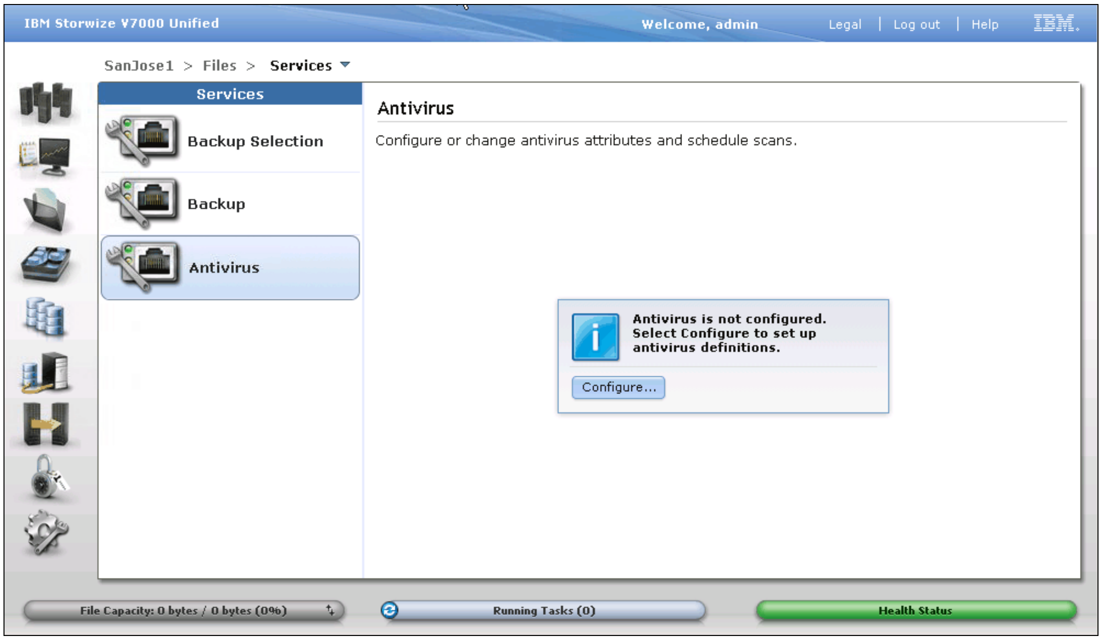


Figure 12-2 Menu path for Antivirus: Files → Services

Figure 12-3 shows the available services. Antivirus is already selected but, in this example, it is not configured yet.



Antivirus

Configure or change antivirus attributes and schedule scans.

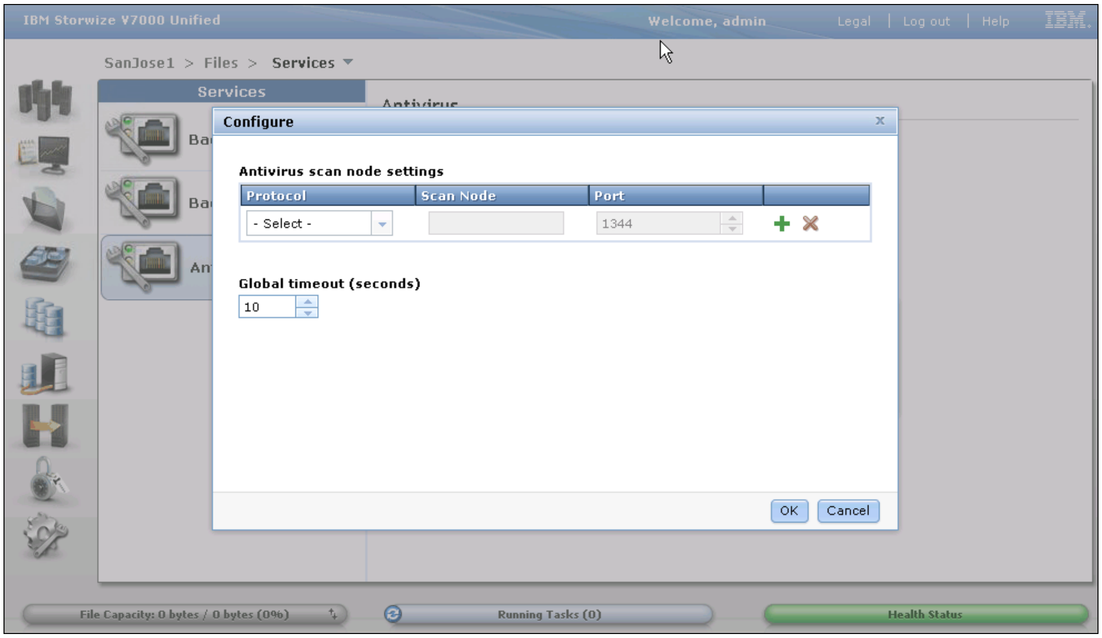
Antivirus is not configured. Select Configure to set up antivirus definitions.

Figure 12-3 Available services and antivirus selection

After selecting **Configure**, specify these three settings according to your needs:

- List of available scan nodes
- Scan protocol that they are using
- Global timeout for every scan

The corresponding GUI panels are shown in Figure 12-4 and Figure 12-5.



Configure

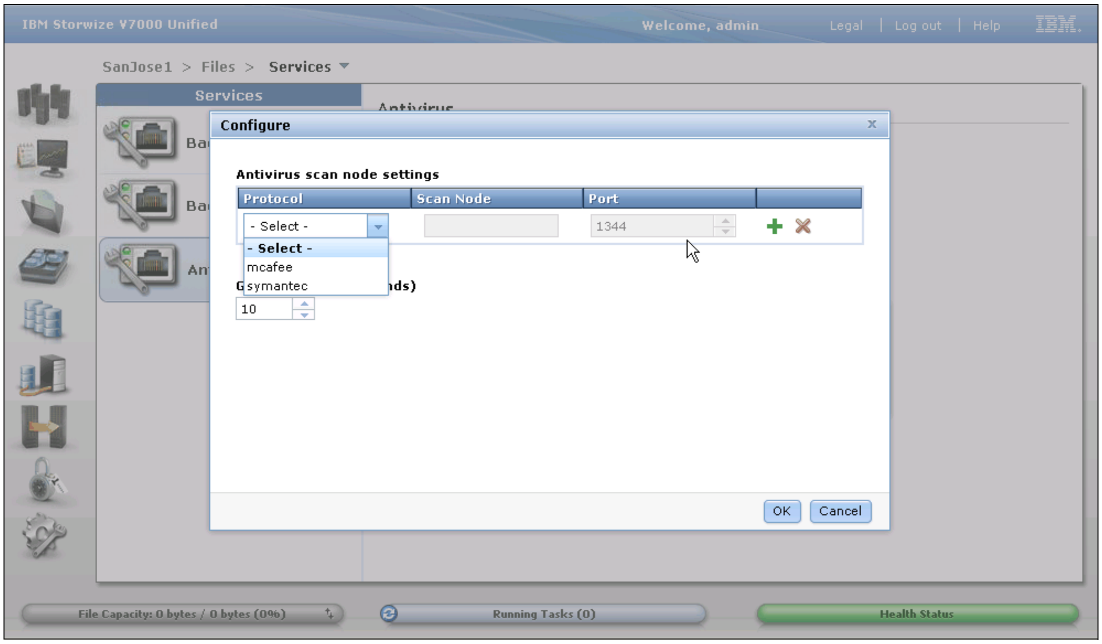
Antivirus scan node settings

Protocol	Scan Node	Port	
- Select -		1344	

Global timeout (seconds): 10

Figure 12-4 Definition of scan nodes, protocols, and timeout value

Figure 12-5 shows the selection of supported scan protocols.



Configure

Antivirus scan node settings

Protocol	Scan Node	Port	
- Select -		1344	

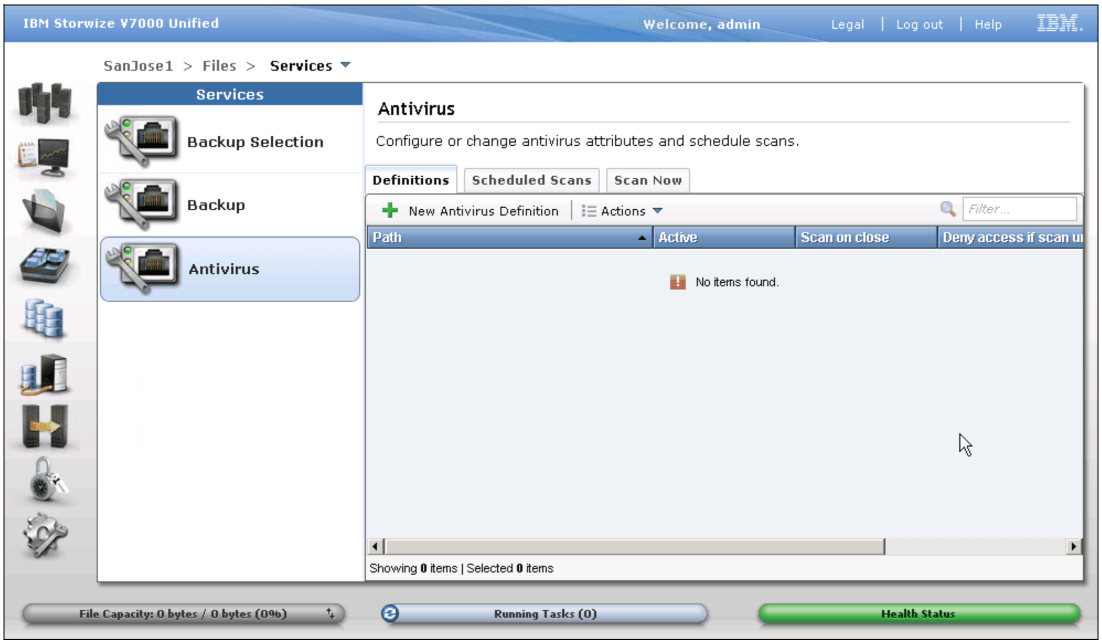
Global timeout (seconds): 10

Figure 12-5 Selection of supported scan protocols

After saving the scan node settings, the main Antivirus window opens, showing three tabs: Definitions, Scheduled Scans, and Scan Now. See Figure 12-6.

The settings for the scan nodes are not displayed on the main window. They are shown or can be changed by using the **Actions** drop-down menu.

Figure 12-6 shows the main window.



Antivirus

Configure or change antivirus attributes and schedule scans.

Path	Active	Scan on close	Deny access if scan u

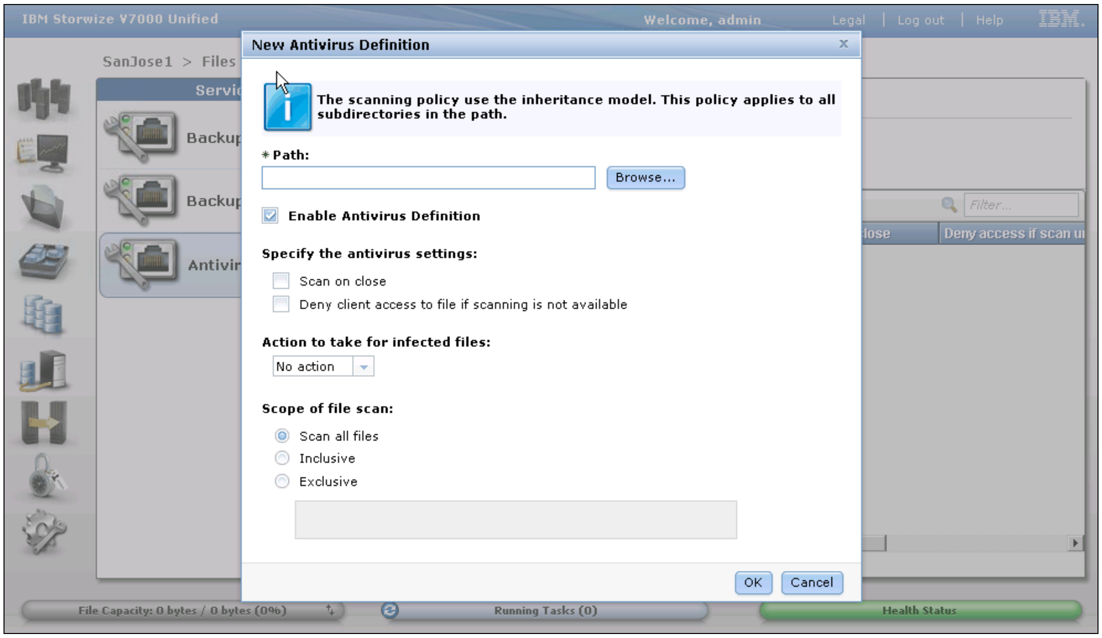
No items found.

Showing **0** items | Selected **0** items

Figure 12-6 Antivirus window that shows the three main tabs: Definitions, Scheduled Scans, Scan Now

Selection of adding a new antivirus definition opens the next window, where you can specify multiple options, as shown in Figure 12-7 on page 224:

- Specify the path to be scanned (browse for available paths)
- Enable or disable each antivirus definition
- Enable scan of files on close after write, if required (provides extra security)
- Deny client access if the file cannot be validated
  - Provides extra security
  - Prevents access even if only the scan nodes are not available or cannot be reached
- Specify which default action to take for infected files (No action, Delete, or Quarantine)
- Define the scope of the scan by choosing one of these options:
  - All files, include files with extensions specified (scan just these)
  - Exclude files with extensions specified (scan all others)



New Antivirus Definition

The scanning policy use the inheritance model. This policy applies to all subdirectories in the path.

* Path:

☑ Enable Antivirus Definition

Specify the antivirus settings:

□ Scan on close

□ Deny client access to file if scanning is not available

Action to take for infected files:

No action

Scope of file scan:

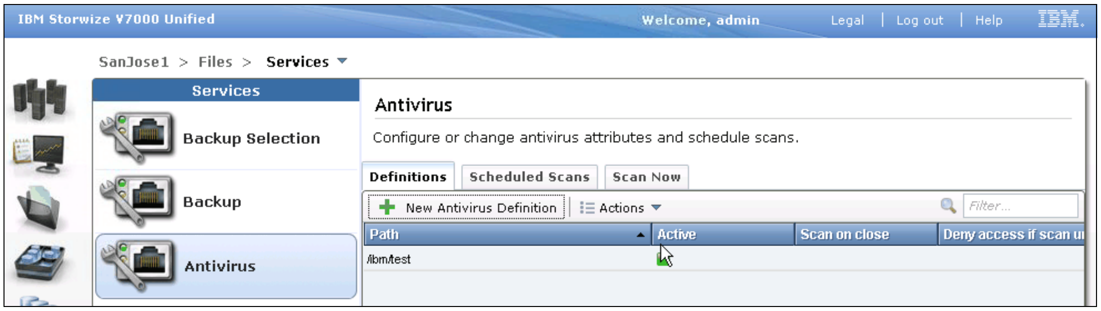
● Scan all files

○ Inclusive

○ Exclusive

Figure 12-7 Create new antivirus definition: Configurable options

When a new definition is created, it is shown on the main Antivirus window (see Figure 12-8).



Antivirus

Configure or change antivirus attributes and schedule scans.

Path	Active	Scan on close	Deny access if scan u
/ibm/test			

Figure 12-8 List of stored antivirus definitions

Beside the antivirus definition, use the **Scheduled Scans** tab to define your scans.

This option can be set up in the corresponding **New Schedule** tab, as shown in Figure 12-9 on page 225.

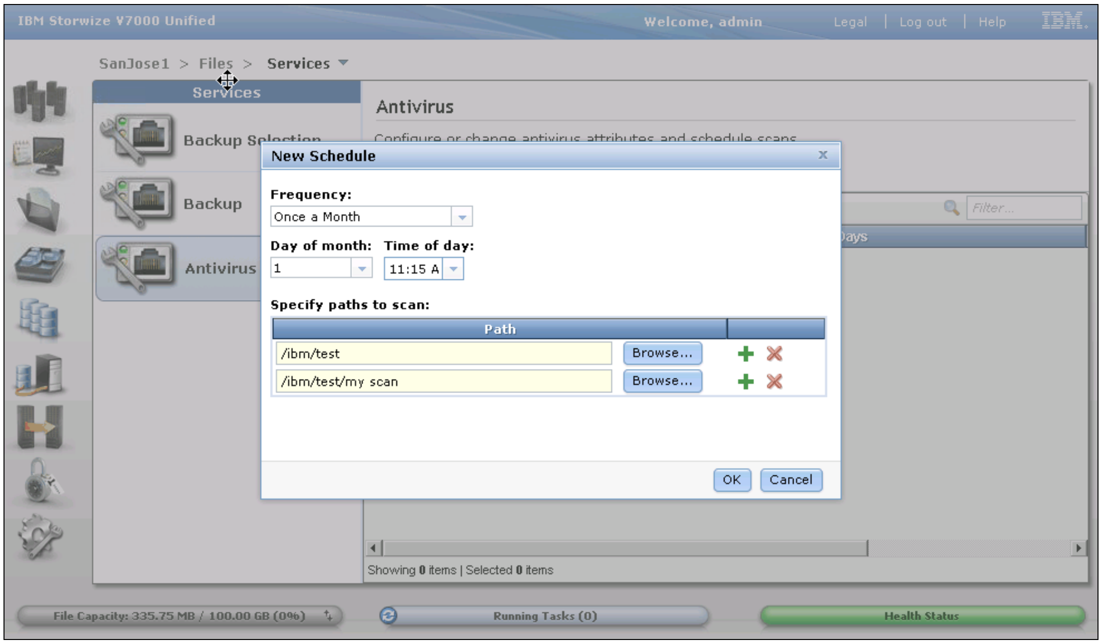
The following options are available for scheduled scans:

- **Frequency:** Choose Once a Day, Multiple Days a Week, Multiple Days a Month, Once a Week, or Once a Month.

> **Note:** Because of the usual change rate for virus definitions of at least once a day for all major vendors of antivirus software, the most common setting for this is expected to be *Once a Day*.

- **Time of day:** Presets have increments of 15 minutes, but any value can be entered.
- **Paths to scan:** Specify the path to use for this scheduled scan definition.

Figure 12-9 on page 225 shows how to define a new scheduled scan.


IBM Storwize V7000 Unified
Welcome, admin
Legal | Log out | Help
IBM.
SanJose1 > Files > Services
Services
Backup Selection
Backup
Antivirus
Antivirus
Configure or change antivirus attributes and schedule scans
Filter...
Days
Browse...
Browse...
OK
Cancel
Showing 0 items | Selected 0 items
File Capacity: 335.75 MB / 100.00 GB (0%)
Running Tasks (0)
Health Status

New Schedule

Frequency: Once a Month

Day of month: 1

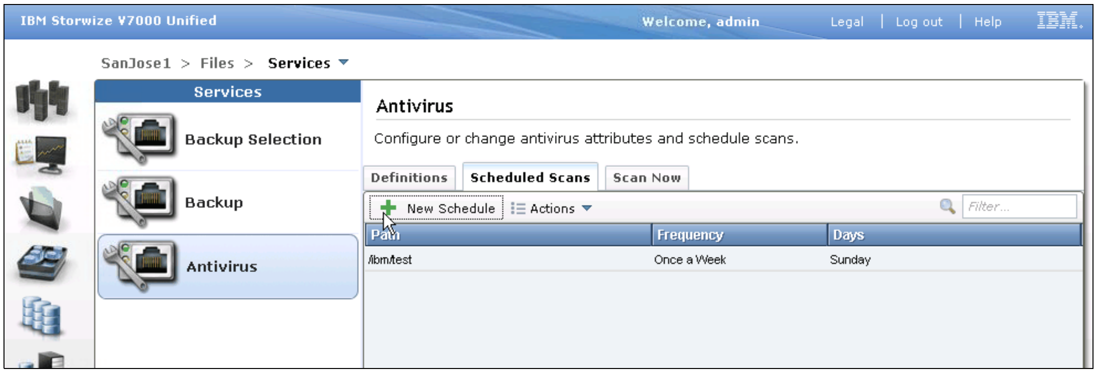
Time of day: 11:15 A

Specify paths to scan:

Path
/ibm/test
/ibm/test/my scan

Figure 12-9 Definition of new scheduled scan

The scheduled scans that are defined are listed on the main page for scheduled scans, which is shown in Figure 12-10.



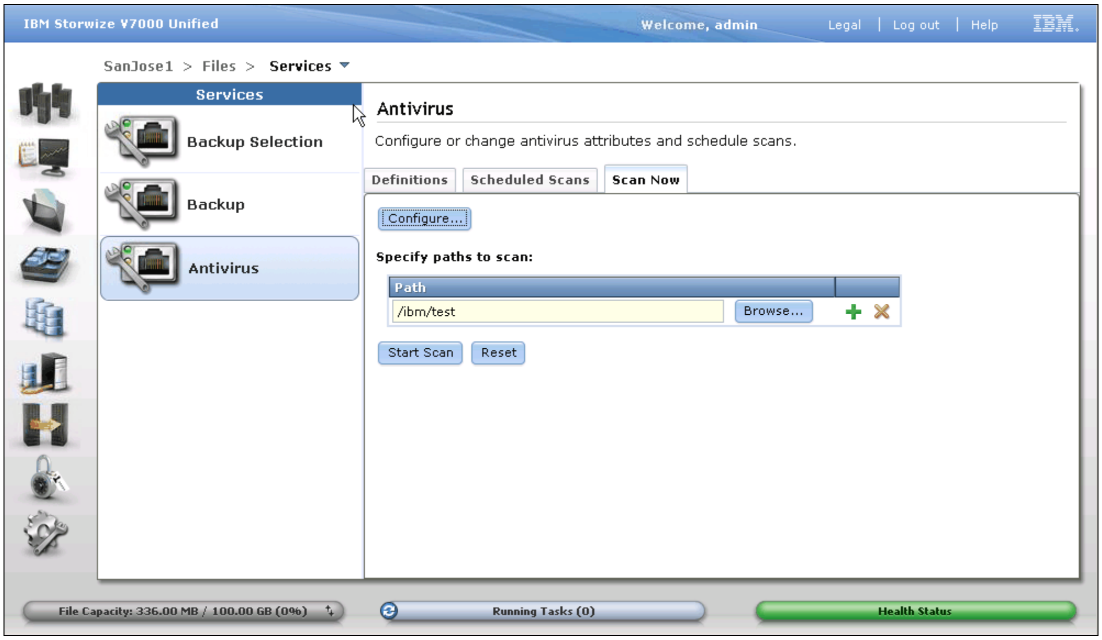
Antivirus

Configure or change antivirus attributes and schedule scans.

Path	Frequency	Days
/ibm/test	Once a Week	Sunday

Figure 12-10 List of stored scheduled scan definitions

To do an immediate scan, you can specify the path in the Scan Now option, as shown in Figure 12-11 on page 226.



Antivirus

Configure or change antivirus attributes and schedule scans.

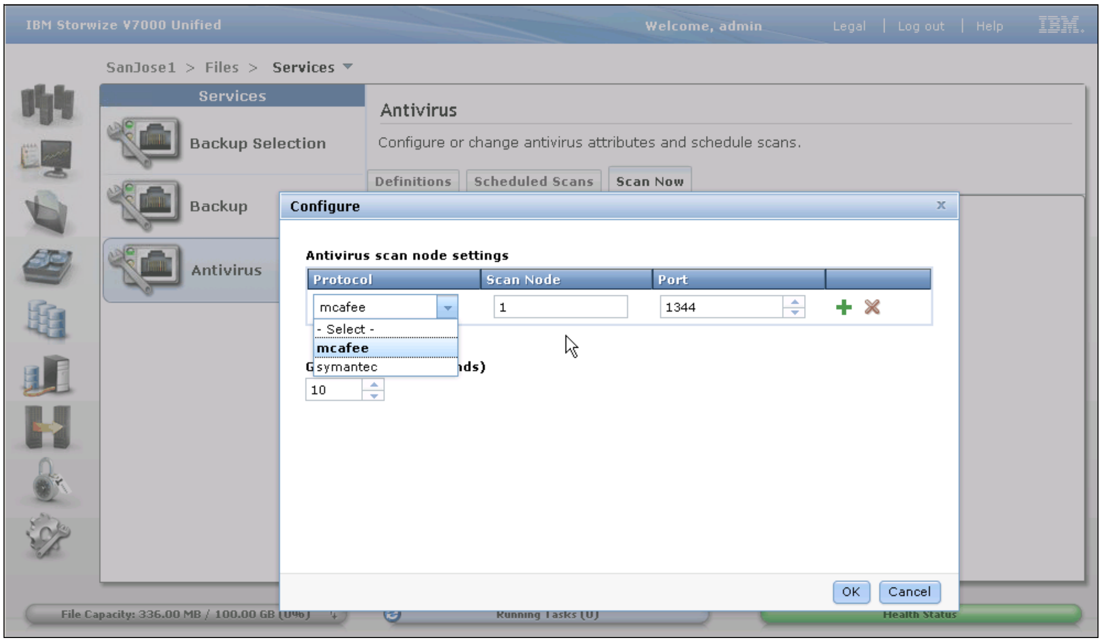
Specify paths to scan:

Path
/ibm/test

File Capacity: 336.00 MB / 100.00 GB (0%)

Figure 12-11 Scan Now

With the Configure option under the Scan Now tab, you can change the protocol that you want to use and the node, port, and timeout value of the search engine, as shown in Figure 12-12.



Antivirus

Configure or change antivirus attributes and schedule scans.

Antivirus scan node settings

Protocol	Scan Node	Port
mcafee	1	1344

G ds)

10

File Capacity: 336.00 MB / 100.00 GB (0%)

Figure 12-12 Scan Now Protocol

After you choose the options, click **OK**. You now have the option to either Scan Now or Reset, as shown in Figure 12-11.

Antivirus setup is now complete. We suggest that you test the antivirus scan engine to ensure that the operation is correct and operates as expected.

13

Performance and monitoring

In this chapter, we describe how to tune the IBM Storwize V7000 Unified Disk System for best performance. We then explore how to monitor performance of the system and describe the approach to identifying and rectifying performance problems.

IBM Real-time Compression: If you are using or considering compression, we strongly suggest that you consult the IBM Redbooks publication titled *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859, especially the chapter on performance guidelines:

`http://www.redbooks.ibm.com/redpieces/abstracts/redp4859.html`

13.1 Tuning the Storwize V7000 Unified system for performance

In the topics that follow, we describe how to tune the Storwize V7000 Unified Disk System for better performance.

13.1.1 Disk and file-system configuration

In this topic, we describe disk and file system configuration.

GPFS block size

Writes from applications are often smaller than the IBM General Parallel File System (GPFS) block size, which defaults to 256 K. However, for sequential accesses, GPFS combines the individual writes from an application and sends a single write request to the underlying Storwize V7000 subsystem for the full GPFS block size. The Storwize V7000 block storage system then stripes the data as it writes it to the underlying managed disk (MDisk). The stripe size can be configured to either 128 K or 256 K; the default is 256 K.

One of the main objectives when deciding on the number of physical drives that make up an MDisk (also known as the array size) is whether it increases the likelihood of achieving full-stride writes. A *full-stride write* in one in which all of the data in the disk array is written in a single chunk. This ensures that the parity is calculated by using only the new data. This avoids the need to read in existing data and parity for the current stripe and combine it with the new data.

With the MDisk example in Figure 13-1, achieving a full-stride write requires the GPFS block size to be significantly larger than 256 K. In this example, it needs to be 1 MB in size (8 x 128 K). For this reason, in addition to the 256 K default, 1 MB and 4 MB GPFS block sizes are supported.

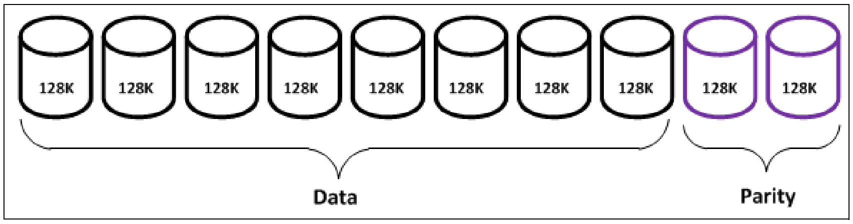


Figure 13-1 MDisk configuration

We suggest using a 1 Mb or 4 MB GPFS block size *only* on sequential workloads. The reason is that a subblock, which is defined as 1/32 of a full GPS block, is the minimum size of data that GPFS can allocate space for. Therefore, using a 1 MB or 4 MB block size on random workloads, where GPFS is unable to combine writes, is likely to result in the subblock being significantly larger than the size of data to be written. This results in poor space use and reduces the effectiveness of caching, so it negatively affects performance.

Number of volumes

The number of volumes that are used to back the GPFS file system can also affect performance. This is because each volume has its own fixed queue depth. Therefore, with only a few volumes, there is risk of I/O operations getting backed up within the GPFS.

We suggest that if greater than or equal to 50% of the total MDisk group capacity is allocated to a file system, you create the number of volumes that is two times the number of MDisks. For example, if 60% of the total MDisk Group capacity is allocated to the file system, and the MDisk Group contains 11 MDisks, create 22 volumes.

If less than 50% of the total MDisk group capacity is allocated to the file system, create the number of volumes that equals the number of MDisks. We suggest that you back a GPFS file system with no fewer than three volumes.

Separation of data and metadata

Solid-state drives (SSDs) can achieve two orders of magnitude more I/O operations per second (IOPS) than traditional spinning disk technology. For example, Near-Line SAS drives achieve approximately 100 - 150 IOPS, whereas enterprise grade SSDs can deliver 10,000 IOPS. However, this extra performance comes at a financial cost, so SSDs need to be used where they offer the greatest benefit.

> **Important:** Spread SSDs across both of the Storwize V7000 chains, so that half are connected though Port 1 and the other half are connected through Port 2. This ensures that the I/O that is generated by the SSDs is serviced equally by both available ports.

To make best use of SSDs, the Storwize V7000 system includes the IBM EasyTier feature. This technology uses the fact that, in general, only a small percentage of the total stored data is accessed frequently. EasyTier automatically identifies this frequently accessed data (known as *hot* data) and migrates it onto the SSDs. This ensures that you can achieve a significant performance boost with only a small number of SSDs.

The I/O originating from the Unified system component is stripped by GPFS onto the underlying Storwize V7000 volumes. It balances I/O accesses to avoid the *hot* and *cold* variation seen in traditional block workloads. Therefore, enabling EasyTier on volumes that back GPFS file systems provides no performance benefit. Instead, it is better to reserve SSDs for the storage of metadata.

Each file in a file system has metadata associated with it. This includes name, size, permissions, creation time, and details about the location of the data. The latency of accessing this type of data can have a disproportionate effect on the performance of the system, especially if the file system contains many small files. Snapshots, backup, and asynchronous replication all require scanning metadata. Therefore, we suggest that you use SDDs when you use these features extensively.

Metadata typically accounts for 5 - 10% of the total space requirements of a file system, so it is cost-effective to put it on SSDs.

SMB configuration

If a file system is to be accessed only by using the Server Message Block (SMB) protocol, the multi-protocol interoperability support is not required. In most cases, disabling this support provides a noticeable improvement in performance.

There are three parameters that provide multi-protocol interoperability support: *leases*, *locking*, and *sharemodes*. Use leases to indicate whether an SMB client is informed if another client accesses the same file at the same time by using a different protocol. Use locking to indicate whether to perform a check before granting a byte range lock to an SMB client. Use sharemodes to indicate whether the share modes are respected by other protocols. (SMB uses share modes to enable applications to specify whether simultaneous access to files is permitted.) All three parameters can be set by using the command line, either during creation of the share (by using the **mkexport** command) or after creating the share (by using the **chexport** command). For example:

```
chexport exportName --cifs “leases=no,locking=no,sharemodes=no”
```

Important: Do not use those SMB settings when the file system is accessed through other protocols, such as NFS. Doing so can corrupt data.

13.1.2 Network configuration

We describe two network configuration considerations.

10 GBit versus 1 GBit interface

For best performance, we strongly advise that clients connect to the Storwize V7000 Unified system by using the 10 GBit interface.

Bonding modes

Storwize V7000 Unified system supports various bonding modes. The default is Mode 0, referred to as *active-backup*. In this mode, only one of the two links is active at anyone time. This is the simplest of the bonding modes, and is designed solely to provide resilience in a link failure.

Although the other modes provide potentially higher throughput than active-backup (because they use both links), they are more complex in nature and have a high probability of reacting adversely to the connected Ethernet switches and associated network infrastructure. Therefore, we strongly suggest that you keep Mode 0 as the default.

13.1.3 Client configuration

The clients that are connected to the Storwize V7000 Unified system play a critical role in determining the performance of the system. For example, if the client is unable to send enough requests to the Storwize V7000 Unified system, performance is constrained by the client. Not only is it important to ensure that there are enough clients to generate the necessary load, but it is also important that the clients are correctly configured.

NFS clients

NFS clients that are commonly used within a UNIX environment offer many parameters on how to connect the NFS File Server. One such parameter that has a major effect on determining overall performance is the use of the *sync mount* option.

Both the client and server side have a sync parameter. The server-side sync is enabled on the Storwize V7000 Unified system. We strongly suggest that you do not change this setting. It ensures that, when receiving a write or commit from the NFS client, the data is stored in the Storwize V7000 subsystem before sending an acknowledgement. This behavior is required when strictly adhering to the NFS protocol.

The client-side sync parameter defines whether each NFS write operation is sent immediately to the V7000 Unified system or whether it remains on the client until a flush operation is performed, such as a commit or close. Setting the sync parameter comes at a significant cost to performance and, often, the extra resilience is not required. Applications that demand full synchronous behavior must open files by using the **O_SYNC** option, which forces the synchronous behavior regardless of the client-side sync parameter setting.

SMB clients

There are two main versions of the SMB protocol. Version 2 of the protocol, introduced in 2007, significantly reduces the number of separate commands and subcommands compared to the original protocol. This significantly reduces the frequency of communication between the client and the file server. Therefore, it offers a significant performance benefit.

By default, the Storwize V7000 Unified system attempts to negotiate the use of Version 2 of the SMB protocol. However, if the client machine is unable to use this later version, the Unified reverts to the earlier and less optimal Version 1. To ensure that Version 2 is used, it is essential that the client machines are running either Microsoft Windows Server 2008 or later or Microsoft Windows 7 or later.

13.2 Monitoring performance of the system

The performance of the V7000 Unified system can be monitored through both the GUI and CLI. You can also use the IBM Tivoli Storage Productivity Center to monitor performance.

13.2.1 Graphical performance monitoring

The Storwize V7000 Unified system provides a quick-look set of graphs that show functions for file and block performance. These graphs have different collection time frames, with block one showing only 5 minutes. They are intended to highlight problem areas in real time and are useful in narrowing down the search during sudden and critical performance hits.

To display the graphs, select **Monitoring** → **Performance**. These three tabs are available for performance monitoring:

- File
- Block
- File Modules

The *File* performance graphs show metrics about client throughput, file throughput, latency, and operations. The time frame for these graphs can be selected from minutes to a year. See Figure 13-2 on page 232 for details.

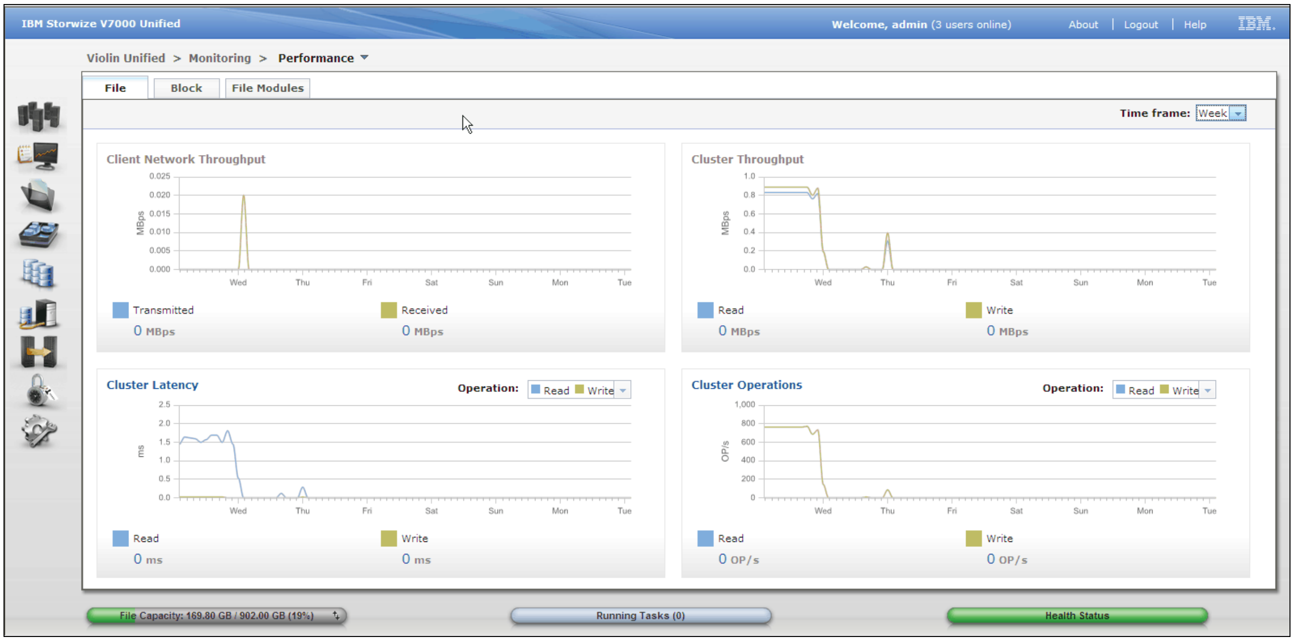


Figure 13-2 File performance

The *Block* performance graphs show real-time statistics that monitor CPU use, volume, interface, and managed disk (MDisk) bandwidth of your system and nodes. The CPU use graphs also show separate system and compression statistics. Each graph represents 5 minutes of collected statistics and provides a means of assessing the overall performance of your system. See Figure 13-3 for details.

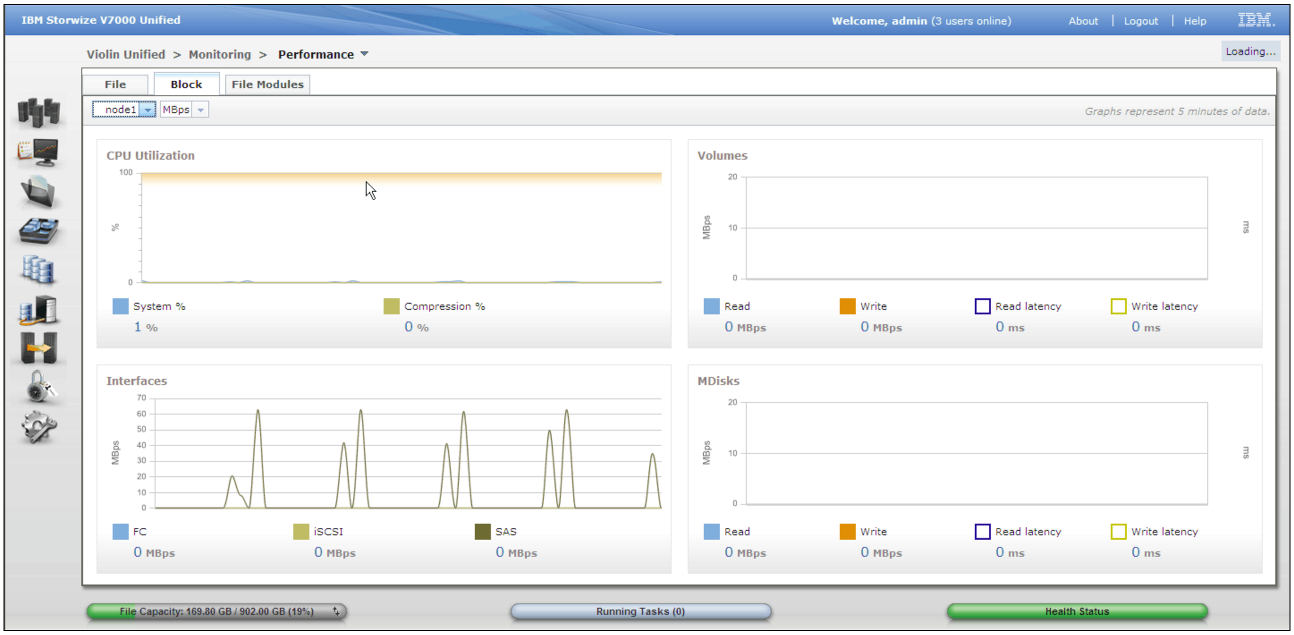
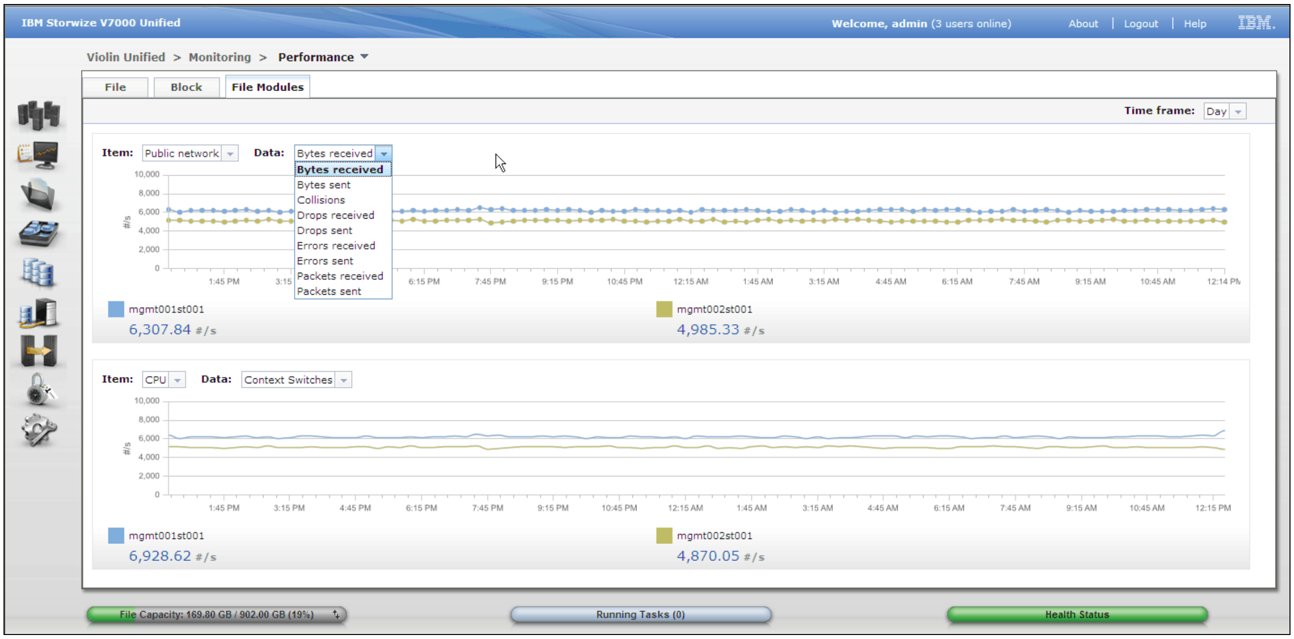


Figure 13-3 Block performance

The *File Modules* performance graphs show the performance metrics for each file module for CPU, memory, and public network. You can also select numerous data types for each item selected, such as collisions, drops, errors, and packets for public network statistics and other related data types for CPU and memory statistics. The time frame for these graphs can be selected from minutes to a year. See Figure 13-4 for details.



Time frame: Day

Item: Public network Data: Bytes received

mgmt001st001 6,307.84 #/s

mgmt002st001 4,985.33 #/s

Item: CPU Data: Context Switches

mgmt001st001 6,928.62 #/s

mgmt002st001 4,870.05 #/s

Figure 13-4 File modules performance

The block storage system is continuously logging raw statistical data to files. Depending on the setting for the statistics interval (default = 5 minutes) all of the statistical counters are collected and saved regularly. Only the last 15 files are kept, and the oldest ones are purged on a first-in last-out (FILO) basis. Tivoli Storage Productivity Center and other performance tools collect these raw files. These stats files are also collected by the support data collection, which collects all files from the last 15 saves. IBM technical support uses this data, if required, to analyze problems.

13.2.2 Command-line interface (CLI) performance monitoring

In addition to the performance graphs, it is possible to monitor performance by using the CLI. The **`lsperfdata`** command provides access to a range of file performance metrics, including the number of file operations that are performed on the cluster. Data is available over a range of time periods, ranging from seconds to a year. Use the **`lsperfdata`** command with addition of the `-l` parameter to list all performance metrics available.

The block performance of the Storwize V7000 subsystem can also be monitored by using the CLI. The **`lssystemstats`** command provides a range of metrics about the performance of the underlying block storage, including the throughput and latency of the volumes, MDisks, and physical drives.

13.2.3 Tivoli Storage Productivity Center

IBM Tivoli Productivity Center now includes the Storwize V7000 Unified storage system. The Tivoli center collects and tracks data from the cluster and can provide details about demand. The areas that you can drill down to include general cluster information, nodes, pools, Network Share Disks (NSDs), file systems, file sets, exports, and selected directories.

Configuration

Tivoli Storage Productivity Center does not use a Common Information Model (CIM) agent, but it can natively connect to the Storwize V7000 Unified system. To configure the cluster to communicate with the center requires one simple step: Create a user ID with full authority for the Storage Productivity Center to use. Assign a password and give the user ID and password details to the Tivoli Storage Productivity Center administrator. All of the remaining configuration is done from Tivoli software, and the configuration details are retrieved and built automatically.

If specific file system scanning is required, you might need to create and export a share to host running the Tivoli Storage Productivity Center agent.

13.3 Identifying and resolving performance problems

As shown in Figure 13-5, there are many components that need to be considered when you are analyzing a performance problem.

At the top of the *storage stack* is the client application that initiates the I/O requests and the client server (also known as the *host*), which is responsible for sending the I/O requests to the Storwize V7000 Unified system.

The next level is the networking, which includes any interconnected switches and routers. Issues at this level usually present in the form of high packet loss rates observed at either the host or the Storwize V7000 Unified system. See 13.3.2, “Network issues” on page 235 for more details about how to identify and resolve issues at this layer.

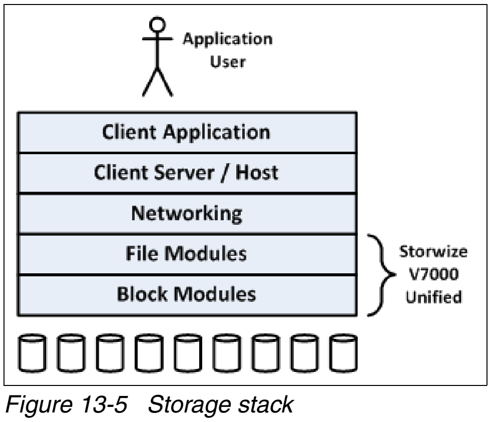


Figure 13-5 Storage stack

Performance issues might also result due to the overuse of the resources in either the file modules or block modules. Identifying and resolving issues at this bottom layer of the storage stack is described in Section 13.3.3, “High latencies” on page 235.

13.3.1 Health status

The health status indicator at the bottom of the V7000 Unified system GUI window, the alerts, and the event logs provide the first point of investigation for a problem. This is described in Chapter 15, “Troubleshooting and maintenance” on page 257.

Issuing the `lshealth` command to expand in detail an unhealthy system can be useful. By adding the `-i` parameter, more detail can be seen on individual components.

> **Important:** It is important to address all issues and maintain the health status of your system as green to prevent issues such as volumes going offline because of lack of available physical storage.

13.3.2 Network issues

Poor performance is often a result of a problem with the network connecting the hosts to the Storwize V7000 Unified system. Using the performance monitoring graphs, select the **File Modules** tab, and then use the drop-down menu to select **Public network**. Then use the **Data Item** to monitor *errors*, *drops*, and *collisions*.

If the data indicates an issue, examine the logs on the Ethernet switches and routers. In addition, if the default bonding Mode 0 (active-backup) is not being used, revert to this mode to determine whether that resolves the problem.

It is also useful to isolate any potential issues by connecting the hosts directly (or as directly as possible) to the Storwize V7000 Unified system. If performance improves when the intermediate networking components are excluded, it indicates that the problem is with how the Storwize V7000 Unified system is interacting with the additional network components.

13.3.3 High latencies

High latencies are often an indication of a bottleneck. The next step is to determine whether the latency is within the file layer or the block layer. Compare the cluster read and write latencies (which are provided as a graph on the File tab page), with the read and write latency of volumes and MDisks (which is provided as a graph on the Block tab page). If the cluster latencies are significantly higher than the volume and MDisk latencies, it indicates that the issue is within the file modules.

File modules

High latencies within the file layer might be due to creating insufficient volumes. See “Number of volumes” on page 228. Another possible cause is high use of the CPUs. This can be monitored by using the performance graphs. Select the **File Modules** tab and use the Item drop-down menu to select **CPU**. Then, use the Data drop-down menu to select **System**, **User**, **I/O Wait**, or **Idle**.

The CLI `lslog` command is useful for showing the activity on the box over the previous day or so. It lists the logs from the event log database and can be used to discover when activities such as backup and asynchronous replication started. It also contains messages that show when the CPU use on either file module crossed the 75% threshold. This indicates that the machine is under high use. If that continues, it can be detrimental to performance.

It is important to ensure that the resource use of the two file modules is balanced, with the file workload sent equally to both nodes. Also, advanced features, such as snapshots, backup, and asynchronous replication, can be resource-intensive. Therefore, it is important to ensure that they are scheduled so that the times when they are running do not overlap.

Block modules

High latencies within the block layer are often an indication that the drives are overloaded. Performance can be improved either by increasing the number of drives available or by upgrading to faster storage, such as SSDs.

14

Backup and recovery

In this chapter, we cover the backup and recovery process of the IBM Storwize V7000 Unified Disk System configuration data and the host user data that is written to the system. These are two distinctly different and unrelated areas, so we cover each one separately. We also describe the processes to restore the system from a backup.

14.1 Cluster backup

In the topics that follow, we describe Storwize V7000 Unified cluster backup.

14.1.1 Philosophy for file and block

The storage cluster is made up of two major components:

- The IBM Storwize V7000 Storage System (subsystem in this case)
- The IBM Storwize V7000 Unified Disk System file modules

Because each subsystem has its own existing and established backup processes with different requirements, they are backed up independently.

Storwize V7000 subsystem (block)

The primary level of backup is that the most current copy of the configuration and status is held by the other operational storage nodes in the cluster. Therefore, if a node fails or needs to join the cluster at any time, it receives all of the required configuration and status information from the config node.

Additionally, the config node is saving regular checkpoint data on the quorum disks. These saves occur about every 10 minutes. The three quorum disks have several functions, including tiebreaking in the event of a split cluster.

A full config XML file is generated every day at 01:00:00. This file is saved on the hard disk drive of the config node and also written to the quorum. This file contains the hardware and storage configuration, including MDisk, volume, and host mapping details, but it does not include the storage extent mapping tables. This backup can be manually triggered at anytime through the command-line interface (CLI).

File modules (file)

Again, like the storage, the Storwize V7000 Unified component relies on the plan that if one file module is operational and online, the other can be restored, even from a complete loss of all system data. The difference here is that the file modules are each unique, so the backup is usable only on the particular module where it was created.

A backup is made automatically at 02:22:00 every day on each file module. This backup is packaged into a single file and copied to the other file module and stored on its hard disk drive so that each file module has a copy of the other’s backup. Previous generations of backups are kept in case of corruption or a need to revert to an earlier version. This backup can be manually triggered at anytime through the CLI.

14.1.2 Storage enclosure backup (Storwize V7000 subsystem)

The storage component is backed up independently from the file modules and uses the inherent processes of the Storwize V7000. The backup process is designed to back up configuration information and the current status of your system, such as cluster setup, user ID setup, volumes, local IBM Metro Mirror information, local IBM Global Mirror information, managed disk (MDisk) groups, hosts, hosts mappings, and nodes.

Three files are created:

`svc.config.backup.xml` This file contains your current configuration data and status.

`svc.config.backup.sh` This file contains a record of the commands that are issued by the backup process.

`svc.config.backup.log` This file contains the backup command output log.

If you want an immediate backup, such as before or after a critical or complex change or before a support data collection, issue this CLI command:

`svcconfig backup` This starts a backup immediately. Successful completion is indicated by a return to the prompt without an error message. While the backup is running, any commands or processes that might change the configuration are blocked.

A manual backup is not normally required because the task runs daily at 01:00:00. It is needed only if changes are made or if the status must be reflected in the support data.

There are high-speed paths between all of the nodes that form the cluster. These paths can be over the SAN. With the Storwize V7000, there is also a high-speed internal bus in the enclosure between the node canisters that also carries these paths. This enclosure link removes the need for SAN paths between the nodes and allows the cluster to operate without a connection to the SAN if you prefer.

These links are used to maintain data integrity between the two nodes in an input/output (I/O) group and are also used to ensure that the cluster status and configuration are available to all nodes at all times. Any changes to the configuration, including the extent maps, are shared over these paths. This means that any node can assume the config role and have fully up-to-date configuration at any time. If a node is disconnected from the cluster (that is, from the other nodes) for any amount of time, it cannot continue handling data. Therefore, it must leave and rejoin to relearn the current configuration to maintain the integrity of the cluster. This takes place automatically.

The code always tries to use three MDisks as quorum drives if at least three are available. One is marked as the active drive. The Storwize V7000 reserves an area of each MDisk to contain the quorum data. You can modify which MDisks are used and which one is active. These MDisks are seen by all nodes in the cluster. The config node periodically writes checkpoint data to the quorums (about every 10 minutes). The quorums also serve other functions, including communication between nodes that are isolated so that a tie can be broken.

If a node is properly shut down, it saves hardened data and gracefully leaves the cluster to go offline. When an offline node starts, it searches for the active cluster to determine whether any other nodes from this cluster are already active and have formed the cluster. If so, it requests and can be granted permission to rejoin. If no cluster exists (that is, no node responds) and the node can see the quorums to confirm that the cluster is not active, this node assumes that it is the first one active and forms the cluster.

If a node fails or is unable to save hardened data for any reason, it has left the cluster and can never rejoin. Any node starting that was a cluster member but has no valid hardened data fails and posts a permanent error, typically 578. The definition for this node in the cluster must now be deleted and re-added as a new node.

If all nodes fail without an appropriate shutdown (that is, all nodes from the cluster failed, and no active nodes remain), the quorum data can be used to rebuild the cluster. This is a rare situation and is known as a *Tier 3 recovery*.

14.1.3 File module backup

A pair of 1 Gb Ethernet connections form the physical path over which a link exists between the two file modules. This link is used for communications and to maintain integrity of the file systems across the cluster. It also is used to break ties and for recovery actions, software loading, and transfer of backup data.

A backup process runs at 02:22:00 each day and creates a packaged single file. A copy of the backup file is then stored on the other node, so that each file module has a copy of the other nodes' backups. The file is given a unique name and stored on the file module in this directory: `/var/sonas/managementnodebackup`

This backup can be manually triggered at any time through the CLI by using this command:

```
backupmanagementnode -v
```

14.2 Cluster recovery

In the topics that follow, we describe Storwize V7000 Unified Disk System cluster recovery.

14.2.1 Storage enclosure recovery (V7000 subsystem)

As with the other products in this family (IBM SAN Volume Controller and Storwize V7000), recovery of a node or cluster from backup is defined in four levels that are known as *tiers*. These tiers are numbered in increasing level of impact and severity. For the Storwize V7000 Unified system, these are the functions of the four tiers:

Tier 1 (T1) Recovers from single failures of hardware or software without loss of availability or data

Tier 2 (T2) Recovers from software failures that occur on nodes, with loss of availability but no loss of data

Tier 3 (T3) Recovers from some double hardware failures, but potentially involves some loss of client data.

Tier 4 (T4) Assumes that all data that is managed by the cluster is lost and provides a mechanism to restore the cluster's configuration to a point that it is ready to be restored from an off-cluster backup (for example, tape backup)

The following descriptions are to assist you in understanding the effect of each recovery and give you an overview of the processes involved. In all cases where action is required, be sure to follow IBM documentation and IBM technical support guidance closely.

Tier 1

A Tier 1 recovery is where the cause of the problem can be resolved by warm-starting the node. The most common trigger for this is a hardware or software error that is detected by the storage node's software, which then triggers an *assert*. An assert is a software-initiated warm start of the node. It does not reboot the operating system, but restarts services and resumes the previous operational state. It also captures a dump for later analysis.

The warm start occurs quickly enough that no connectivity is lost and paths are maintained. No data is lost, and the cache destages when the node is recovered. There is not likely to be any effect on the file modules. All I/O operation is eventually honored. Most asserts are noticed only in the event log or by an alert that is posted.

Note: Report any assert to IBM Support for investigation.

Tier 2

The Tier 2 recovery level is much the same as for Tier 1, but the node failed and must be re-added. This process is automated. When the storage node reboots, the cluster performs recovery procedures to rebuild the node's configuration and add it back into the cluster. During this process, most configuration tasks are blocked. This recovery is normally indicated by a 1001 error code. This recovery takes about 10 minutes.

It is important to carefully follow the procedures that are explained in the information center and as directed by the maintenance procedures and IBM Support.

The Storwize V7000 storage component recovers and becomes fully operational again without intervention. No data is lost; the cache is recovered and destaged from the partner node. Any I/O being directed to the affected node at the time of failure would fail. Depending on multipath driver support on the hosts, this I/O either fails or is retried on the other node in the I/O group.

File services need manual intervention to recover. Wait for the Storwize V7000 subsystem to recover and display a message on the graphical user interface (GUI) console. Call IBM Support for assistance and guidance. Start with the event log and perform maintenance on the error. Ensure that you follow all directions and the information center carefully.

Recovery of the file services is likely to require reboots of the file modules, one at a time, and checking the status of file systems. Follow the IBM Support guidance and the online maintenance procedures. When the recovery is complete, the IBM Support representative asks for logs to confirm system health before advising your staff to release the change locks.

Tier 3

Tier 3 recovery is required if no storage nodes remain in the Storwize V7000 Unified cluster. The cluster is then rebuilt from the last good checkpoint, as stored on the quorum disks. This is a rare situation, but if it is required, direct assistance from IBM Support is necessary.

The Storwize V7000 storage component can be recovered by the user by following the procedures in the information center. IBM Support can provide assistance if needed.

When the Storwize V7000 storage component is recovered, direct IBM Support is necessary to recover the file services. This requires collection and transmission of log files and remote access to the CLI, which can be achieved by using IBM Assist On-site (AOS) or any acceptable process that allows support access. IBM Support representatives investigate the status of various components in the file modules and repair them, as needed. This process can take some time.

Note: It is important that no one performs any actions onsite without IBM Support guidance.

There is potential for data loss if data existed in a node's cache at the time of failure. All access to data on this cluster is lost until the recovery is complete.

Tier 4

This rare process is unlikely and is not automatically started. It is directly driven by IBM Support after all attempts to recover data fail or are ruled out. All user data is considered lost, and the Storwize V7000 Unified system is reinitialized and restored to the last known configuration.

On completion of the recovery process, the storage component has the MDisks, pools, and volumes defined. All previously mapped volumes are mapped to the host definitions. No user data in volumes is recovered, so it must be restored from backup.

The file modules are reloaded and reinitialized. The user must provide all of the configuration data that is used to initially build the system, which is used to rebuild the configuration to the same point. All file system configuration is also re-entered, including file systems, file sets, and shares.

When file services are resumed, user data can then be restored from a backup. The process that is used depends on the backup system in use. This is covered in section 14.3, “Data backup” on page 242.

14.3 Data backup

The file data that is stored on the cluster can be backed up, using conventional means, by the servers or a server-based backup system. Alternatively, the file systems that are mounted in the IBM General Parallel File System (GPFS) file system on the cluster can be backed up directly. Currently, two methods of data backup are supported for backing up the file systems on the Storwize V7000 Unified: Tivoli Storage Manager and Network Data Management Protocol (NDMP). Tivoli Storage Manager uses proven IBM backup and restore tools. NDMP is an open standard protocol for network-attached storage (NAS) backups.

> **Warning:** Do not attempt to use NDMP with hierarchical storage management (HSM) or Tivoli Storage Manager. This is not supported and can cause an outage.

14.3.1 Data backup philosophy

The following topics explain the data backup philosophy.

Server

This method reads the data from the cluster by using the same method that the servers use to read and write the data. The data might be read by the server that wrote it or by a stand-alone server that is dedicated to backup processes that have share access to the data. These methods are beyond the scope of this book.

Tivoli Storage Manager

The storage agent is included with the cluster software. When it is enabled, it runs on the file modules. When it is configured, Tivoli Storage Manager can be scheduled to back up file systems from the GPFS system by running on the file modules to external disk or tape devices. This data backup and management is controlled by Tivoli Storage Manager.

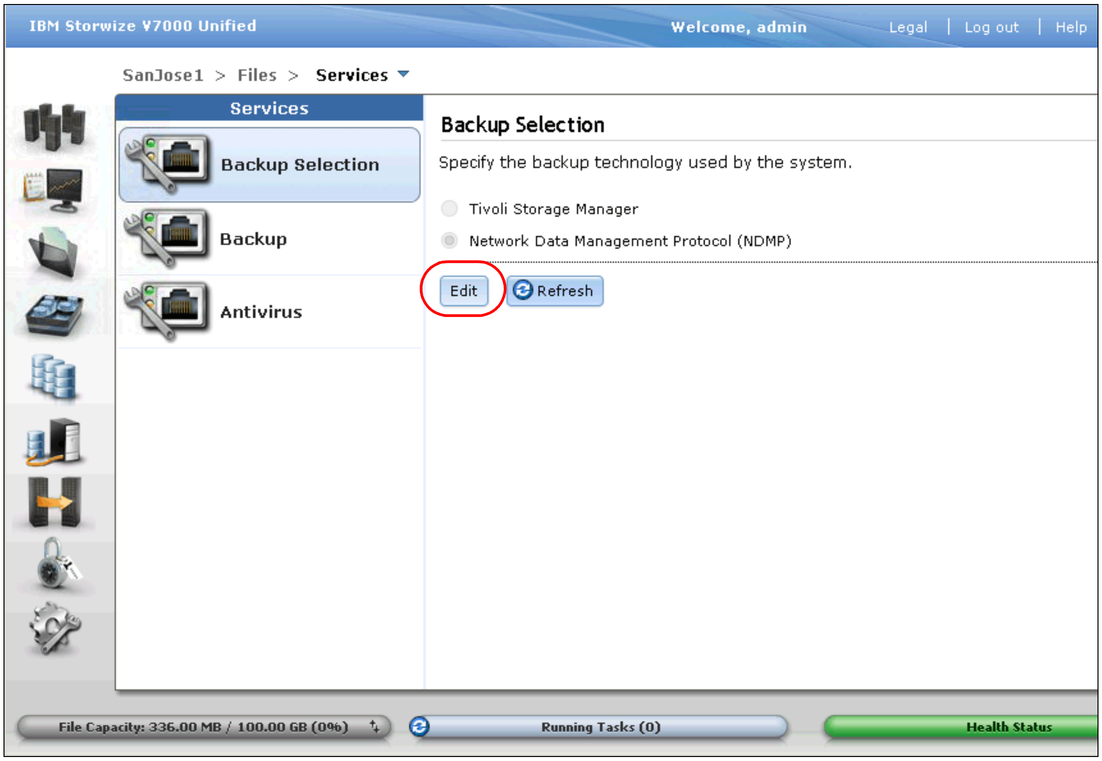
Network Data Management Protocol

A Data Management Application (DMA) is installed on an external server. When enabled and configured, the NDMP agent runs on the file modules and communicates with the DMA. Backup scheduling and management are controlled in the DMA. Data is prepared by the cluster and sent to the DMA server, which then stores the backup data on an external disk or on tape.

14.3.2 Tivoli Storage Manager

Tivoli Storage Manager is able to use the IBM Active Cloud Engine, which is incorporated in the Storwize V7000 Unified system.

To configure Tivoli Storage Manager, you must enable it as the backup technology and create a Tivoli Storage Manager server definition on the Storwize V7000 Unified system. As the first step, select **Files** → **Services** and click **Backup Selection**. This displays the window that is shown in Figure 14-1.



Backup Selection

Specify the backup technology used by the system.

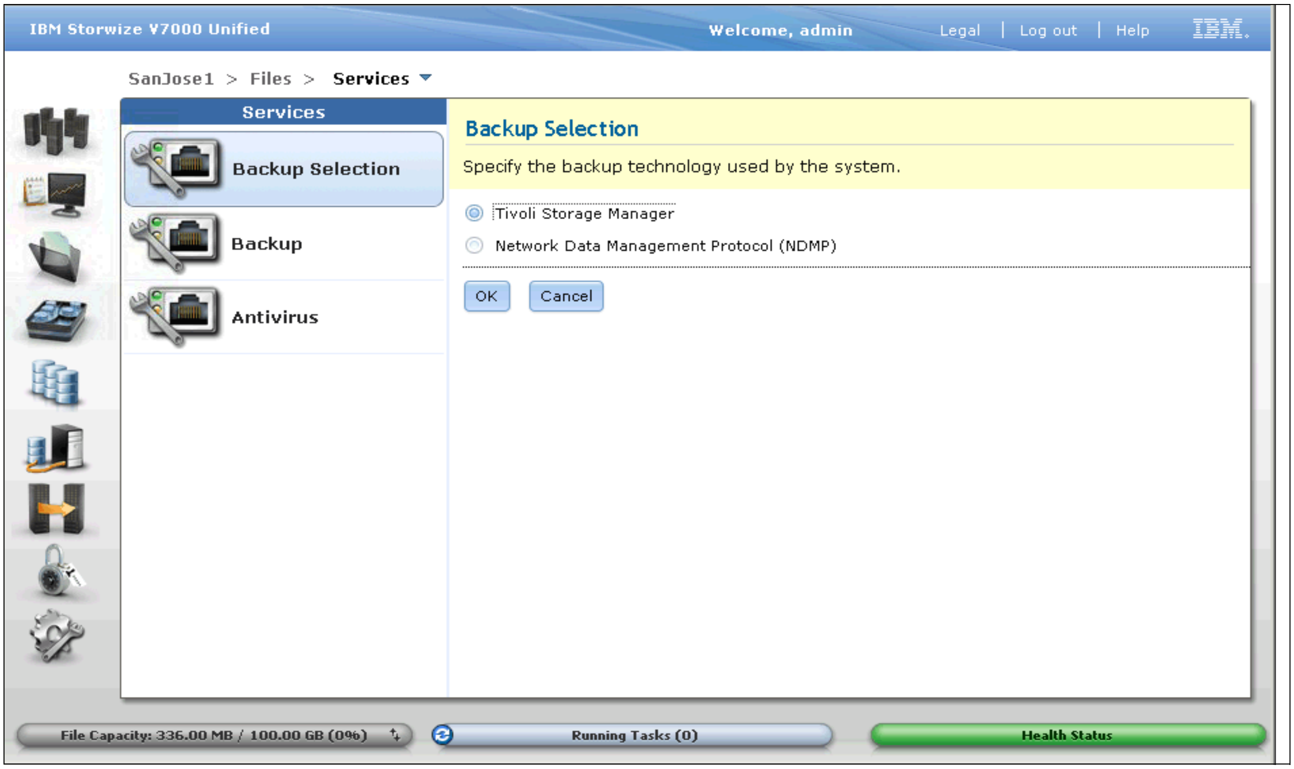
○ Tivoli Storage Manager

○ Network Data Management Protocol (NDMP)

Figure 14-1 Backup Selection tab

Confirm that the Tivoli Storage Manager option is selected. If not (as shown in Figure 14-1), use the **Edit** function to change it.

Figure 14-2 on page 244 shows the Backup Selection tab after changing the backup technology to Tivoli Storage Manager.



Backup Selection

Specify the backup technology used by the system.

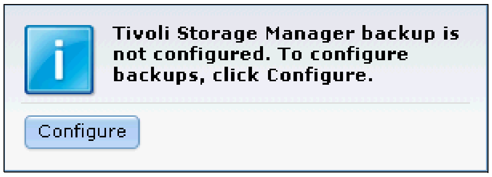
● Tivoli Storage Manager

○ Network Data Management Protocol (NDMP)

OK Cancel

Figure 14-2   Select Tivoli Storage Manager protocol

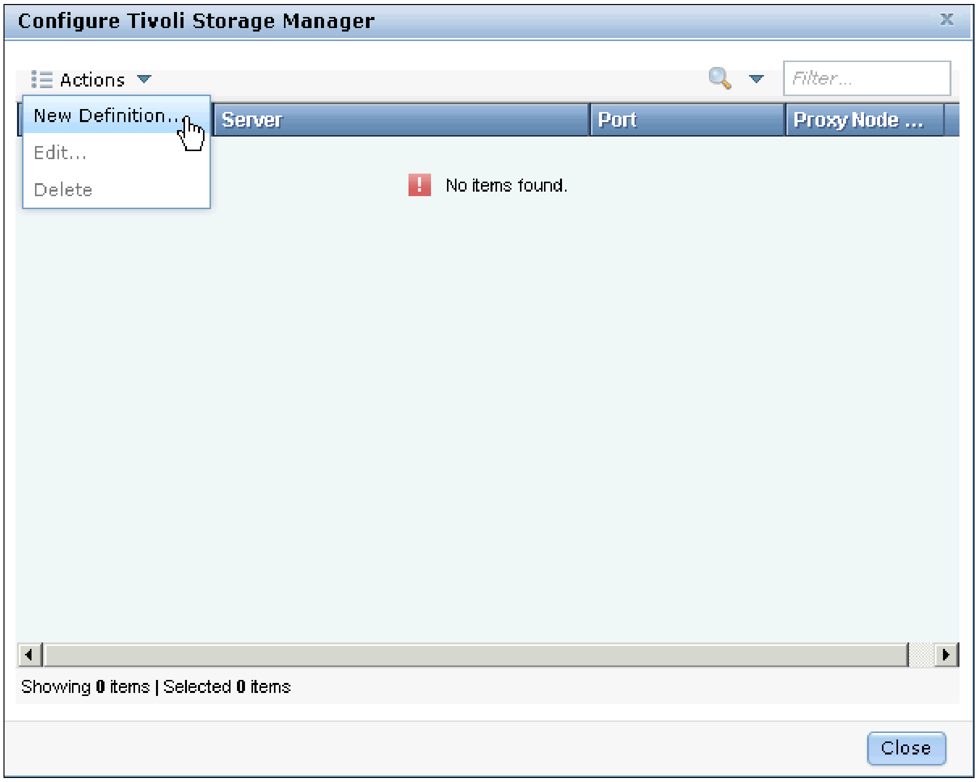
When you click the **Backup** icon, a message indicates that Tivoli Storage Manager is not configured (see Figure 14-3). Click **Configure**.



Tivoli Storage Manager backup is not configured. To configure backups, click Configure.

Figure 14-3   Tivoli Storage Manager backup is not configured

This starts the Tivoli Storage Manager configuration applet, which shows that there are no definitions. From the **Actions** pull-down menu, select **New Definition**, as shown in Figure 14-4 on page 245.



No items found.

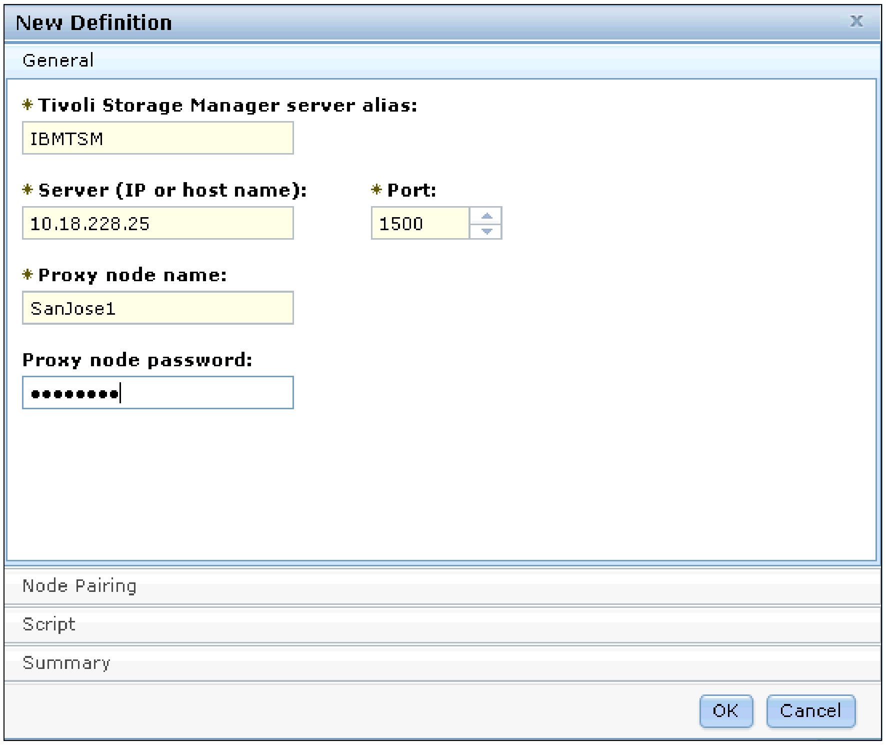
Showing **0** items | Selected **0** items

Figure 14-4 Add new definition

The **New Definition** window is displayed. This window has four tabs:

- General
- Node pairing
- Script
- Summary

The **General** tab is displayed first. Complete the Tivoli Storage Manager server and proxy details, as shown in Figure 14-5 on page 246.



* Tivoli Storage Manager server alias: IBMTSM

* Server (IP or host name): 10.18.228.25

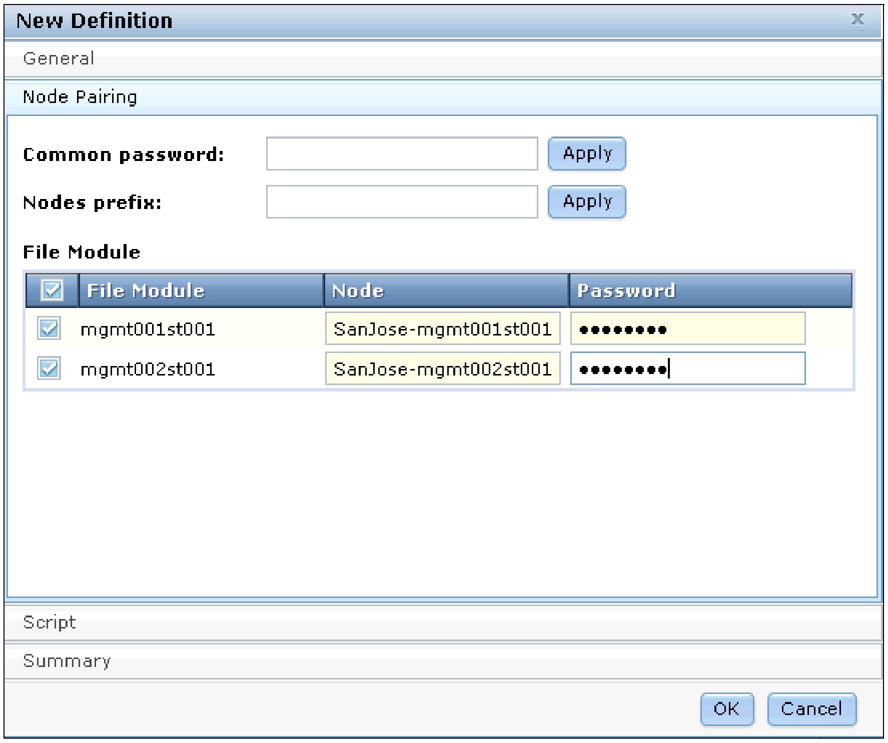
* Port: 1500

* Proxy node name: SanJose1

Proxy node password: ••••••••

Figure 14-5 New Definition window, General tab

Now click the **Node Pairing** tab. As Figure 14-6 shows, the two file modules are listed. Add a prefix by typing the prefix in the “Nodes prefix” field, and then click **Apply**. Also, enter the password in the “Common password” field and click **Apply**.



Common password:

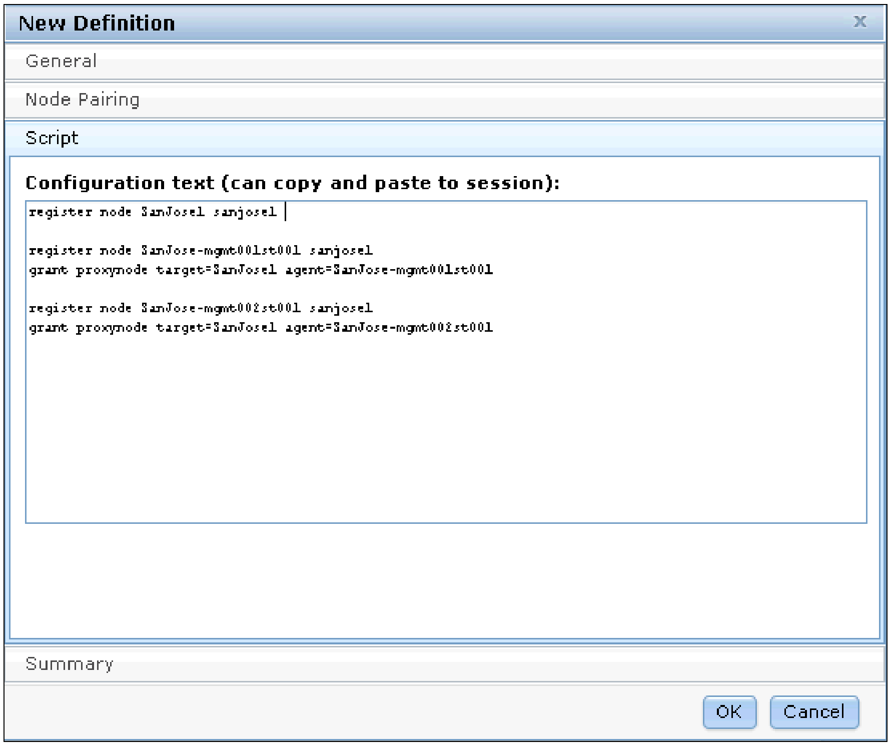
Nodes prefix:

File Module

☑	File Module	Node	Password
☑	mgmt001st001	SanJose-mgmt001st001	••••••••
☑	mgmt002st001	SanJose-mgmt002st001	••••••••

Figure 14-6 New Definition window, Node Pairing tab

Before clicking OK, you must copy this definition to the Tivoli Storage Manager server. Click the **Script** tab shown in Figure 14-7 to see the commands that are required to configure Tivoli Storage Manager. These are provided for your convenience. They can be copied and pasted to the Tivoli Storage Manager console.



Configuration text (can copy and paste to session):

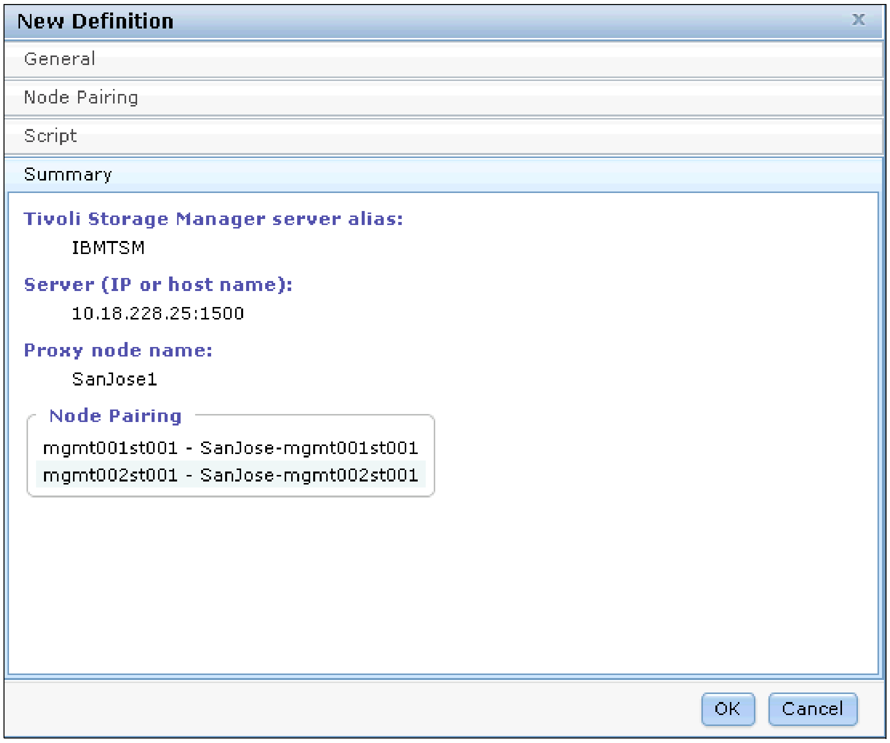
```
register node SanJose1 sanjose1

register node SanJose-mgmt001st001 sanjose1
grant proxynode target=SanJose1 agent=SanJose-mgmt001st001

register node SanJose-mgmt002st001 sanjose1
grant proxynode target=SanJose1 agent=SanJose-mgmt002st001
```

Figure 14-7 New Definition window, Script tab

Clicking the **Summary** tab displays the new configuration for your review (see Figure 14-8 on page 248).



Tivoli Storage Manager server alias:
IBMTSM

Server (IP or host name):
10.18.228.25:1500

Proxy node name:
SanJose1

Node Pairing

mgmt001st001 - SanJose-mgmt001st001
mgmt002st001 - SanJose-mgmt002st001

Figure 14-8 New Definition window, Summary tab

When the commands shown in the Script tab have completed successfully, click **OK** to build the definition.

You are now able to use Tivoli Storage Manager to perform backups of the data on this cluster. Management of the storage manager and the processes that are required to back up the data on the cluster are handled by Tivoli software. Explaining that process is and not within the scope of this book. Consult Tivoli Storage Manager technical support for assistance, if required.

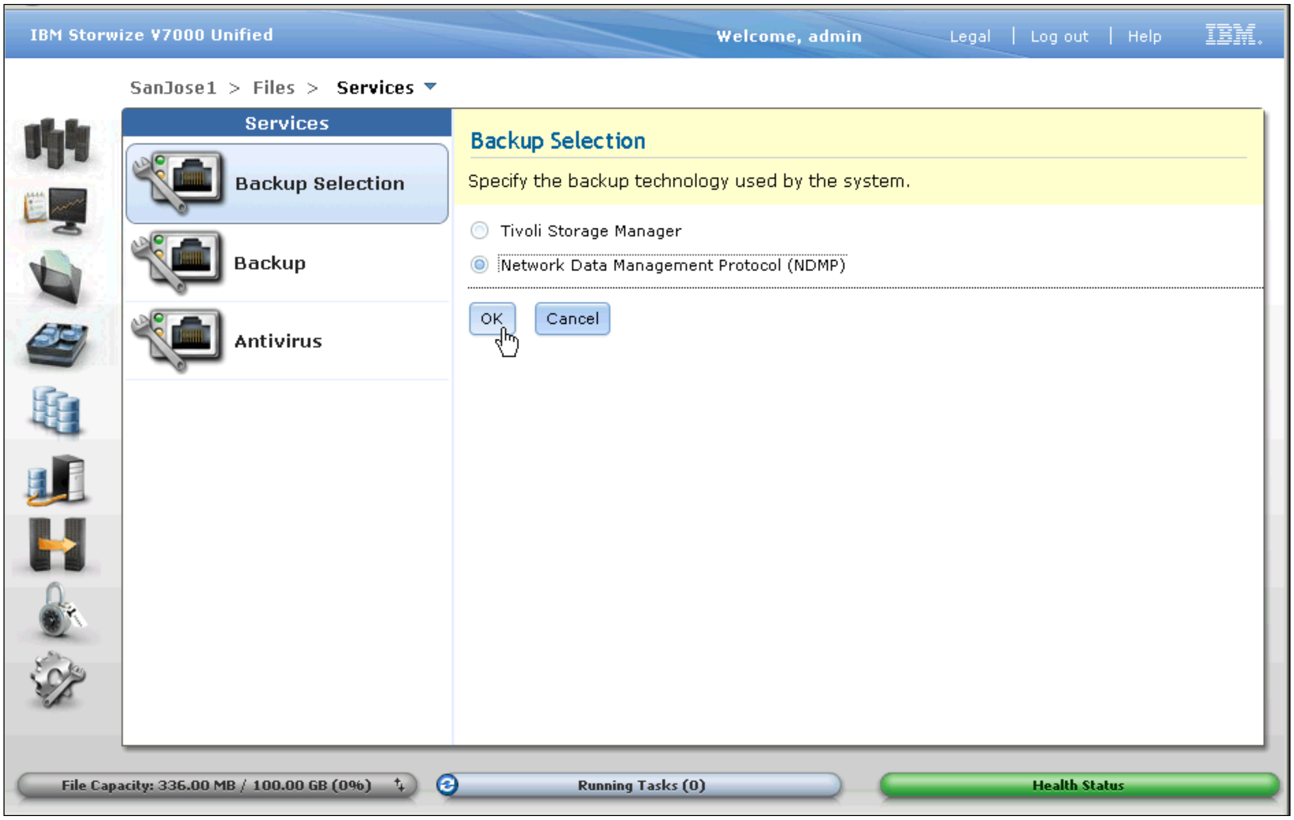
14.3.3 Network Data Management Protocol

Configuring the Network Data Management Protocol (NDMP) agent on the Storwize V7000 Unified system requires three basic steps:

1. Create the network group definition.
2. Set the configuration values.
3. Activate the node group.

The process begins by selecting the backup technology to use. Click **Files** → **Services** and then **Backup Selection**, as shown in Figure 14-9 on page 249.

Confirm that the **Network Data Management Protocol (NDMP)** option is selected. If not, use the edit function to change it.



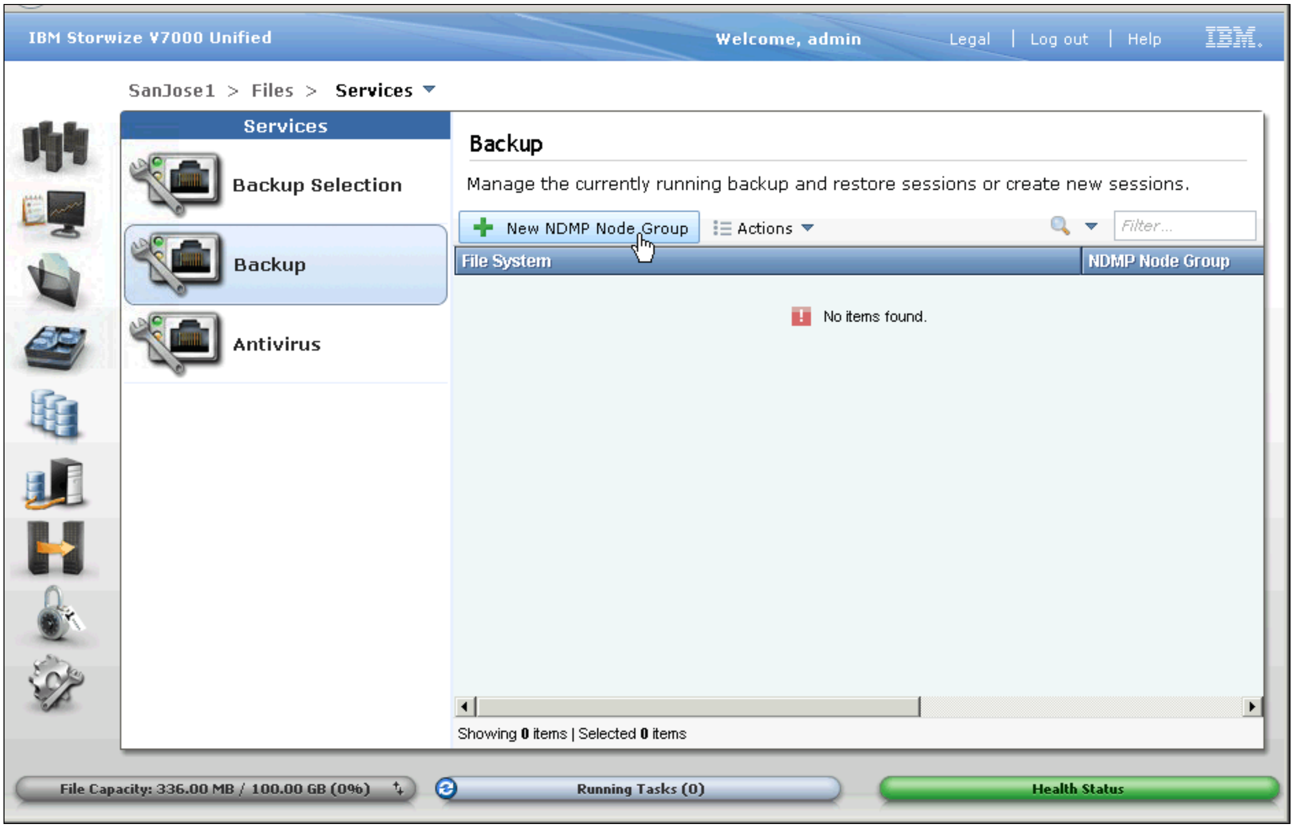
Backup Selection

Specify the backup technology used by the system.

○ Tivoli Storage Manager
● Network Data Management Protocol (NDMP)

Figure 14-9 Backup Selection tab: NDMP selected

Next, click the **Backup** icon to display the backup management window. Click **New NDMP Node Group** to create the network group. We show this in Figure 14-10 on page 250.



Backup

Manage the currently running backup and restore sessions or create new sessions.

No items found.

Showing 0 items | Selected 0 items

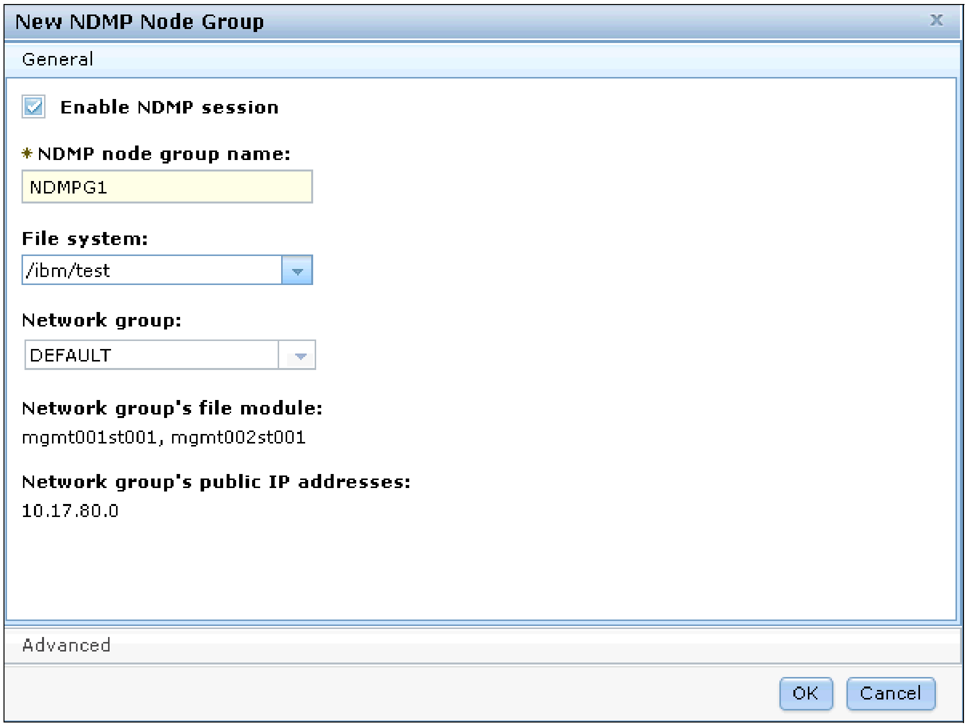
Figure 14-10 New NDMP Node Group selection

This action opens the window that is shown in Figure 14-11 on page 251.

Ensure that **the Enable NDMP session** field is selected, and enter a name for the group. Next, select the file systems to include in this group and the default network group. Use the following CLI command to create the group:

```
cfgndmp ndmpg1 --create
```

In this example, *ndmpg1* is the name of the group.



New NDMP Node Group

General

☑ Enable NDMP session

* NDMP node group name: NDMPG1

File system: /ibm/test

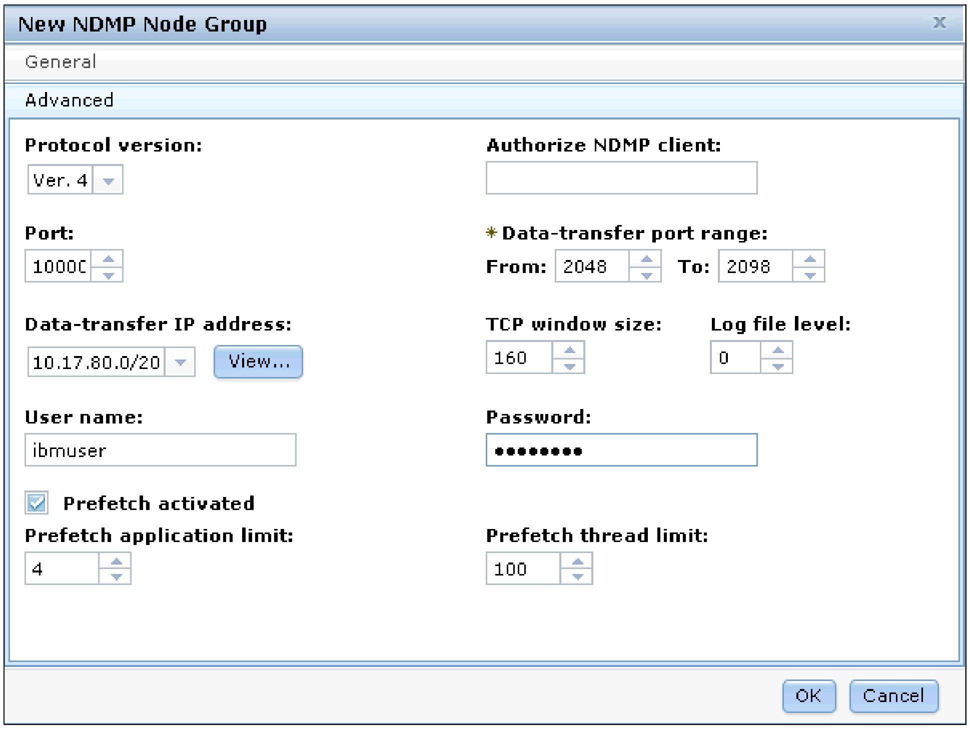
Network group: DEFAULT

Network group's file module: mgmt001st001, mgmt002st001

Network group's public IP addresses: 10.17.80.0

Figure 14-11 New NDMP Node Group window, General tab

Click the **Advanced** tab to show the remaining configuration parameters. Set these parameters as detailed in Table 14-1 on page 252 and as shown in Figure 14-12.



New NDMP Node Group

Advanced

Protocol version: Ver. 4

Authorize NDMP client:

Port: 10000

* Data-transfer port range: From: 2048 To: 2098

Data-transfer IP address: 10.17.80.0/20

TCP window size: 160

Log file level: 0

User name: ibmuser

Password: ●●●●●●●●

☑ Prefetch activated

Prefetch application limit: 4

Prefetch thread limit: 100

Figure 14-12 New NDMP Node Group: Advanced tab

Click **Close** when you are ready to continue.

Table 14-1 shows the NDMP group configuration.

Table 14-1 NDMP group configuration

Definition	Default	CLI command	Comment
NETWORK_GROUP_ATTACHED	“”	`cfgndmp ndmpg1 --networkGroup <xxx>`	The group of nodes to be attached with the NDMP group (optional).
DEFAULT_PROTOCOL_VERSION	4	`cfgndmp ndmpg1 --protocol 4`	Do not change this setting.

AUTHORIZED_NDMP_CLIENTS	“”	`cfgndmp ndmpg1 --limitDMA 192.168.0.3`
		`cfgndmp ndmpg1 --freeDMA`
			Default = null, no restrictions.

NDMP_PORT	10000	`cfgndmp ndmpg1 --dmaPort 10000`	
DATA_TX_PORT_RANGE		`cfgndmp ndmpg1 --dataTransferPortRange 10020-10025`	

DATA_TX_IP_ADDRS	“”	`cfgndmp ndmpg1 --limitDataIP 17.0.0.0/24`
		`cfgndmp ndmpg1 --freeDataIP`
			Default = null, no restrictions.

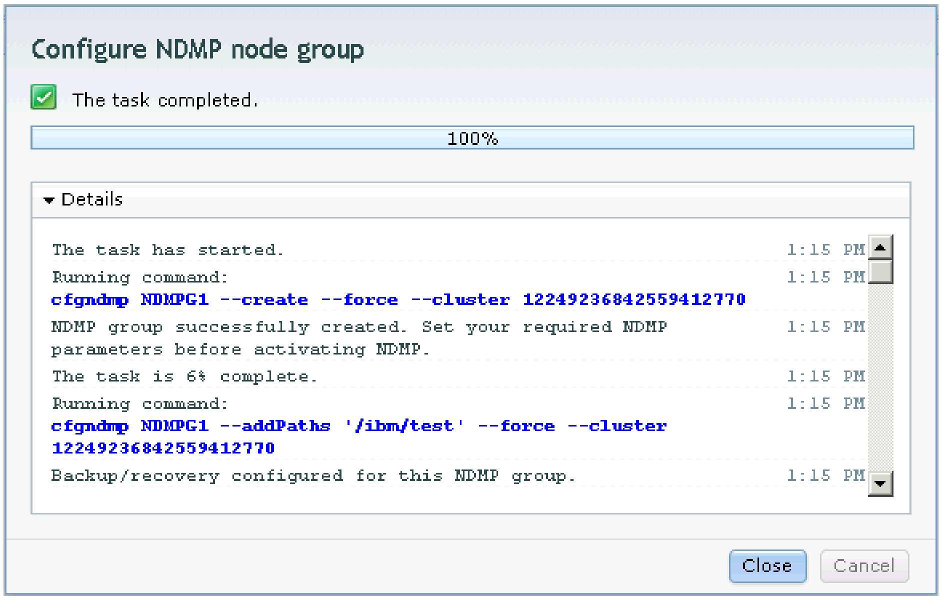
NDMP_TCP_WND_SZ	160	`cfgndmp ndmpg1 --tcpSize 160`	
LOG_FILE_TRACE_LEVEL	0	`cfgndmp ndmpg1 --logLevel 3`	
NDMP_USER_NAME	ndmp	`cfgndmp ndmpg1 --userCredentials ibmuser%mypass%mypass`	Specifies user name of “ibmuser” and password of “mypass” (repeated).
NDMP_PASSWORD	ndmp		

FILESYSTEM_PATHS		`cfgndmp ndmpg1 --addPaths /ibm/gpfs1`
		`cfgndmp ndmpg1 --removePaths /ibm/gpfs1`
			Define the paths to each file system that you want to include in this backup group.

ENABLE_NEW_SESSIONS	allow	`cfgndmp ndmpg1 --allowNewSessions`
		`cfgndmp ndmpg1 --denyNewSessions`	

Prefetch settings			
		`cfgndmpprefetch ndmpg1 —-activate`	Enable prefetch if wanted. NDMP must be deactivated to change prefetch status.
PF_APP_LIMIT	4	`cfgndmpprefetch ndmpg1 —-applimit 4`	Maximum sessions using prefetch at one time (1 - 10). NDMP must be deactivated to change prefetch status.
PF_NUM_THREADS	100	`cfgndmpprefetch ndmpg1 -—numThreads 180`	Threads per node to be used for prefetching (50 - 180). NDMP must be deactivated to change prefetch status.

When you have finished all entries, click **OK** to create the group. A pop-up window shows the progress. Upon completion, you can review the details, as shown in Figure 14-13 on page 253.



Configure NDMP node group

☑ The task completed.

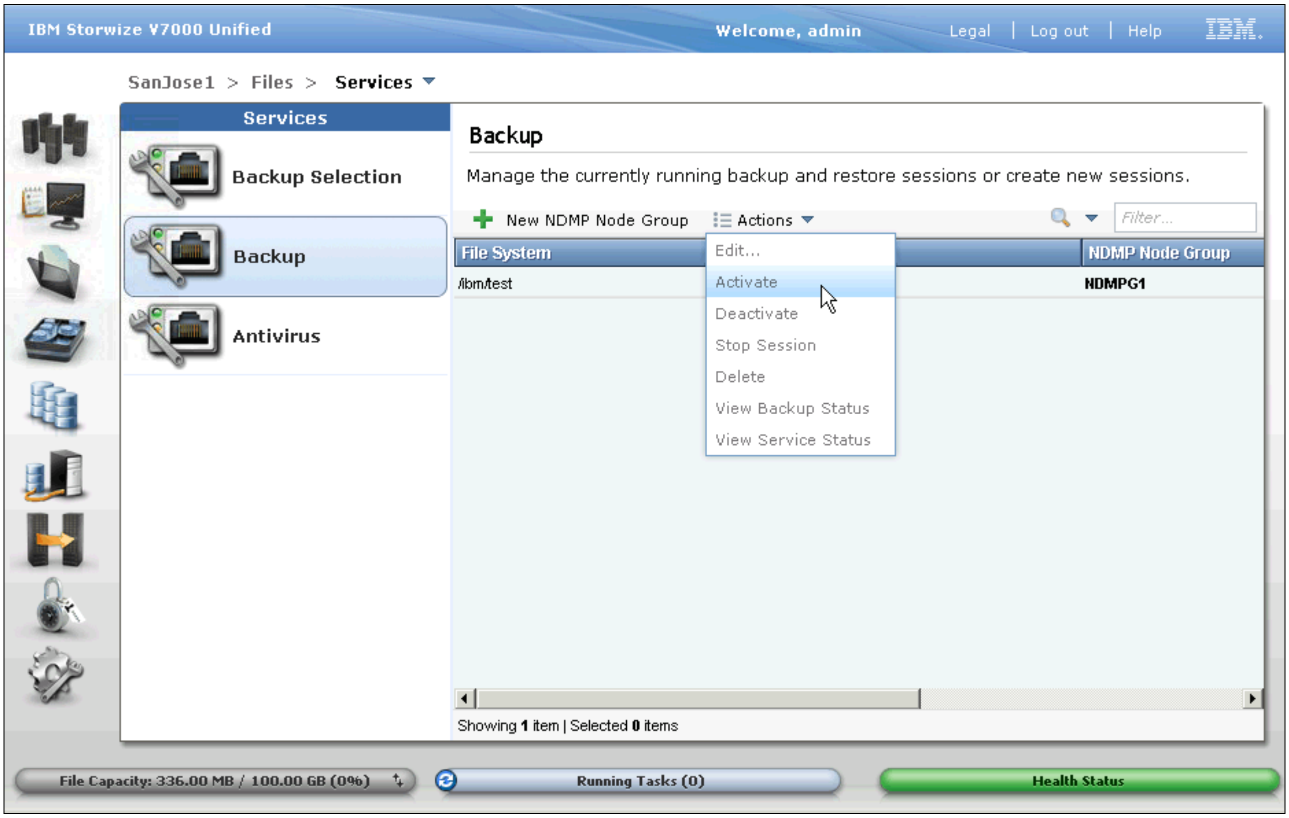
100%

▾ Details

```
The task has started.                                          1:15 PM
Running command:                                               1:15 PM
cfgndmp NDMPG1 --create --force --cluster 12249236842559412770
NDMP group successfully created. Set your required NDMP        1:15 PM
parameters before activating NDMP.
The task is 6% complete.                                       1:15 PM
Running command:                                               1:15 PM
cfgndmp NDMPG1 --addPaths '/ibm/test' --force --cluster
12249236842559412770
Backup/recovery configured for this NDMP group.                1:15 PM
```

Figure 14-13 Configure NDMP node group window, task completed

This returns you to the backup tab and now there is a line entry on the window for this new group. Highlight the line, then use the **Actions** pull-down menu or right-click to manage the group (as shown in Figure 14-14).



Backup

Manage the currently running backup and restore sessions or create new sessions.

File System	NDMP Node Group
/ibm/test	NDMPG1

Showing **1** item | Selected **0** items

File Capacity: 336.00 MB / 100.00 GB (0%)

Figure 14-14 Backup window, actions menu with Activate selected

You can edit the group's parameters. This takes you back through the same panels as the Create task. You can also activate and deactivate the NDMP group and stop a session. By using the **View Backup Status** option, you can show all currently running sessions. The **View Service Status** option shows you the status of the NDMP service on each node.

> **Caution:** Do not change the NDMP settings or prefetch configuration while a backup is being performed. This causes current backups to fail when NDMP recycles.

The following CLI command activates the group:

```
cfgndmp ndmpg1 --activateSnapshot
```

> **Caution:** Activate NDMP well before a backup is scheduled to run. Activating NDMP immediately before a backup might affect the backup while NDMP does start-up housekeeping. If snapshots are expired and require deletion, that might prevent creation of a new snapshot.

Prefetching

This feature is designed to greatly improve backup performance of NDMP with small files. The process predicts the file sequence for the NDMP group that is being backed up and reads the files into the node's cache. This has the effect of a high-read cache hit rate.

Prefetching can be turned on for up to 10 NDMP sessions. The threads setting is provided to allow tuning of the backup workload across the nodes that are based on the nodes cache memory and performance. Prefetch is designed to work with small files, under 1 MB. Files over 1 MB are not prefetched. Therefore, turning this feature on for groups with predominantly large files has little effect.

Prefetching is disabled by default. Activating it causes any currently running NDMP sessions to cease.

Making backups

When the NDMP sessions are configured and NDMP is active, you can make backups. This is done by using your NDMP-compliant DMA on an external server. Configuration and management of the DMA is beyond the scope of this book.

More information about managing NDMP is available in the "Managing Network Data Management Protocol" section of the IBM Storwize V7000 Unified 1.4.2 Information Center:

http://ibm.co/NTMv38

14.4 Data recovery

Although we do not show a data recovery scenario in this chapter, we describe the steps that are required to perform a recovery.

Various recovery procedures are available and can be used in case of the following issues:

- User ID and system access
- File module-related issues
- Control enclosure-related issues
- Restoring data
- Upgrade recovery

For details, see the IBM Storwize V7000 Unified 1.4.2 Information Center:

http://bit.ly/1qXdM19

14.4.1 Tivoli Storage Manager

Because the Storwize V7000 Unified software contains a Tivoli Storage Manager client, the restoration of a file system must be performed on the cluster that you want to recover from your Tivoli Storage Manager server.

The specific configuration and implementation of Tivoli Storage Manager differs from site to site. Therefore, we describe the process generically. Consult the relevant manuals and your Tivoli Storage Manager administrator before proceeding with a restore operation. The basic steps are as follows:

1. Ensure that there are no backups or restores currently running by issuing this command:

   `lsbackupfs`

2. You can restore a single file system or part of it using this command:

   `startrestore`

   A file pattern must be given. Using wildcards, all or part of a file system can be restored, even to a specific file. It is also possible to filter to a time stamp that restores files as they were, based on the backups available. *Overwrite* can be enabled or disabled. Review the command description in the information center and the examples that are given before using this command.

3. The **`lsbackupfs`** command can be used again to monitor progress.

See the “Restoring Tivoli Storage Manager data” section of the IBM Storwize V7000 Unified 1.4.2 Information Center for a full example:

http://ibm.co/1net3hi

14.4.2 Asynchronous data recovery

Recovering a file system with asynchronous replication requires that you configure and start a replication relationship from the target site to the source site.

For more information and an example, see “Restoring asynchronous data” in the IBM Storwize V7000 Unified 1.4.2 Information Center:

http://ibm.co/1crUbVx

14.4.3 Network Data Management Protocol

Like the backup process, restoring data with NDMP is performed on the DMA server and is beyond the scope of this book. This process requires that the NDMP agent that is running on the cluster is already defined. In a running system that requires backup of a file system or part of it, the agent is ready. In the case of a full rebuild, it might be necessary to define and configure the agent before you can proceed.

More information about managing NDMP is available in the IBM Storwize V7000 Unified 1.4.2 Information Center in the “Managing Network Data Management Protocol” section:

http://ibm.co/NTMv38

15

Troubleshooting and maintenance

In this chapter, we explain how errors and events are reported and logged and what tools are available to analyze and recover from a problem. We describe the procedures to follow and the methods and tools that are provided with the IBM Storwize V7000 Unified Disk System to enable IBM Support to assist you.

We also cover common maintenance procedures, including software upgrades, and address compression-related recovery, such as recovering from a volume that goes offline.

15.1 Maintenance philosophy

Many events or problems that occur in your Storwize V7000 Unified system environment require little or no user action. This is because the system employs a “self-healing” approach so that, where possible, automatic recovery is triggered for many events. Also, when the cause of a problem abates, recovery procedures automatically run and the event warning is closed. This happens, for example, when a storage or host path is lost and later recovered.

When a problem occurs, an entry that describes the problem by using errors codes is entered into the event log. If the event required action cannot be automatically resolved, it is marked as *unfixed*. Only unfixed events require action. Alerts can be configured to send email or a Simple Network Management Protocol (SNMP) alert, and this can be filtered according to the type of event. Recovery actions are taken by running the built-in guided maintenance procedures, which the user starts from the event display.

If the problem cannot be resolved by using guided procedures, the user is prompted to call IBM Service by using the local procedure. Depending on how you set up your *Call Home* feature, the Storwize V7000 Unified system might also send an alert to IBM, and an IBM Support representative might call you first. The primary purpose of this Call Home function is to get information about a possible problem to IBM promptly. It serves as a backup for problem notification. However, it remains your responsibility to be aware of the status and health of your system and to call IBM if service is required, unless you make special arrangements beyond the standard maintenance agreement.

All user and technical manuals are incorporated in a single interactive repository that is called the *information center*. This web-based system is described in 15.4, “Information center” on page 274.

Support of the Storwize V7000 Unified system is provided primarily by using the IBM remote support model, where the first contact is with the IBM Remote Technical Services (RTS) desk in each region. This proven process gives the fastest response and immediate engagement of specialized support personnel. If required, higher levels of support can be engaged quickly. An IBM Service Support Representative (SSR) is dispatched to your site only if there are actions necessary that require assistance of an IBM representative.

Most of the parts in the Storwize V7000 Unified system are designated *client replace*. If such a part needs replacement, the client-replaceable unit (CRU) is sent by courier to your site, and the RTS specialist gives instructions for the replacement process.

If there is a requirement for IBM Support to observe behavior or to log in to perform complex procedures, the specialist uses the IBM Assist On-site connection. When it is configured, this connects directly to one of your Storwize V7000 Unified system workstations. This tool is web-based. By using secure authorization, it provides a simple remote KVM function to the cluster or to a workstation in the your environment. It uses only normal HTML-based ports that are typically not blocked and also allows selected authorized specialists within IBM to jointly access the session if needed. The client maintains control of the workstation and can observe all activity. This tool greatly speeds up resolution time by getting a specialist onto the system quickly.

15.2 Event logs

Logging in the Storwize V7000 Unified system is in the event log. An event might be critical failures of a component that affects the data access or might be a record of a minor configuration change. The two major components of the Storwize V7000 Unified system both maintain an event log, but these logs are stored and handled independently. The format of the data in the log and the tools that are available to check and act on the log entries also differ significantly, so we describe them separately.

All log entries are tagged with a status that indicates the impact or severity of the event. This quickly highlights the importance of an event and allows for sorting and filtering for display.

Many trace and debug logs are also recorded at a low level within both components. Many of these logs wrap and must be collected close to a problem event. The logs are unformatted and not visible to the user. Many are collected by the *Download Support Package* data collection process. If more logs must be collected or generated, IBM Support provides instructions or connects to the device.

There are two places to look for these events. The first item to look at is the lower-right bar, which shows the overall health status of the system as green, yellow, or red. As shown in Figure 15-1, there are tabs for File and Block events. For file events, you also need to check the status for each file module by using **Monitoring** → **System Details** → ***select a File Module***, as shown in Figure 15-2 on page 260.

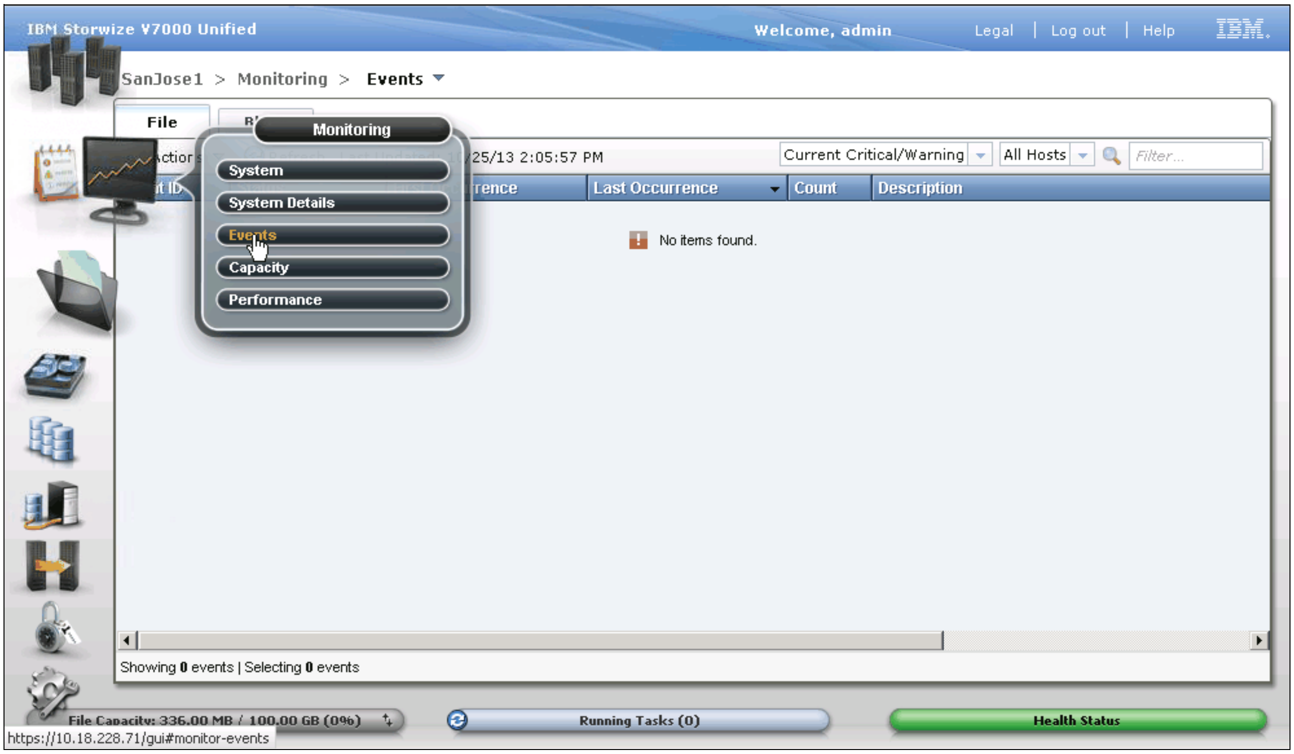
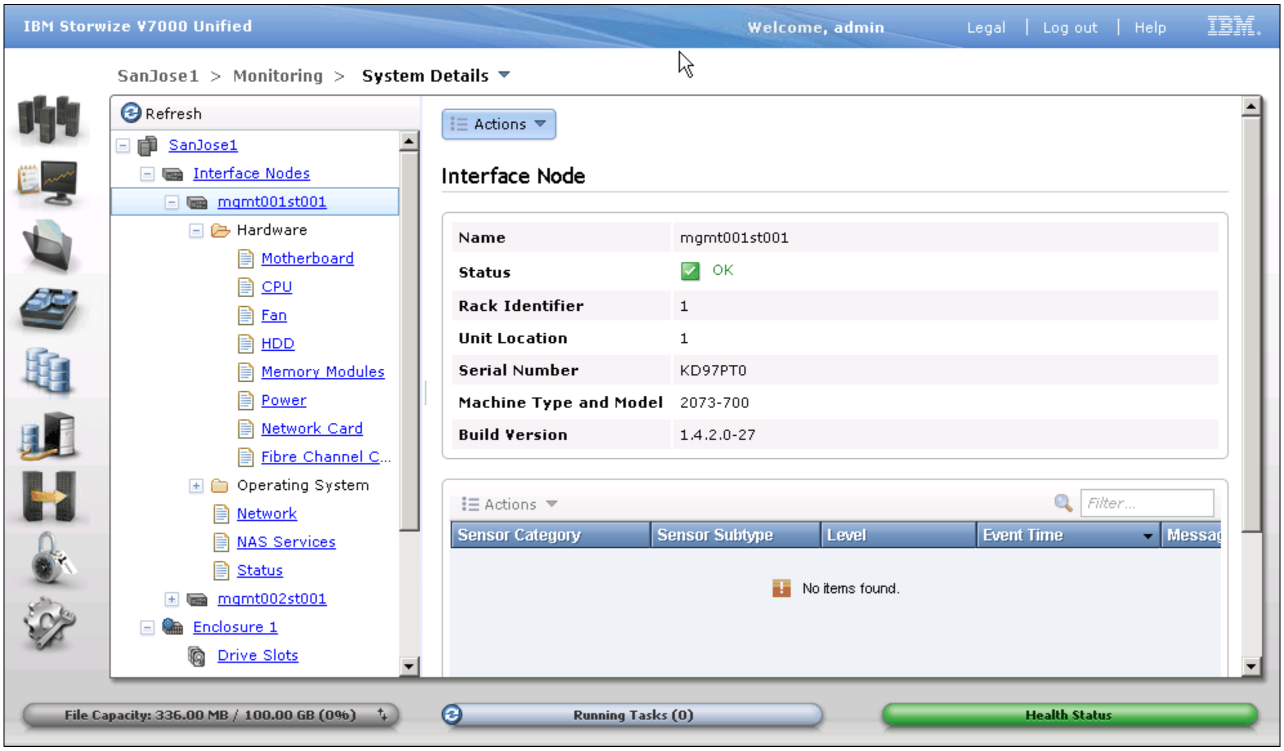


Figure 15-1 Block and file events

Figure 15-2 shows file module events.



Interface Node

Name	mgmt001st001
Status	✓ OK
Rack Identifier	1
Unit Location	1
Serial Number	KD97PT0
Machine Type and Model	2073-700
Build Version	1.4.2.0-27

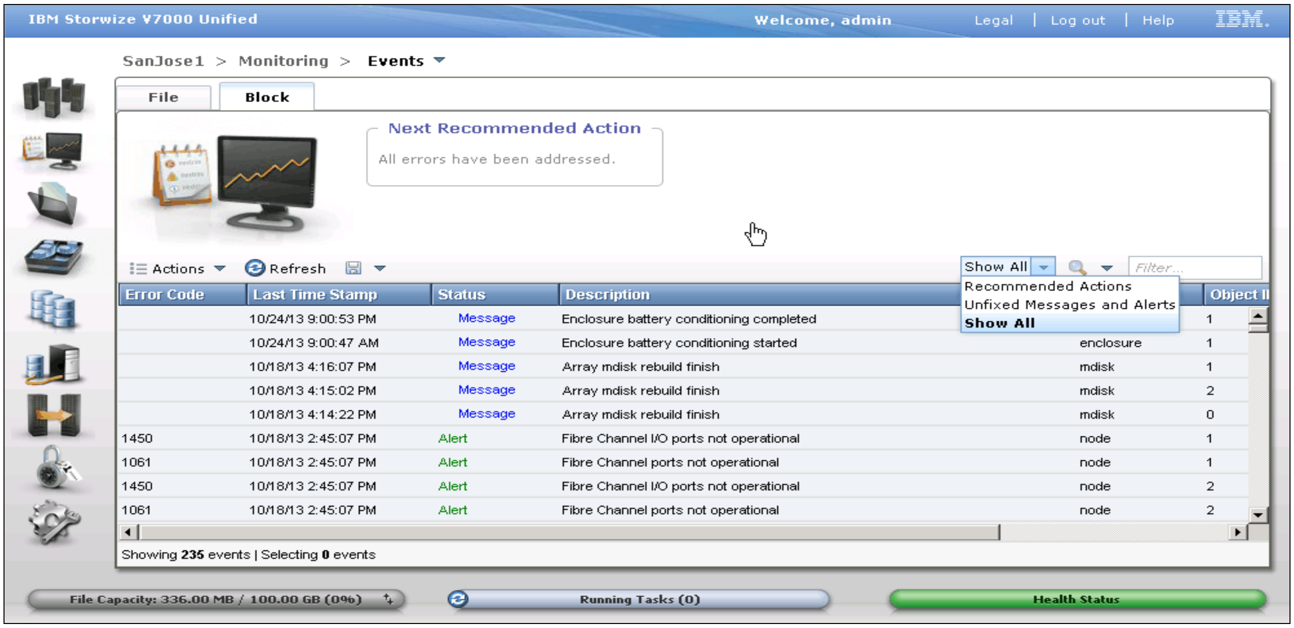
Sensor Category	Sensor Subtype	Level	Event Time	Messag

No items found.

Figure 15-2 File module events

15.2.1 Storwize V7000 storage controller event log (block)

To display the events, select **Monitoring** → **Events** and click the **Block** tab. This action opens the window that is shown in Figure 15-3.



Next Recommended Action

All errors have been addressed.

Error Code	Last Time Stamp	Status	Description		Object I
	10/24/13 9:00:53 PM	Message	Enclosure battery conditioning completed		1
	10/24/13 9:00:47 AM	Message	Enclosure battery conditioning started	enclosure	1
	10/18/13 4:16:07 PM	Message	Array mdisk rebuild finish	mdisk	1
	10/18/13 4:15:02 PM	Message	Array mdisk rebuild finish	mdisk	2
	10/18/13 4:14:22 PM	Message	Array mdisk rebuild finish	mdisk	0
1450	10/18/13 2:45:07 PM	Alert	Fibre Channel I/O ports not operational	node	1
1061	10/18/13 2:45:07 PM	Alert	Fibre Channel ports not operational	node	1
1450	10/18/13 2:45:07 PM	Alert	Fibre Channel I/O ports not operational	node	2
1061	10/18/13 2:45:07 PM	Alert	Fibre Channel ports not operational	node	2

Showing **235** events | Selecting **0** events

Figure 15-3 Event logs: Block

There are options to make this window more meaningful. You can sort on the event status. The pull-down window gives a choice of Recommended Actions, Unfixed Messages and Alerts, and Show All. The following list describes these options:

Recommended Actions Events that need actions performed, usually by running guided maintenance.

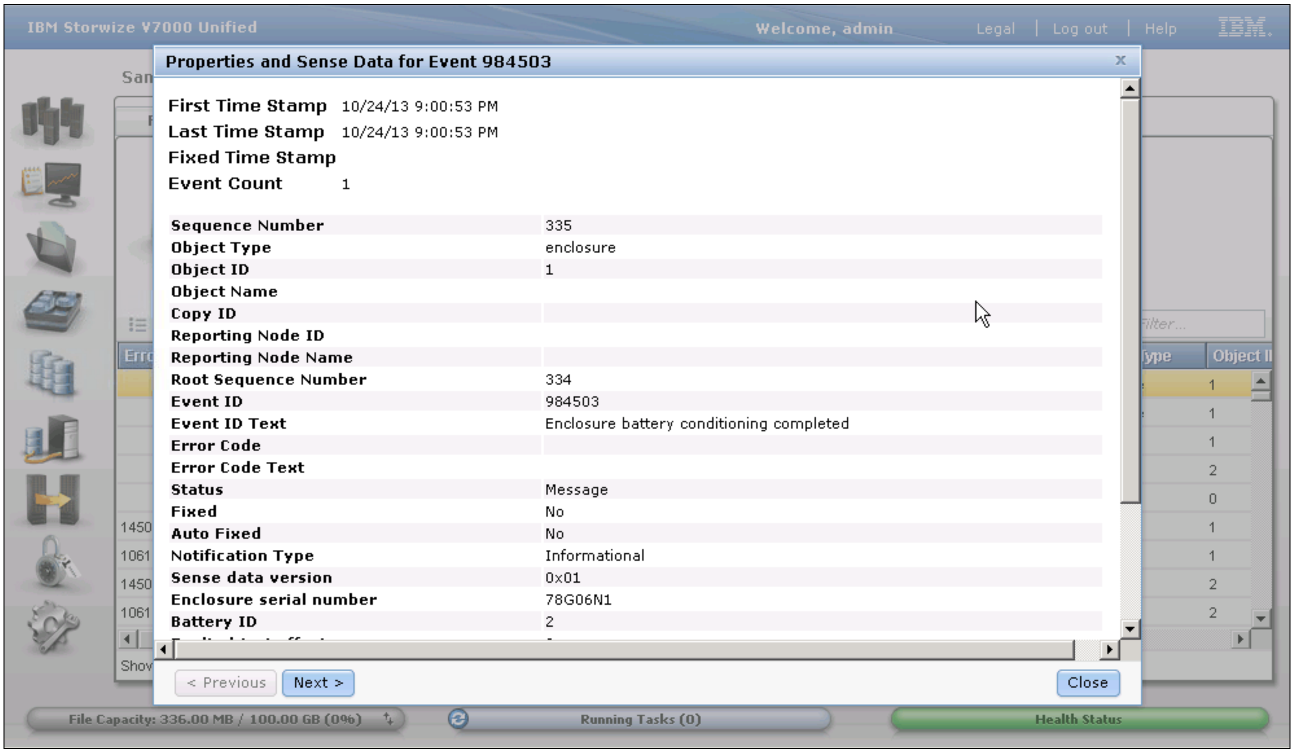
Unfixed Messages and Alerts This lists all events that are marked as *unfixed*.

Show All Lists all entries in the event log.

A filter option is also provided, although it is needed only if the log is cluttered with a high volume of events. Next to the magnifying glass is a pull-down arrow. Select the field to use as the filter. Then, in the filter box, enter the value of your filter.

Details

To display the event details, use the actions to select properties. An example of a minor event is shown in Figure 15-4.



Properties and Sense Data for Event 984503

First Time Stamp	10/24/13 9:00:53 PM
Last Time Stamp	10/24/13 9:00:53 PM
Fixed Time Stamp	
Event Count	1

Sequence Number	335
Object Type	enclosure
Object ID	1
Object Name	
Copy ID	
Reporting Node ID	
Reporting Node Name	
Root Sequence Number	334
Event ID	984503
Event ID Text	Enclosure battery conditioning completed
Error Code	
Error Code Text	
Status	Message
Fixed	No
Auto Fixed	No
Notification Type	Informational
Sense data version	0x01
Enclosure serial number	78G06N1
Battery ID	2

Figure 15-4 Block event: Properties

Two time stamps are logged: the first occurrence and the last. These are read with the event count. If the event count is one, the time stamps are the same, so the event occurred only once. If the same event occurs more than once, the count increases, and the last time stamp shows when the last one occurred.

The *Sequence number* uniquely identifies this event. This number increases by an increment for every new event for the life of the cluster.

The *Object type*, *ID,* and *name* identify the resource that this event refers to.

The *Reporting node* is the node that detected and logged this event.

If this event is related to another event, the *root sequence number* is the event that triggered this one. Therefore, this event is also considered in problem determination.

The *Event ID* is a number that uniquely relates to the event that is being logged and can be used to search the manuals or support knowledge bases to get more detail about the event. *Event ID text* that describes the event might be associated with the event ID.

An *Error code* might be displayed and might include associated text. It identifies the error that caused this event to be logged. All events that require guided maintenance have error codes.

Status gives the type of event and its severity.

The *Notification type* gives the priority of the event and the level of notification. This is used by the alert tasks in the cluster to determine what alerts to generate, based on the configured rules.

The *Fixed* field is important. Unfixed events can cause recovery routines to not be run if locks exist on processes as a result of an unfixed problem. Unfixed problems require action. There should be no unfixed events in the log unless they are for a current issue.

More information might also be supplied, including *Sense data version*, which varies based on the event.

Important: Regularly check the event log for unfixed problems. Although high-priority events cause alerts that are based on your configuration, it is possible to miss important events. Ensure that you act on all unfixed problems. If thin provisioning or compression is used, it is important that any out-of-space warnings be corrected immediately before that causes volumes to go offline because they lack sufficient resources.

Actions

Select a log entry by clicking it (which highlights it in yellow). Then, either right-click or use the Actions pull-down menu to display the following choices of actions:

Run Fix Procedures If the event is an error or qualifies for guided maintenance (and therefore has an error code logged), this option is available. Otherwise, it is disabled. This is the correct action for any unfixed problem that needs maintenance. This action starts directed maintenance procedures.

Mark as Fixed This option changes the fix status of an event to Fixed = yes. This is useful when multiple events are logged as a result of a problem and the problem is being fixed by using another event. It is also useful when you are confident that the cause has been corrected and there is no need to run maintenance. However, make that decision with caution, because you might inadvertently bypass cleanup routines. If a problem is informational and fix procedures are not available, use this option to remove the event from the unfixed list.

Mark as Unfixed If an event was incorrectly marked as fixed, this marks it as *unfixed*.

Filter by Date Gives you a choice of start and end dates to customize the display.

Show entries with... Expands to give a choice of *minutes*, *hours*, or *days previous* to limit the events that are displayed.

Reset Date Filter Resets the Show entries with... the ... choice.

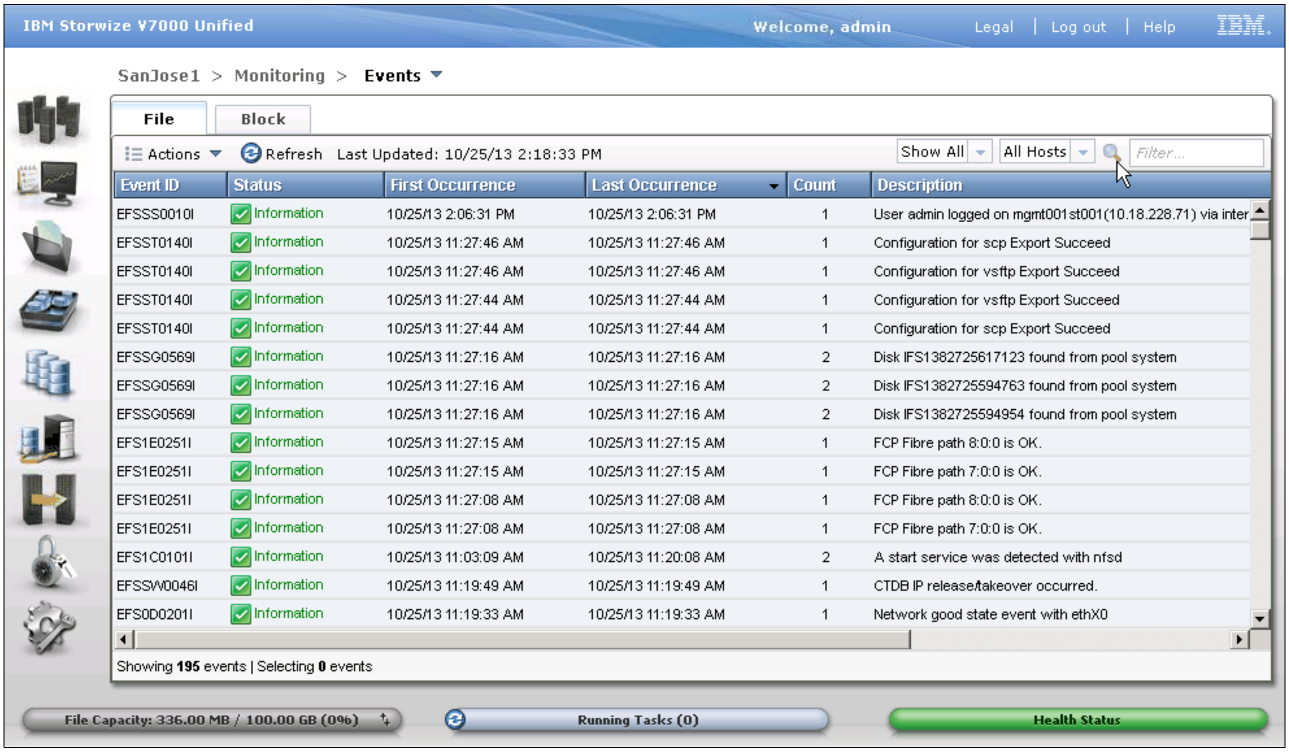
Clear Log Use Clear Log with great caution. This deletes the event contents from the log, which might compromise IBM Support's ability to diagnose problems. There is no need to clear the log. There is plenty of space to store the data, and the cluster performs its own housekeeping periodically to purge outdated or unimportant entries. The Refresh button refreshes the display.

Properties Displays the details of the event and the control fields as described in "Details" on page 261.

> **Suggestion:** Do not clear the event log. There is no practical gain in doing so, other than in special cases for site security reasons.

15.2.2 V7000 Unified system File module event log file

The *File* event log is cluster-wide and is common for both of the file modules. To display the events, select **Monitoring** → **Events** and click the **File** tab, as shown in Figure 15-5.



Last Updated: 10/25/13 2:18:33 PM

Event ID	Status	First Occurrence	Last Occurrence	Count	Description
EFSSS0010I	✓ Information	10/25/13 2:06:31 PM	10/25/13 2:06:31 PM	1	User admin logged on mgmt001st001(10.18.228.71) via inter
EFSST0140I	✓ Information	10/25/13 11:27:46 AM	10/25/13 11:27:46 AM	1	Configuration for scp Export Succeed
EFSST0140I	✓ Information	10/25/13 11:27:46 AM	10/25/13 11:27:46 AM	1	Configuration for vsftp Export Succeed
EFSST0140I	✓ Information	10/25/13 11:27:44 AM	10/25/13 11:27:44 AM	1	Configuration for vsftp Export Succeed
EFSST0140I	✓ Information	10/25/13 11:27:44 AM	10/25/13 11:27:44 AM	1	Configuration for scp Export Succeed
EFSSG0569I	✓ Information	10/25/13 11:27:16 AM	10/25/13 11:27:16 AM	2	Disk IFS1382725617123 found from pool system
EFSSG0569I	✓ Information	10/25/13 11:27:16 AM	10/25/13 11:27:16 AM	2	Disk IFS1382725594763 found from pool system
EFSSG0569I	✓ Information	10/25/13 11:27:16 AM	10/25/13 11:27:16 AM	2	Disk IFS1382725594954 found from pool system
EFS1E0251I	✓ Information	10/25/13 11:27:15 AM	10/25/13 11:27:15 AM	1	FCP Fibre path 8:0:0 is OK.
EFS1E0251I	✓ Information	10/25/13 11:27:15 AM	10/25/13 11:27:15 AM	1	FCP Fibre path 7:0:0 is OK.
EFS1E0251I	✓ Information	10/25/13 11:27:08 AM	10/25/13 11:27:08 AM	1	FCP Fibre path 8:0:0 is OK.
EFS1E0251I	✓ Information	10/25/13 11:27:08 AM	10/25/13 11:27:08 AM	1	FCP Fibre path 7:0:0 is OK.
EFS1C0101I	✓ Information	10/25/13 11:03:09 AM	10/25/13 11:20:08 AM	2	A start service was detected with nfsd
EFSSW0046I	✓ Information	10/25/13 11:19:49 AM	10/25/13 11:19:49 AM	1	CTDB IP release/takeover occurred.
EFS0D0201I	✓ Information	10/25/13 11:19:33 AM	10/25/13 11:19:33 AM	1	Network good state event with ethX0

Showing **195** events | Selecting **0** events

Figure 15-5 File event

Like the block display, there are options to make this window more meaningful. You can sort on the event status and currency:

Current Critical/Warning Unresolved events with a status of Critical or Warning.

Critical/Warning All events in the log with a status of Critical or Warning.

Show All Lists all entries in the event log.

You can filter the display to show the events that are generated by either or both file modules.

A filter option is also provided so that you can further reduce the display by including a search word or phrase. Simply type the word and press **Enter**. A Reset option clears the filter.

Actions

Unlike the block event log, there are no actions to fix or clear a specific log. This display is for viewing only. Highlight a log entry by clicking it. Then, either right-click or use the **Actions** pull-down menu to display the choices.

The following actions are available:

Filter by Date Gives a choice of *start* and *end* dates to customize the display.

Show entries within... Expands to give a choice of *minutes*, *hours*, or *days previous* to limit the displayed events.

Reset Date Filter Resets the preceding choice.

Clear Log Clears all entries from the event log. Again, use with caution.

Properties This gives more information about the event. Also in this window is a hyperlink to the Storwize V7000 Unified Information Center, which links directly to a description of the error code, as shown in Figure 15-8 on page 267.

In the topics that follow, we describe guided maintenance and the procedures.

15.2.3 Block

Any event that needs a maintenance action can run *with guided maintenance* enabled and is set to *unfixed*. To display unfixed problems, use the procedure that is explained in 15.2.1, “Storwize V7000 storage controller event log (block)” on page 260.

> **Important:** It is important to complete maintenance on all unfixed events in the cluster. Unfixed events can prevent maintenance routines from running and create contention for maintenance resources, which can prevent problems from auto recovering.

Check the event log regularly to ensure that someone acts on all unfixed events. Of particular importance are “out of space” conditions when compression is being used. Those must be addressed before physical storage is exhausted.

There are several types of actions for an unfixed event:

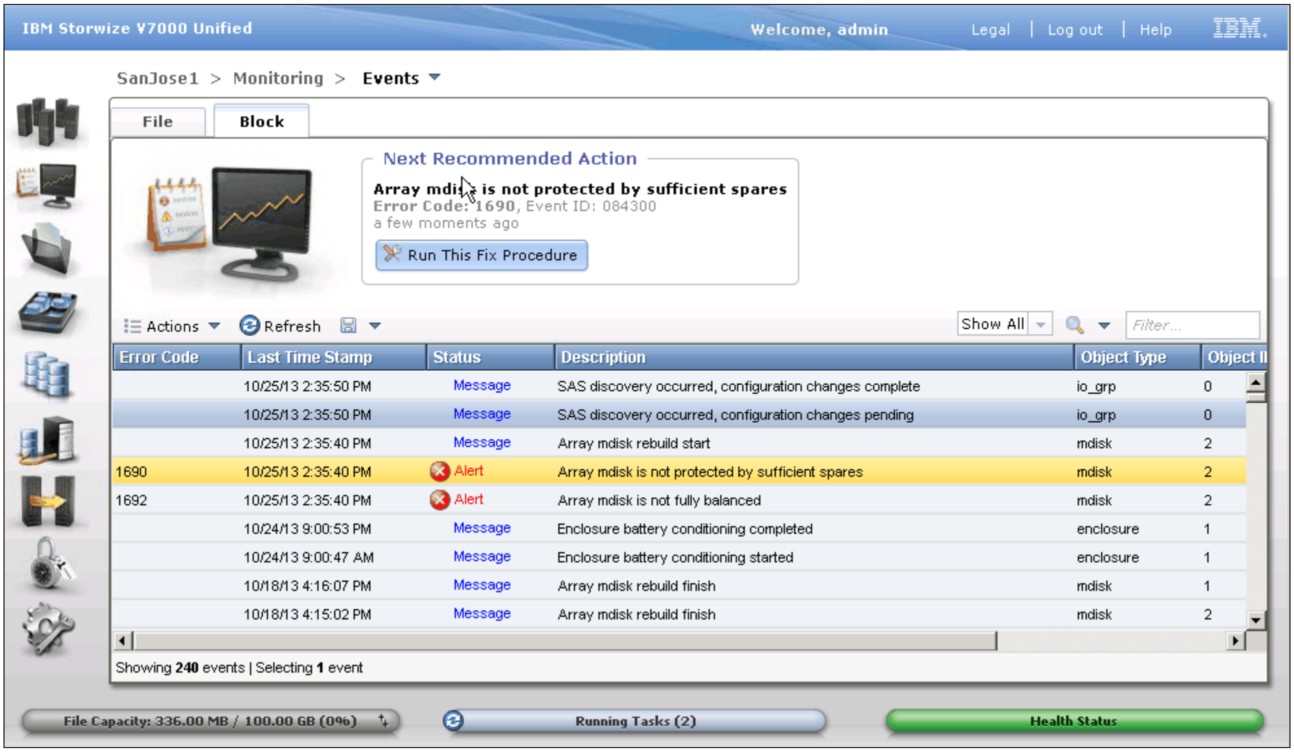
1. Use the *Run Fix Procedures* action to perform guided maintenance on the event. This is the preferred option.
2. Mark the event as *fixed*. This is done if the event is informational and fix procedures are not enabled. In this case, the event is intended to be read with another event or is purely for your information and is marked as unfixed to get your attention.

   Events can also be marked as *fixed* when there are several similar events and the problem was resolved by running maintenance on another entry. Use this option with caution, because it bypasses the guided maintenance routines, which include cleanup and discovery processes. This could leave a resource in a degraded or unusable state.
3. The Storwize V7000 system uses a self-healing approach where possible. Many events that are known to be transient trigger an autorun of the maintenance procedures and recovery. Subsequent events that clearly negate an earlier one automatically mark the earlier event as fixed. For example, a *node offline* message remains in the log, but gets marked as fixed if a *node online* message for the same node is logged 10 minutes later.

Run fix procedure

To display the block event log, select **Recommended Actions** from the filter pull-down menu. If there are any events that require running guided maintenance, they are listed in the window. The software also makes a suggestion in the Event list about which event to work on first. This is typically based on starting with the error code with the lowest number. The shortcut icon in this window starts the guided maintenance of the recommended event.

An example of this process is shown in Figure 15-6.



Next Recommended Action

Array mdisk is not protected by sufficient spares
Error Code: 1690, Event ID: 084300
a few moments ago

Run This Fix Procedure

Error Code	Last Time Stamp	Status	Description	Object Type	Object II
	10/25/13 2:35:50 PM	Message	SAS discovery occurred, configuration changes complete	io_grp	0
	10/25/13 2:35:50 PM	Message	SAS discovery occurred, configuration changes pending	io_grp	0
	10/25/13 2:35:40 PM	Message	Array mdisk rebuild start	mdisk	2
1690	10/25/13 2:35:40 PM	Alert	Array mdisk is not protected by sufficient spares	mdisk	2
1692	10/25/13 2:35:40 PM	Alert	Array mdisk is not fully balanced	mdisk	2
	10/24/13 9:00:53 PM	Message	Enclosure battery conditioning completed	enclosure	1
	10/24/13 9:00:47 AM	Message	Enclosure battery conditioning started	enclosure	1
	10/18/13 4:16:07 PM	Message	Array mdisk rebuild finish	mdisk	1
	10/18/13 4:15:02 PM	Message	Array mdisk rebuild finish	mdisk	2

Showing **240** events | Selecting **1** event

Figure 15-6 Block events: Error

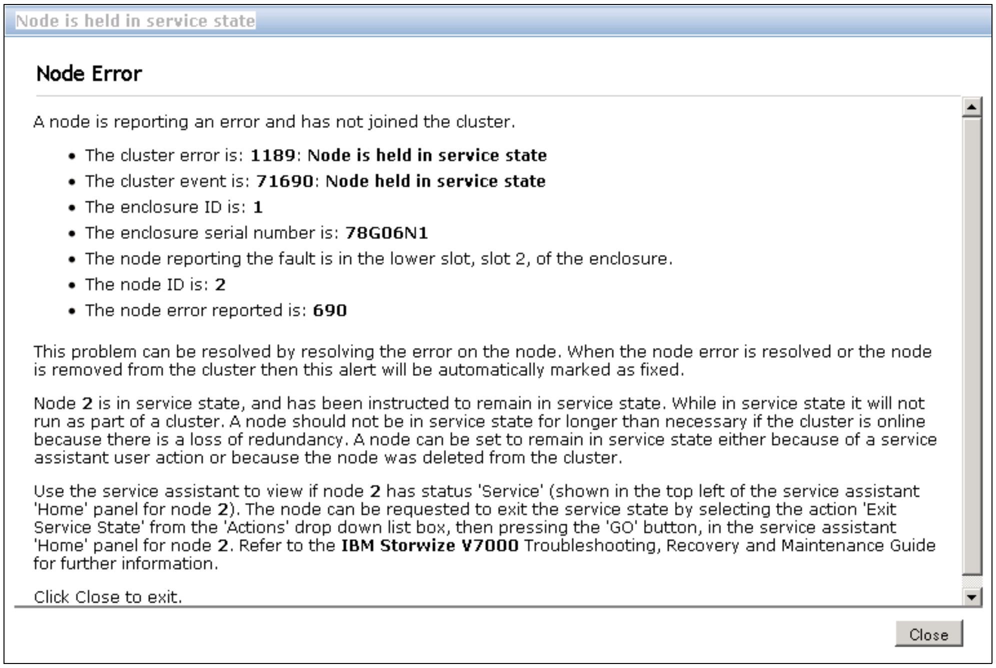
To manually select the event, click it in the event list. Then, use the **Actions** pull-down menu or right-click to see the action list and select **Run This Fix Procedure**.

This starts the guided maintenance. The panels that display are unique to the error and vary from a single window that provides information about how to resolve the error to a series of questions that require responses and actions. For a part failure, the guided maintenance steps through diagnosing and identifying the failed part, preparing the cluster for its isolation, powering off components if required, and guiding you through replacement and testing. The procedure confirms that the original error is now cleared and puts the appropriate resources online. Finally, it marks the event as fixed, which closes the problem report.

For most errors, the panels are interactive and ask for confirmation of status or configuration. *Be careful* answering these questions. Ensure that the requested state is correct before you confirm. For example, for a pathing error, you might be asked whether all zoning and pathing are currently operational and operating as intended. If you answer Yes, even though one path failed, the maintenance process assumes that previously missing paths are no longer in the configuration and does not look for them again.

In the example shown in Figure 15-7 on page 266, the guided maintenance does not take any action on the status because it does not identify the reason that it is set. Service mode is typically set manually by the user, so it is set for a reason. It can also be set by a failed node.

Therefore, there is not an accompanying event of higher importance. The panel that is displayed gives details about what the state is and where to go to clear it. For this particular error, the process leaves the event marked as unfixed, advising that it will automatically be marked as fixed when the state clears.



Node Error

A node is reporting an error and has not joined the cluster.

- The cluster error is: **1189**: **Node is held in service state**
- The cluster event is: **71690**: **Node held in service state**
- The enclosure ID is: **1**
- The enclosure serial number is: **78G06N1**
- The node reporting the fault is in the lower slot, slot 2, of the enclosure.
- The node ID is: **2**
- The node error reported is: **690**

This problem can be resolved by resolving the error on the node. When the node error is resolved or the node is removed from the cluster then this alert will be automatically marked as fixed.

Node **2** is in service state, and has been instructed to remain in service state. While in service state it will not run as part of a cluster. A node should not be in service state for longer than necessary if the cluster is online because there is a loss of redundancy. A node can be set to remain in service state either because of a service assistant user action or because the node was deleted from the cluster.

Use the service assistant to view if node **2** has status 'Service' (shown in the top left of the service assistant 'Home' panel for node **2**). The node can be requested to exit the service state by selecting the action 'Exit Service State' from the 'Actions' drop down list box, then pressing the 'GO' button, in the service assistant 'Home' panel for node **2**. Refer to the **IBM Storwize V7000** Troubleshooting, Recovery and Maintenance Guide for further information.

Click Close to exit.

Figure 15-7 Guided maintenance: 1189 error

In the example in Figure 15-7, there are three distinctly different codes:

Event code This is the reference code (5 digits) of the event, which uniquely identifies the type of event that occurred and provides details.

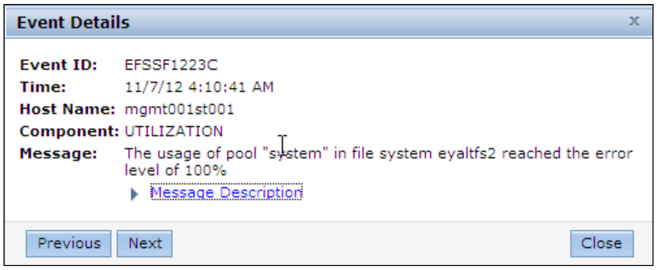
Error code An error code (4 digits) is posted only if an error occurs. This code is used by the guided maintenance and IBM Support personnel to repair the fault.

Node error code This code (3 digits) is related to the node module, not the cluster software.

15.2.4 File

The code that is running in the file modules does not incorporate a guided maintenance system. All maintenance and recovery procedures are guided from the information center. Use the error code as a starting point and search for the code in the information center. The information center gives appropriate actions for each error.

By viewing the event log and using the Actions pull-down menu (or right-clicking) to display the properties, you can see the event details, as shown in the example in Figure 15-8 on page 267. This gives more information about the event. Also in this window is a hyperlink to the Storwize V7000 Unified Information Center, which links directly to a description of the error code.

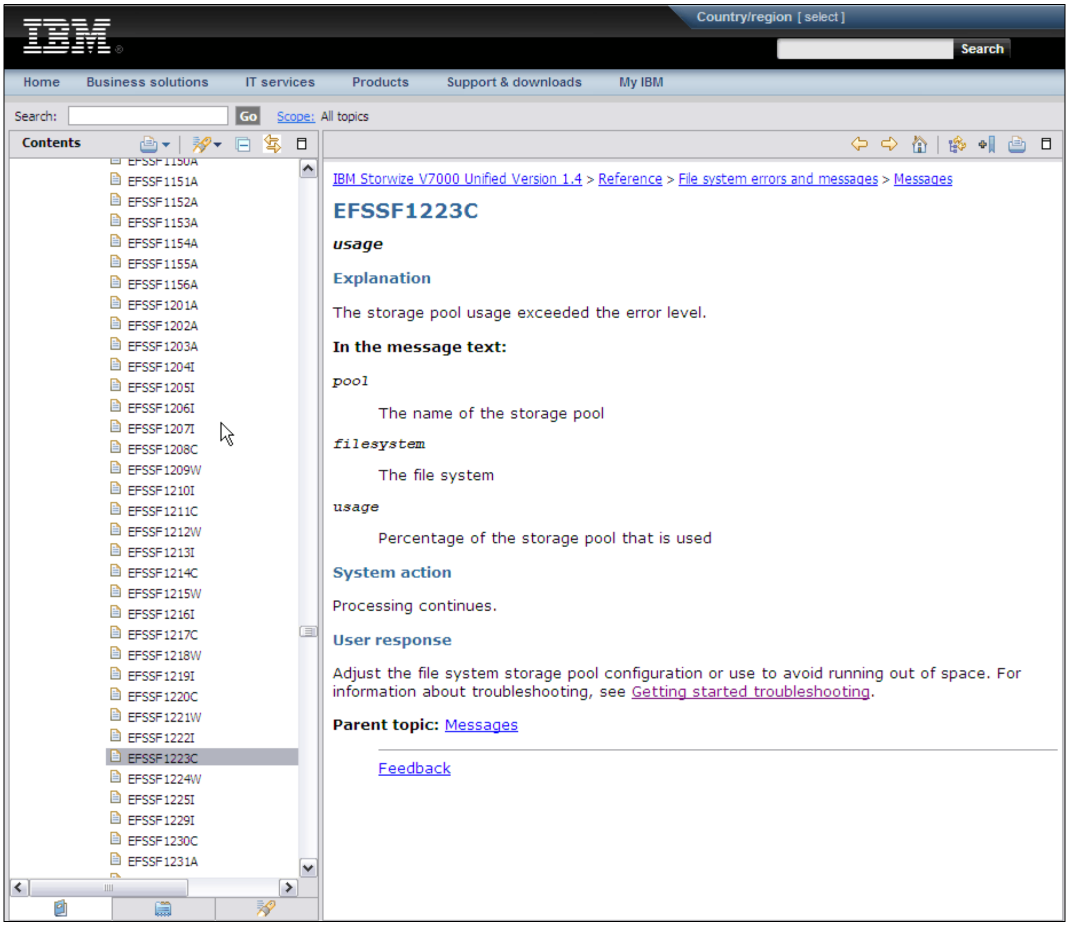


Event ID: EFSSF1223C
Time: 11/7/12 4:10:41 AM
Host Name: mgmt001st001
Component: UTILIZATION
Message: The usage of pool "system" in file system eyaltfs2 reached the error level of 100%

Message Description

Figure 15-8 File event: Details

In this example, the hyperlink launches the page that is shown in Figure 15-9.



IBM Storwize V7000 Unified Version 1.4 > Reference > File system errors and messages > Messages

EFSSF1223C

usage

Explanation

The storage pool usage exceeded the error level.

In the message text:

pool

The name of the storage pool

filesystem

The file system

usage

Percentage of the storage pool that is used

System action

Processing continues.

User response

Adjust the file system storage pool configuration or use to avoid running out of space. For information about troubleshooting, see Getting started troubleshooting.

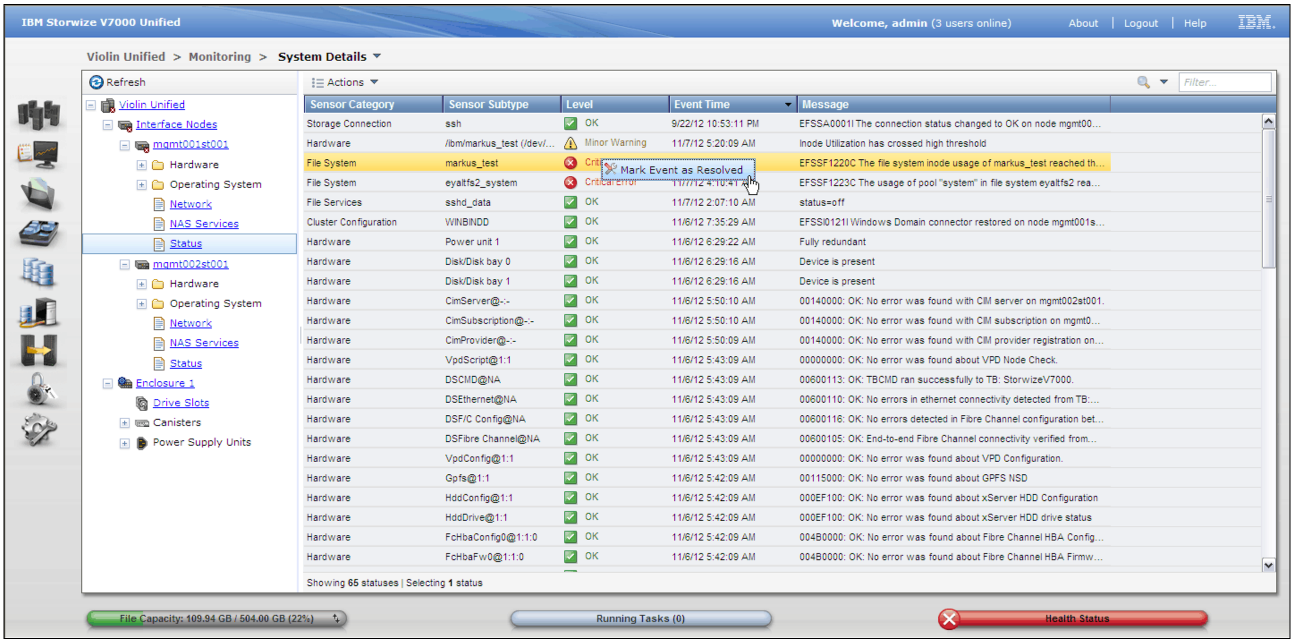
Parent topic: Messages

Feedback

Figure 15-9 Information center

Use the procedures that are outlined in the information center to resolve the problem. When the issue has been resolved, you need to **Mark Event as Resolved** by using the following procedure:

For File type events, check the status for each file module by using **Monitoring** → **System Details** → ***select a file module***. This is the window that is used to mark the error as being fixed, as shown in Figure 15-10.



Sensor Category	Sensor Subtype	Level	Event Time	Message
Storage Connection	ssh	✓ OK	9/22/12 10:53:11 PM	EFSSA0001I The connection status changed to OK on node mgmt00...
Hardware	/ibm/markus_test (/dev/...	⚠ Minor Warning	11/7/12 5:20:09 AM	Inode Utilization has crossed high threshold
File System	markus_test	☒ Criti		EFSSF1220C The file system inode usage of markus_test reached th...
File System	eyaltfs2_system	☒ Critical Error	11/7/12 4:10:41 A	EFSSF1223C The usage of pool "system" in file system eyaltfs2 rea...
File Services	sshd_data	✓ OK	11/7/12 2:07:10 AM	status=off
Cluster Configuration	WINBINDD	✓ OK	11/6/12 7:35:29 AM	EFSSI0121I Windows Domain connector restored on node mgmt001s...
Hardware	Power unit 1	✓ OK	11/6/12 6:29:22 AM	Fully redundant
Hardware	Disk/Disk bay 0	✓ OK	11/6/12 6:29:16 AM	Device is present
Hardware	Disk/Disk bay 1	✓ OK	11/6/12 6:29:16 AM	Device is present
Hardware	CimServer@-:-	✓ OK	11/6/12 5:50:10 AM	00140000: OK: No error was found with CIM server on mgmt002st001.
Hardware	CimSubscription@-:-	✓ OK	11/6/12 5:50:10 AM	00140000: OK: No error was found with CIM subscription on mgmt0...
Hardware	CimProvider@-:-	✓ OK	11/6/12 5:50:09 AM	00140000: OK: No error was found with CIM provider registration on...
Hardware	VpdScript@1:1	✓ OK	11/6/12 5:43:09 AM	00000000: OK: No error was found about VPD Node Check.
Hardware	DSCMD@NA	✓ OK	11/6/12 5:43:09 AM	00600113: OK: TBCMD ran successfully to TB: StorwizeV7000.
Hardware	DSEthernet@NA	✓ OK	11/6/12 5:43:09 AM	00600110: OK: No errors in ethernet connectivity detected from TB:...
Hardware	DSF/C Config@NA	✓ OK	11/6/12 5:43:09 AM	00600116: OK: No errors detected in Fibre Channel configuration bet...
Hardware	DSFibre Channel@NA	✓ OK	11/6/12 5:43:09 AM	00600105: OK: End-to-end Fibre Channel connectivity verified from...
Hardware	VpdConfig@1:1	✓ OK	11/6/12 5:43:09 AM	00000000: OK: No error was found about VPD Configuration.
Hardware	Gpfs@1:1	✓ OK	11/6/12 5:42:09 AM	00115000: OK: No error was found about GPFS NSD
Hardware	HddConfig@1:1	✓ OK	11/6/12 5:42:09 AM	000EF100: OK: No error was found about xServer HDD Configuration
Hardware	HddDrive@1:1	✓ OK	11/6/12 5:42:09 AM	000EF100: OK: No error was found about xServer HDD drive status
Hardware	FcHbaConfig0@1:1:0	✓ OK	11/6/12 5:42:09 AM	004B0000: OK: No error was found about Fibre Channel HBA Config...
Hardware	FcHbaFw0@1:1:0	✓ OK	11/6/12 5:42:09 AM	004B0000: OK: No error was found about Fibre Channel HBA Firmw...

Showing **65** statuses | Selecting **1** status

Figure 15-10 Mark event as resolved

A specific example for resolving a volume out-of-space condition and marking the event as resolved is covered in the next section, 15.2.5, “Working with compressed volumes out-of-space conditions” on page 268.

15.2.5 Working with compressed volumes out-of-space conditions

The most important metric to monitor is the use of the physical space in a storage pool. This metric relates to the actual physical storage capacity already used for storing compressed data that is written to the pool. It is important to ensure that physical allocation does not go over the suggested threshold (the default is set to 80%). To reduce the used space, the storage pool size needs to be increased. Adding physical capacity to the pool reduces its used space. Whenever a threshold is exceeded and an alert is generated, the system points you to the procedures outlined in the information center to resolve the problem. If a corrective action to reduce space used is not performed before the storage pool reaches 100% used, Network Shared Disks (NSDs) can go offline, and that can cause the file system to go offline.

> **Note:** The steps that follow were verified while writing this book. However, we suggest that you check the information center, because it might offer updated steps to resolve volume out-of-space conditions.

The following steps are required to correct an unmounted file system as a result of a compressed volume (NSD) going offline:

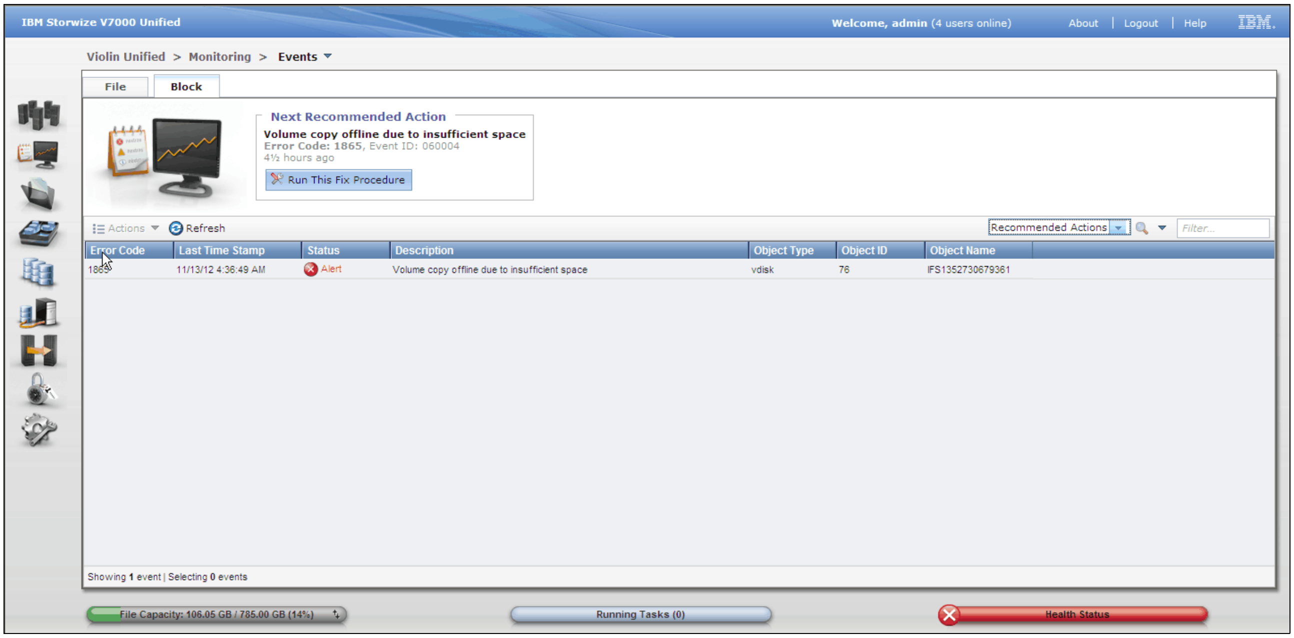
1. For a Block event, run the fix procedure for the NSD that went offline, and follow the steps that are provided by the fix procedure for the block event. The NSD offline condition must be addressed before the file system can be mounted.
2. Start the NSD when space is available.
3. Determine whether there are any stale NFS file handles.
4. Perform single node reboots to clear any stale NFS file handles.
5. Mount the file system.

Proceed with the following detailed steps.

Run fix procedure (block)

To display the block event log, select **Recommended Actions** from the filter pull-down menu. If there are any events that require guided maintenance to be run, they are listed in the window. The error for a volume copy offline because of insufficient space is *Error 1865*. The software also makes a recommendation in the Event list about which event to work on first. This is usually based on starting with the error code with the lowest number. The shortcut icon in this window starts the guided maintenance of the recommended event.

Figure 15-11 shows an example.


IBM Storwize V7000 Unified
Welcome, admin (4 users online)
About | Logout | Help
IBM.
Violin Unified > Monitoring > Events
File
Block
Run This Fix Procedure
Actions
Refresh
Recommended Actions
Filter...
File Capacity: 106.05 GB / 785.00 GB (14%)
Running Tasks (0)
Health Status

Next Recommended Action

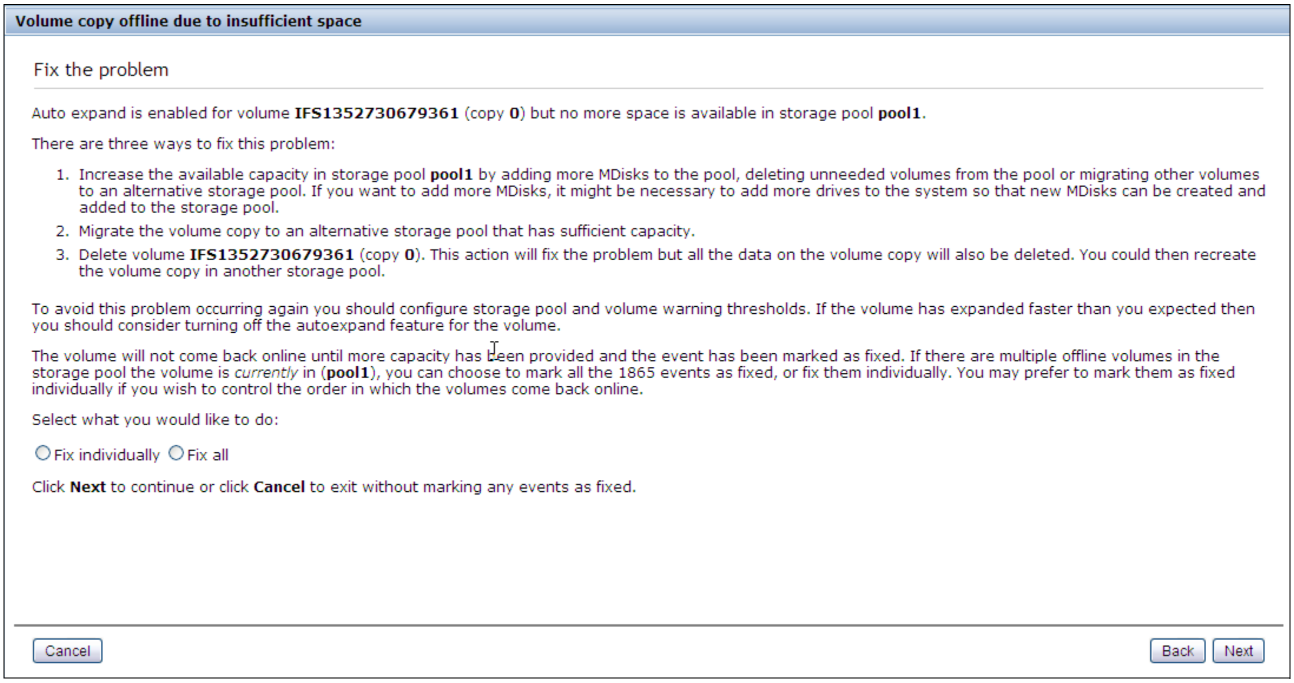
Volume copy offline due to insufficient space
Error Code: 1865, Event ID: 060004
4½ hours ago

Error Code	Last Time Stamp	Status	Description	Object Type	Object ID	Object Name
1865	11/13/12 4:36:49 AM	Alert	Volume copy offline due to insufficient space	vdisk	76	IFS1352730679361

Showing **1** event | Selecting **0** events

Figure 15-11 Run fix procedure for NSD offline

The guided maintenance procedure coaches you on how to resolve the volume-offline condition, as the example in Figure 15-12 on page 270 shows.



Volume copy offline due to insufficient space

Fix the problem

Auto expand is enabled for volume **IFS1352730679361** (copy **0**) but no more space is available in storage pool **pool1**.

There are three ways to fix this problem:

1. Increase the available capacity in storage pool **pool1** by adding more MDisks to the pool, deleting unneeded volumes from the pool or migrating other volumes to an alternative storage pool. If you want to add more MDisks, it might be necessary to add more drives to the system so that new MDisks can be created and added to the storage pool.
2. Migrate the volume copy to an alternative storage pool that has sufficient capacity.
3. Delete volume **IFS1352730679361** (copy **0**). This action will fix the problem but all the data on the volume copy will also be deleted. You could then recreate the volume copy in another storage pool.

To avoid this problem occurring again you should configure storage pool and volume warning thresholds. If the volume has expanded faster than you expected then you should consider turning off the autoexpand feature for the volume.

The volume will not come back online until more capacity has been provided and the event has been marked as fixed. If there are multiple offline volumes in the storage pool the volume is *currently* in (**pool1**), you can choose to mark all the 1865 events as fixed, or fix them individually. You may prefer to mark them as fixed individually if you wish to control the order in which the volumes come back online.

Select what you would like to do:

○ Fix individually ○ Fix all

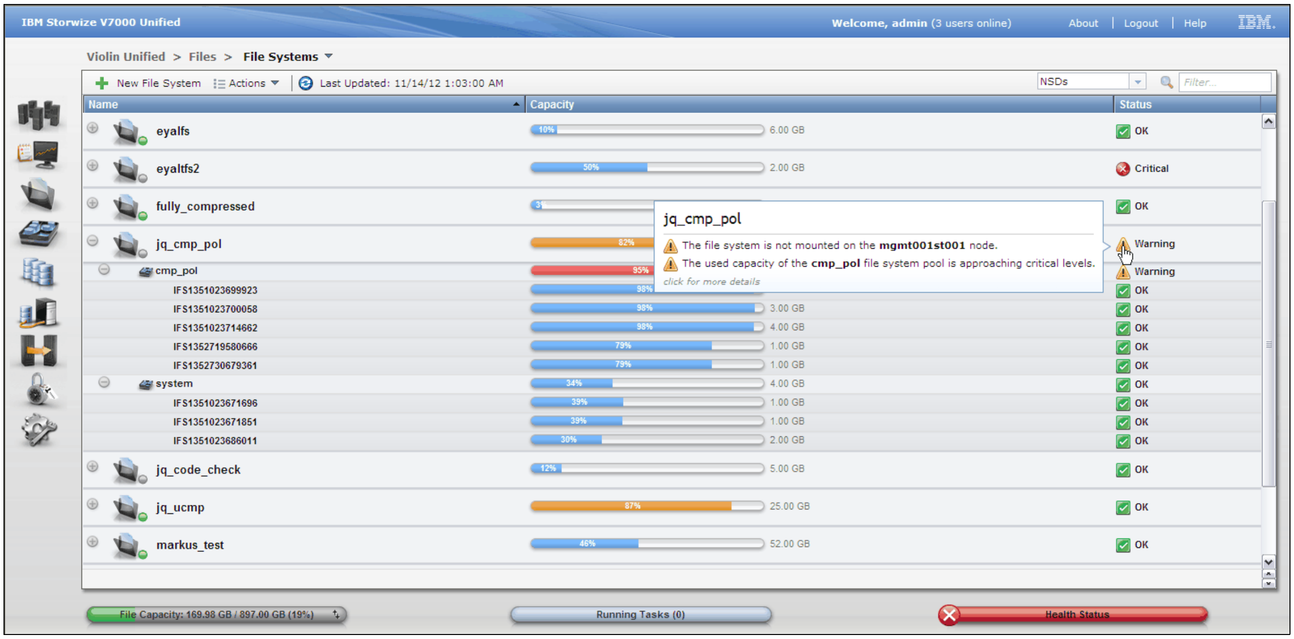
Click **Next** to continue or click **Cancel** to exit without marking any events as fixed.

Figure 15-12 Fix the problem

When the issue that caused the NSD to go offline is resolved, you must specify that the problem is fixed. This causes the NSD to come back online, and the 1865 error is cleared.

After the NSD is back online, the file system remains unmounted. Now you must follow the steps that are provided by the file events.

The first step is to verify the state of the file system, as shown in Figure 15-13, which indicates that the file system is not mounted.



IBM Storwize V7000 Unified
Welcome, admin (3 users online)
About | Logout | Help
IBM.
Violin Unified > Files > File Systems
New File System
Actions
Last Updated: 11/14/12 1:03:00 AM
NSDs
Filter...

Name	Capacity		Status
eyalfs	10%	6.00 GB	OK
eyaltfs2	50%	2.00 GB	Critical
fully_compressed	3%		OK
jq_cmp_pol	82%		Warning
cmp_pol	95%		Warning
IFS1351023699923	98%		OK
IFS1351023700058	98%	3.00 GB	OK
IFS1351023714662	98%	4.00 GB	OK
IFS1352719580666	79%	1.00 GB	OK
IFS1352730679361	79%	1.00 GB	OK
system	34%	4.00 GB	OK
IFS1351023671696	39%	1.00 GB	OK
IFS1351023671851	39%	1.00 GB	OK
IFS1351023686011	30%	2.00 GB	OK
jq_code_check	12%	5.00 GB	OK
jq_ucmp	87%	25.00 GB	OK
markus_test	46%	52.00 GB	OK

jq_cmp_pol

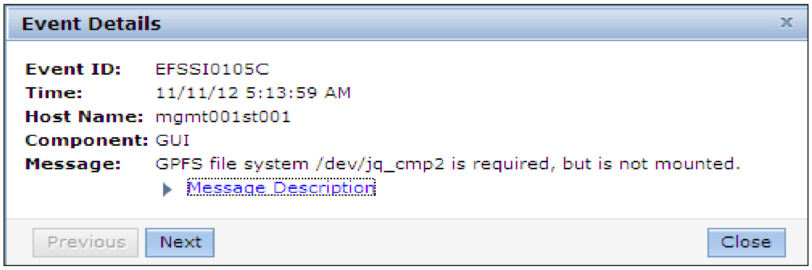
The file system is not mounted on the **mgmt001st001** node.

The used capacity of the **cmp_pol** file system pool is approaching critical levels.

click for more details

Figure 15-13 File system is not mounted

By viewing the file event log for the file system-unmounted error and using the Actions pull-down menu (or right-clicking) to display the properties, you can see the event details, as shown in Figure 15-14. There is a Message Description hyperlink that connects to the Storwize V7000 Unified Information Center. This link takes you directly to a description of the error code and instructions to get the file system checked and remounted.



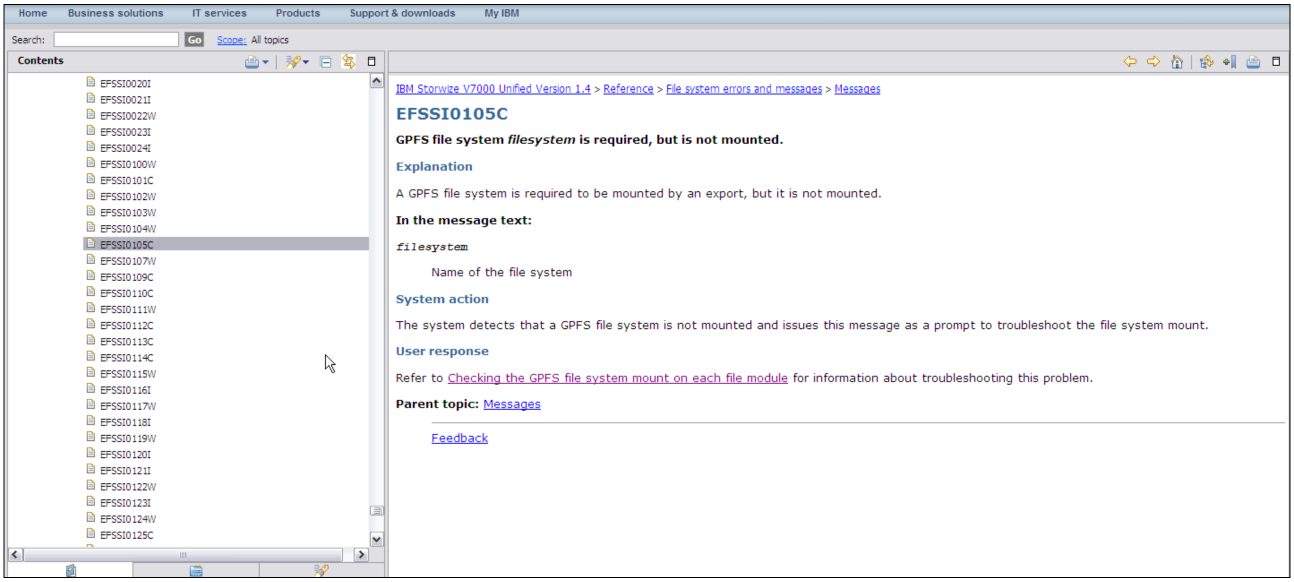
Event Details

Event ID: EFSSI0105C
Time: 11/11/12 5:13:59 AM
Host Name: mgmt001st001
Component: GUI
Message: GPFS file system /dev/jq_cmp2 is required, but is not mounted.

Figure 15-14 "File system not mounted" notice

Note: The steps that follow were verified while writing this book. However, we suggest that you check the information center, because it might offer updated steps to resolve a file system unmounted condition.

The first page that the hyperlink shows in the information center is shown in Figure 15-15.



EFSSI0020I
EFSSI0021I
EFSSI0022W
EFSSI0023I
EFSSI0024I
EFSSI0100W
EFSSI0101C
EFSSI0102W
EFSSI0103W
EFSSI0104W
EFSSI0105C
EFSSI0107W
EFSSI0109C
EFSSI0110C
EFSSI0111W
EFSSI0112C
EFSSI0113C
EFSSI0114C
EFSSI0115W
EFSSI0116I
EFSSI0117W
EFSSI0118I
EFSSI0119W
EFSSI0120I
EFSSI0121I
EFSSI0122W
EFSSI0123I
EFSSI0124W
EFSSI0125C

IBM Storwize V7000 Unified Version 1.4 > Reference > File system errors and messages > Messages

EFSSI0105C

GPFS file system *filesystem* is required, but is not mounted.

Explanation

A GPFS file system is required to be mounted by an export, but it is not mounted.

In the message text:

filesystem

Name of the file system

System action

The system detects that a GPFS file system is not mounted and issues this message as a prompt to troubleshoot the file system mount.

User response

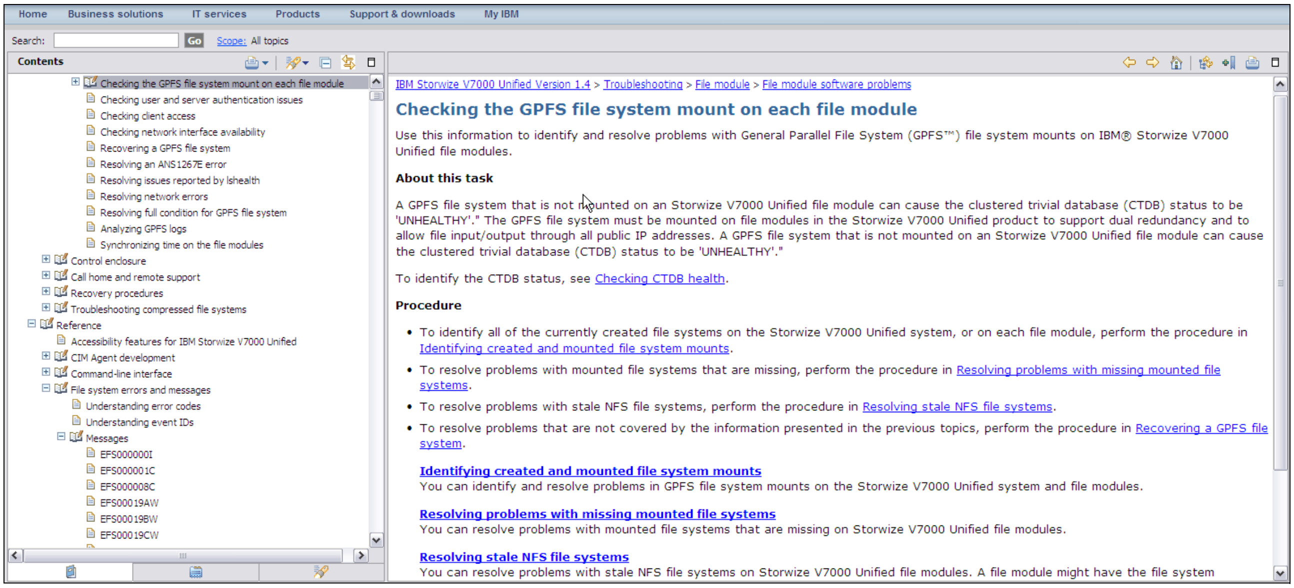
Refer to Checking the GPFS file system mount on each file module for information about troubleshooting this problem.

Parent topic: Messages

Feedback

Figure 15-15 Event EFSSI0105C

Select the **User response** link to get details and steps that are needed to bring the file system online again. Ensure that the steps are performed in the same sequence and completed. The high-level steps are shown in Figure 15-16 on page 272.


Home
Business solutions
IT services
Products
Support & downloads
My IBM
Search:
Go
Scope: All topics
Contents
Checking the GPFS file system mount on each file module
Checking user and server authentication issues
Checking client access
Checking network interface availability
Recovering a GPFS file system
Resolving an ANS1267E error
Resolving issues reported by lshealth
Resolving network errors
Resolving full condition for GPFS file system
Analyzing GPFS logs
Synchronizing time on the file modules
Control enclosure
Call home and remote support
Recovery procedures
Troubleshooting compressed file systems
Reference
Accessibility features for IBM Storwize V7000 Unified
CIM Agent development
Command-line interface
File system errors and messages
Understanding error codes
Understanding event IDs
Messages
EFS000000I
EFS000001C
EFS000008C
EFS00019AW
EFS00019BW
EFS00019CW
IBM Storwize V7000 Unified Version 1.4 > Troubleshooting > File module > File module software problems

Checking the GPFS file system mount on each file module

Use this information to identify and resolve problems with General Parallel File System (GPFS™) file system mounts on IBM® Storwize V7000 Unified file modules.

About this task

A GPFS file system that is not mounted on an Storwize V7000 Unified file module can cause the clustered trivial database (CTDB) status to be 'UNHEALTHY'." The GPFS file system must be mounted on file modules in the Storwize V7000 Unified product to support dual redundancy and to allow file input/output through all public IP addresses. A GPFS file system that is not mounted on an Storwize V7000 Unified file module can cause the clustered trivial database (CTDB) status to be 'UNHEALTHY'."

To identify the CTDB status, see Checking CTDB health.

Procedure

- To identify all of the currently created file systems on the Storwize V7000 Unified system, or on each file module, perform the procedure in Identifying created and mounted file system mounts.
- To resolve problems with mounted file systems that are missing, perform the procedure in Resolving problems with missing mounted file systems.
- To resolve problems with stale NFS file systems, perform the procedure in Resolving stale NFS file systems.
- To resolve problems that are not covered by the information presented in the previous topics, perform the procedure in Recovering a GPFS file system.

Identifying created and mounted file system mounts
You can identify and resolve problems in GPFS file system mounts on the Storwize V7000 Unified system and file modules.

Resolving problems with missing mounted file systems
You can resolve problems with mounted file systems that are missing on Storwize V7000 Unified file modules.

Resolving stale NFS file systems
You can resolve problems with stale NFS file systems on Storwize V7000 Unified file modules. A file module might have the file system

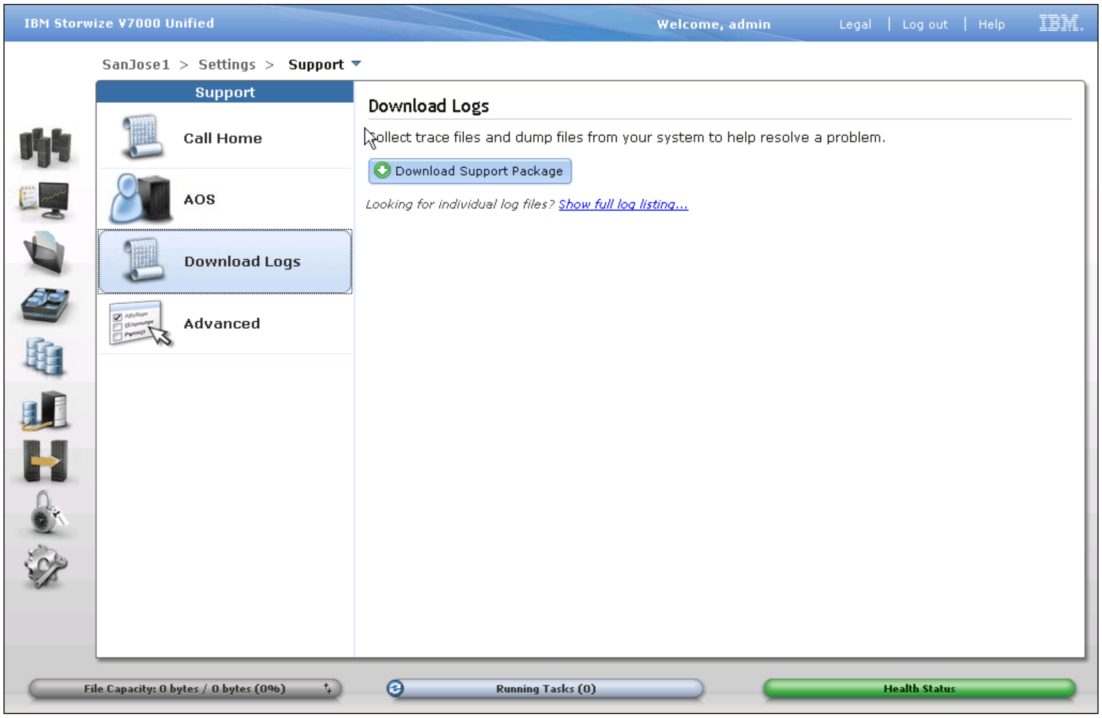
Figure 15-16 Checking the General Parallel File System (GPFS)

15.3 Collect support package

> **Note:** The workstation that you are using to connect to the cluster needs to have Internet access to be able to display the information center.

The three modules of the Storwize V7000 Unified system each have their own data collection processes. Each file module collects its own log and configuration files, and the storage control enclosure also collects its own snap file. The cluster triggers these collections and combines them into a single file so that only one operation is required.

This is triggered from the cluster graphical user interface (GUI). Select **Settings** → **Support** → **Download Logs** to open the Download Logs window that is shown in Figure 15-17 on page 273.



Download Logs

Collect trace files and dump files from your system to help resolve a problem.

Download Support Package

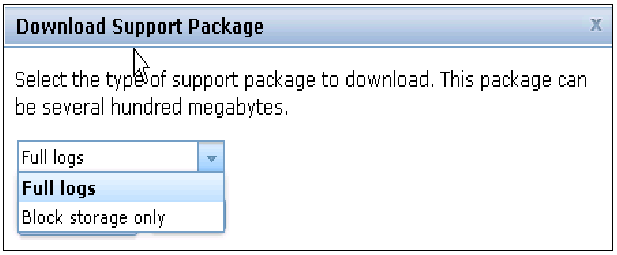
Looking for individual log files? Show full log listing...

Figure 15-17 Download Logs window

Here, you can get a listing of the packages that are currently stored on the cluster and you can initiate a current data collection. If the file list is not displayed, select **Show full log listing** to see the file list. Then, select which node to view from the pull-down menu.

> **Caution:** Each module stores only the data that it collects. The data file is stored by the current management node. Ensure that you are looking at the listing for the correct node.

To collect the logs, click **Download Support Package**. When you are asked for the type of logs that are required, as shown in Figure 15-18, choose **Full logs** and then click **Download** to get all of the logs. A progress window opens. Wait for the entire task to complete and confirmation that it was successful. Then, close the window.



Select the type of support package to download. This package can be several hundred megabytes.

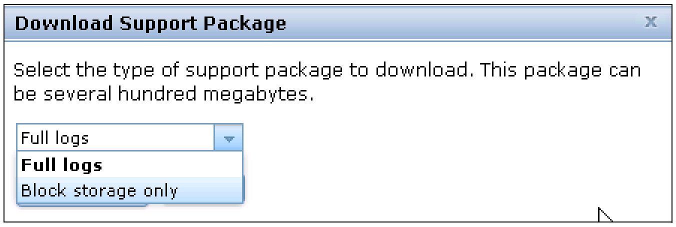
Figure 15-18 Download logs: Full logs

Locate the resultant file in the file list. The file name includes a date-time stamp that you can use to identify the recent log. Highlight the file, and either right-click or use the **Action** pull-down menu, and then select **Download**. You can save this file to your workstation, and it is ready to upload to IBM Support. Avoid renaming this file.

> **Note:** The files remain on the file module indefinitely. It is your job to maintain them and to delete old copies. There is no concern about how many remain on the module other than confusion. Our suggestion is to delete files only after the problem is resolved, and always keep the last one.

If you are using compression, the receive-any control element (RACE) module maintains internal diagnostic information, which is kept in memory in each of the V7000 nodes in a compression I/O group. This information is available as part of the support package for the block storage. The full logs contain the standard logs plus the *most recent* statesave from each node. This is fine if the system observed a failure. For all other diagnostic purposes, such as a performance issue, the standard logs plus *new* statesaves is required.

To get the *standard logs plus new state saves*, select **Block storage only**, as shown in Figure 15-19.

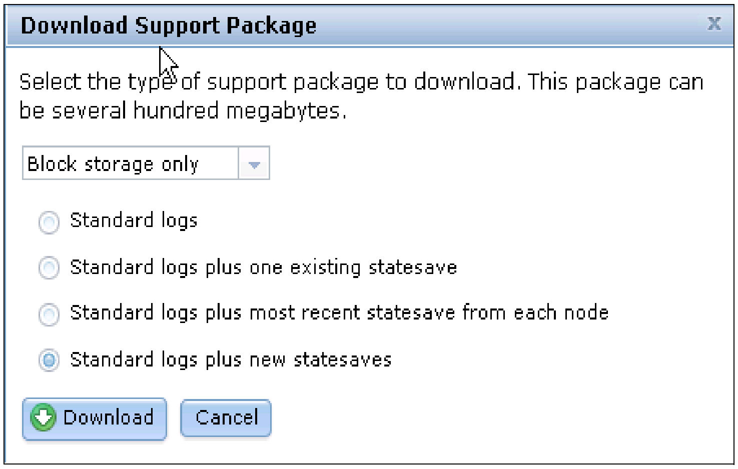


Download Support Package

Select the type of support package to download. This package can be several hundred megabytes.

Figure 15-19 Block storage only

Then, select **Standard logs plus new statesaves**, as shown in Figure 15-20.



Download Support Package

Select the type of support package to download. This package can be several hundred megabytes.

Block storage only

○ Standard logs

○ Standard logs plus one existing statesave

○ Standard logs plus most recent statesave from each node

● Standard logs plus new statesaves

Figure 15-20 Standard logs plus new statesaves

15.4 Information center

All documentation for the Storwize V7000 Unified system is compiled in the online *IBM Storwize V7000 Unified Information Center*. This includes installation and configuration, troubleshooting, and administration. Figure 15-21 on page 275 shows the information center home page:

http://publib.boulder.ibm.com/infocenter/storwize/unified_ic/index.jsp

As Figure 15-21 shows, the page has three main areas: search bar, left pane, and right pane. The left pane has three tabs at the bottom to choose the view that is displayed: contents (default), index, and search results. The right pane displays the selected detail page. The search bar can be used for a context search of the entire information center by using rich search arguments.

15.4.1 Contents

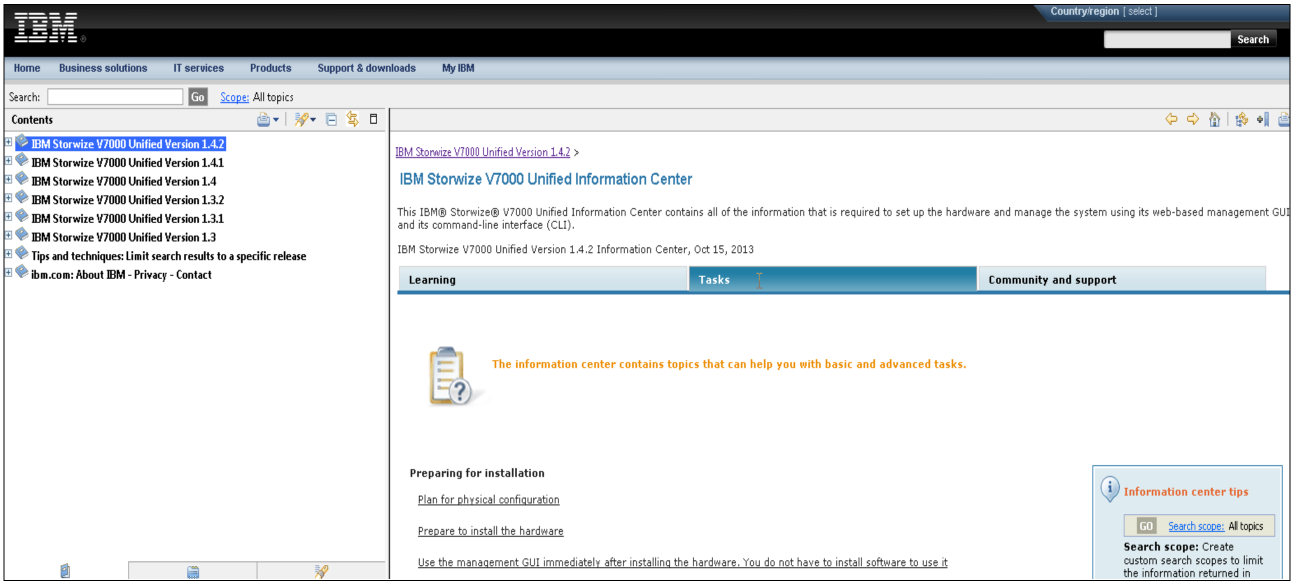
This tab in the left pane displays an expandable map view of the headings and subheadings of the information. You can use this view to see all information available under a subject heading and to go to the information quickly.

15.4.2 Index

In this tab, there is an alphabetical listing of keywords that you can scroll through or search by using the Search field at the top of the panel.

15.4.3 Search results

The results of a search that is performed in the search toolbar at the top of the page are listed in this tab. For each result, the section heading that matches the search argument is given, along with an abbreviated summary of the text in that section. By clicking the heading, which is a URL, the page is displayed in the right panel. See Figure 15-21.



- IBM Storwize V7000 Unified Version 1.4.2
- IBM Storwize V7000 Unified Version 1.4.1
- IBM Storwize V7000 Unified Version 1.4
- IBM Storwize V7000 Unified Version 1.3.2
- IBM Storwize V7000 Unified Version 1.3.1
- IBM Storwize V7000 Unified Version 1.3
- Tips and techniques: Limit search results to a specific release
- ibm.com: About IBM - Privacy - Contact

IBM Storwize V7000 Unified Version 1.4.2 >

IBM Storwize V7000 Unified Information Center

This IBM® Storwize® V7000 Unified Information Center contains all of the information that is required to set up the hardware and manage the system using its web-based management GUI and its command-line interface (CLI).

IBM Storwize V7000 Unified Version 1.4.2 Information Center, Oct 15, 2013

The information center contains topics that can help you with basic and advanced tasks.

Preparing for installation

Plan for physical configuration

Prepare to install the hardware

Use the management GUI immediately after installing the hardware. You do not have to install software to use it

Search scope: Create custom search scopes to limit the information returned in

Figure 15-21 IBM Storwise V7000 Unified Disk System Information Center: Home

The best way to locate information in the information center is to construct a good search argument for the information that you want to see. Enter that into the Search field at the top of the page. Scan the results for the section that meets your requirements, and display each one in the right window by clicking the URL. If you want to look at several, use your browser’s functions to open the links in a new tab or window. Hovering your mouse cursor over the URL displays a pop-up window that shows the source of the item. Be careful about multiple similar hits, where some are marked *previous version*.

15.4.4 Offline information center

For situations where Internet access is not possible or poor, a downloadable version of the information center is available. Go to the Download section of the Support page to locate the link to the files. The download version that is current at the time of publication is *V1.4.0 Installable Information Center for IBM Storwize V7000 Unified*:

http://www.ibm.com/support/docview.wss?uid=ssg1S4001035

Note: The download is very large, over 1 GB. When installed, it requires a substantial amount of storage space.

There are two modes available for installing the offline information center: Either on your workstation for personal use or on a server for use by anyone in your organization. The `readme` file gives setup instructions for each of those methods.

Caution: When downloaded, the information in the information center is not updated and might not reflect the most current detail. In general, the information that is provided is correct and valid, but if there is any doubt, particularly with recovery procedures, consult the online system to confirm. To maintain concurrency, the file needs to be downloaded and the offline information center reinstalled regularly.

The only way to learn what the information center offers is to use it. Spend time looking at what information is available and how to find it. Practice with searches and locating procedures and references. In a short time, you will be comfortable with the layout of the information and be able to locate topics quickly.

15.5 Call Home feature and alerts

Although the cluster automatically detects and logs events and errors, alerts are necessary to ensure that the user is made aware of an issue promptly. The cluster incorporates three methods for alerts. For the configuration of these services, see 11.8.3, “Support” on page 182.

15.5.1 Simple Network Management Protocol

This agent uses the well-known Simple Network Management Protocol (SNMP). Alerts are sent to the designated SNMP server. System health reports and error alerts are determined by the individual implementation of the SNMP process within each enterprise, and those instructions are beyond the scope of this book.

The cluster is configured with the address of the SNMP server. The severity of each alert type can be set to filter the alerts that are presented. Multiple server targets can be defined and configured individually.

15.5.2 Email

This is a versatile alert method that uses email to send an alert. It can be sent to your email address, a group email address, or even a server address.

First, the cluster is configured to allow email alerts. The SMTP server that it sends the email to must be defined. Next, each recipient needs to be defined. As many email recipients as you want can be configured. Each definition needs the email address and can be individually set up for the type of event to alert and the severity of each type. This offers flexibility for users to receive only the alerts that interest them.

> **Tip:** Resist the urge to enable all levels of alerts. This generates a high volume of alerts because it sends notices of every minor change in the system. We suggest “Critical only” for most users and “Critical/Warning” for normal monitoring.

15.5.3 Call Home

Call Home is one of the IBM Support tools included to assist IBM representatives in supporting your system. By using the same process as the email alerts, email is sent directly to an IBM server on the web, which includes a brief log of the error. If your system is properly set up in the IBM Support system, this generates a problem management record (PMR). A call is then queued to the Support team for your region.

If you did not already called IBM Support, depending on the nature and severity of the error, a Support representative might call you. This call home process gives IBM Support teams timely awareness of the problem and allows them to begin analysis. It also serves as a backup for alerts that are sent to you.

> **Important:** Although IBM Support might receive a direct and early alert for an error, this does not constitute a service call. It is the client’s responsibility to ensure that alert notices within their enterprise are effective and that there are responses to critical events. Always place a service call unless someone from IBM Support has already called you. In most cases, the call home improves IBM responsiveness and leads to faster analysis and recovery.

15.6 IBM Support remote access

There might be times during an outage that IBM Support needs to connect to the cluster. This might be to speed up problem determination and recovery by allowing IBM Support staff to issue commands and see responses directly. Or, there might be commands and access required that can be performed only by IBM.

IBM Support teams use the *Assist On-site* (AOS) tool, which they have used for many years to support of IBM System x and storage products. This involves a daemon that is loaded onto your workstation or system. It connects to a secure server within IBM. Then IBM Support specialists can access this server internally, view your console, and use the keyboard and mouse. Access within IBM is strictly controlled and restricted. If required, when a session is established, other approved IBM specialists can also monitor the session. This enables a cooperative approach, which can hasten recovery time.

For a Windows workstation, you access a web page and register. A small temporary agent is downloaded to provide the client-side daemon. Access is through well-known Ports 80 and 433 and is secure. A session-unique pass key, which is passed verbally, must be entered to complete the connection. Your specialist and the IBM specialist share the video, mouse, and keyboard, and you maintain full control.

With the Storwize V7000 Unified system, permanent client software is included in the package. This is disabled by default and must be enabled and configured by the user. There are two modes of operation: lights on and lights out. *Lights on* indicates that your site is staffed at all times. Any connection attempt requires authorization by the client that is connecting locally to the cluster GUI and approval of the connection. *Lights out* implies that the cluster is in an unstaffed location (at least for some of the time). Connection occurs without the need for local approval. Use this setting to define how you want IBM Support to connect, based on your local security policies. We suggest lights on for protection and awareness, provided that approval through the GUI can be given in a timely manner, if needed.

Setup of the AOS connection is covered in the Chapter 11, “Implementation” on page 141.

15.7 Changing parts

If a part or component of the Storwize V7000 Unified system fails, your first indication is an entry in the event log and, probably, an alert, depending on your alert settings. Depending on the nature of the failure and the impact, you might run guided maintenance procedures for the event (in the case of block storage). Or you might research the error code in the information center. If the procedures indicate that a part must be changed, you need to call IBM, using your local procedures. If at any time during your analysis of the problem you feel uncomfortable or the problem is pervasive, immediately call IBM for assistance.

The IBM Support representative is likely to ask you for a data collection to be able to review the logs, check other related information in the data, and research the problem in internal databases. The IBM specialist might ask you to perform other recovery actions, to gather additional data, or to run commands. If it is determined that a part needs replacement, the representative arranges for shipment of the part and, depending on which part, gives you instructions on the replacement procedures or sends a service representative.

The Storwize V7000 Unified system is designed so that nearly all parts can be replaced by the client’s team. This speeds up recovery and reduces cost. Replacement is typically under the guided maintenance procedures, which are based on reports in the error log. IBM calls these parts client-replaceable units, or CRUs.

If the part is not a CRU or if more technical problem determination is required, IBM Support dispatches a service representative.

It is important to ensure that you are familiar with the safety procedures that are described in the information center and have a good understanding of the process before you change any parts. If at any time you feel unsure, call IBM Support for guidance.

CRU parts are typically sent by courier to your location. Depending on procedures for your country, this shipment includes a process to return the removed part, which it is important to do promptly.

15.8 Preparing for recovery

When a disaster strikes, getting the problem diagnosed and systems recovered is generally time-sensitive. Many tasks and procedures are encountered or needed only at the worst time possible. If you are unprepared for any of these or have not tested the procedures, that frequently leads to an outage that is longer than it needed to be.

The following are several key items that, if prepared beforehand, greatly improve the diagnosis and recovery time.

15.8.1 Superuser password

This is the password for the *superuser* user ID, which is used to log on to the Storwize V7000 storage enclosure directly. In the unified cluster, this password is seldom if ever used, because all functions are performed from the V7000 Unified GUI and CLI. But if certain problems occur on the storage component, this password becomes essential. If it is not handy, that delays service actions.

Store the password securely, but ensure that it can be accessed quickly.

15.8.2 Admin password

Although this user ID might be in regular use, the password needs to be made available on short notice when service is required. If the user ID and password are held aside and all users have their own IDs, a user ID and password for an account with full authority needs to be available for service use. Alternative methods of producing a valid ID and password need to be in place if the authorized person is not present or available.

15.8.3 Root password

As described in the Implementation section, the root password to the file modules is widely known to IBM Support personnel. If this poses a risk, we suggest that you ask IBM personnel to change it at installation time. If this is done, the password needs to be secured and able to be produced on demand. Currently, several service and recovery procedures require root access. If it is not available, IBM Support's ability to recover a problem is compromised.

IBM intends to negate the need for field support to use this level of access in the future. When this happens, this action will no longer be required.

15.8.4 Service IP addresses

An often overlooked step during implementation is to set the IP addresses for service access to the storage nodes and the file modules. The storage node service IP addresses have default settings, but these are unlikely to match your network, where local and direct physical access are necessary to connect. Set them to assigned IPs in the local LAN so that they are accessible from your network. These addresses need to be documented and easily accessed.

Considerable time is wasted if these addresses are not known or not in the local networks.

15.8.5 Test the GUI connection to the Storwize V7000 storage enclosure

As part of the installation process, confirm that you can connect to the Storwize V7000 storage nodes directly. Ensure that this is tested regularly thereafter. You need to be able to connect to and log on to the storage system GUI over the management address and to both the Service Assistant GUIs over the two service IP addresses.

Also, set up and test both of the CLI connections.

15.8.6 Assist On-site

The IBM software includes a tool that allows simple access to the Storwize V7000 Unified system. Assist On-site (AOS) needs to be configured, so ensure that this is done and tested before it is needed to ensure that no time is lost if there is a need for IBM Support to connect to the cluster. This includes resolving any firewall and security issues. We encourage you to do this configuration and testing during the implementation.

15.8.7 Backup config saves

Before any maintenance activity, do a config backup and then offload it either as files or by doing a normal data collection. This greatly assists IBM Support in the event of a problem and might increase the likelihood of a successful recovery.

For the storage units, issue this CLI command:

```
svcconfig backup
```

This command gathers and saves a current version of the `config.xml` file. Then, do a data collection as normal, offload, and save the file safely. It is wise to perform this activity at regular intervals (for example, monthly) as part of routine housekeeping.

15.9 Software

All of the software that is used in the Storwize V7000 Unified system is bundled into a single package. Each bundle is uniquely named by using the IBM standard for software name and version control. The bundle is identified by its Version.Release.Modification.Fix (VRMF) number (for example, GA release was 1.4.0.0). These numbers increase as new code is released. The meaning of each field is as follows:

Version Increments increase only for a major change to the functions and features of the software product.

Release Increments increase as new functions or features are released.

Modification Increments increase when a function or feature is modified.

Fix This number increases in increments for each group of program temporary fixes (PTFs) that are released. Notice that PTFs do not change function. They fix code problems only.

15.9.1 Software package

There are five different software packages that are used with the Storwize V7000 Unified Disk System:

- **Full software installation DVD.** This is needed only in the case of a full rebuild of a cluster. It is a bootable DVD that reloads all of the file module software, management software, updates hardware (HW), basic input/output system (BIOS), and utilities, if needed, and updates Storwize V7000 software if required. This process initializes the system and *destroys previous configurations*. This is an ISO image and can be obtained only from IBM Support for a specific situation.
- **Software update.** This file is used to concurrently upgrade the file module software, management software, update HW BIOS and utilities if required, and update Storwize V7000 software if required. This file can be downloaded from IBM and is installed by using the upgrade tools in the Storwize V7000 Unified cluster.
- **Storwize V7000 control enclosure software update.** This update file is included in the Storwize V7000 Unified package and is managed and installed by the file module management software. It is automatically updated if required. Do not attempt to upgrade the storage control enclosure manually unless directed to do so by IBM Support.
- **File module HW BIOS and utilities.** This firmware is managed by the Storwize V7000 Unified software and is automatically upgraded if required. *Do not* attempt to manually upgrade this firmware.
- **Install test utility.** This small package is loaded in the same way as the normal software update. It installs a test utility that then checks for known issues or conditions that can potentially cause the software Apply to fail. It is important to run this utility first.

> **Caution:** Do not use the DVD to upgrade an operational system. The DVD is used only to install and initialize a new system or to recover a system. Installing the software from the DVD *destroys* the existing configuration.

Download

If your Storwize V7000 Unified system has a connection to the Internet and sufficient bandwidth is available to perform the download, log on to the management GUI and, from the upgrade software option, download the upgrade file. If connectivity is not available or you do not want a direct download, then download the files from the web as described here.

Web download

All software packages are downloaded from IBM Fix Central. To access this site, you need an IBM ID. If you do not already have one, there is a link on the first page of the site to register.

Select these two packages from the list of available software.

1. Select and download the latest UpgradeTestUtility. This is a small file that contains the test utility that you install first to check for known problems.
2. Select and download the latest Storwize V7000 Unified software bundle. This includes the main upgrade file. It also includes the license agreement, the release notes, and a small file with the MD5 checksum value. It is important to read the release notes carefully. This file constitutes the “read me” file and special instructions for this release.

You can get the latest software from the Fix Central website:

`http://www.ibm.com/support/fixcentral/`

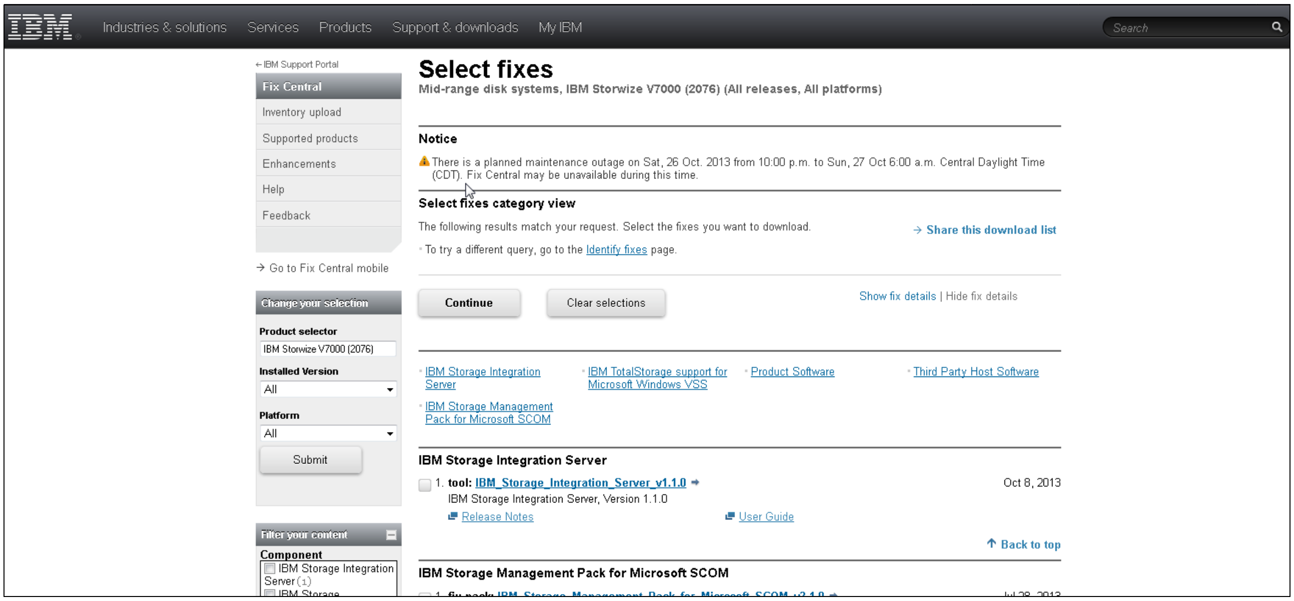
In the **Find product** tab of Fix Central, type the following into the **Product selector** field:

```
IBM Storwise V7000 Unified
```

Then, select your **Release** and **Platform** and click **Continue**.

The next window shows a list of downloads for the Storwize V7000 Unified system. Click the latest version to see the downloadable files page. Scroll down and review, as required, the information that is given on this page, which includes details and links for compatibility tables, interoperability, and restrictions. Package download links are at the bottom. Choose the applicable line and select a download option (HTTP or, if available, FTP).

Read and accept the terms and conditions in the pop-up window. This action then takes you to the Download page that is shown in Figure 15-22.



Select fixes

Mid-range disk systems, IBM Storwize V7000 (2076) (All releases, All platforms)

Notice

There is a planned maintenance outage on Sat, 26 Oct. 2013 from 10:00 p.m. to Sun, 27 Oct 6:00 a.m. Central Daylight Time (CDT). Fix Central may be unavailable during this time.

Select fixes category view

The following results match your request. Select the fixes you want to download.

- To try a different query, go to the Identify fixes page.

Product selector: IBM Storwize V7000 (2076)

Installed Version: All

Platform: All

- IBM Storage Integration Server
- IBM TotalStorage support for Microsoft Windows VSS
- Product Software
- Third Party Host Software
- IBM Storage Management Pack for Microsoft SCOM

IBM Storage Integration Server

□ 1. tool: IBM_Storage_Integration_Server_v1.1.0 — Oct 8, 2013

IBM Storage Integration Server, Version 1.1.0

Release Notes   User Guide

IBM Storage Management Pack for Microsoft SCOM

Component

□ IBM Storage Integration Server (1)

□ IBM Storage

Figure 15-22 Software download

Ensure that the package that you need is selected and click **Continue**. Follow the prompts to save the package.

15.9.2 Software upgrade

> **Important:** Always read the release notes before continuing.

There are two methods of installing software on the Storwize V7000 Unified system.

The normal process is to perform an upgrade, which is applied concurrently. This involves a stepped process of stopping and upgrading one file module, and then bringing it back online. After a settling time, the other file module is handled in the same way. If required, the control enclosure is also upgraded. This involves stopping and upgrading one node, waiting 30 minutes, and then doing the other node. All of this is concurrent with the Storwize V7000 Unified system operation, and client access is not interrupted.

File system access switches between the modules, and host access continues uninterrupted if the hosts have network connectivity to both file modules. For block access over the SAN, all hosts must have operational multipathing and at least one path to each node canister.

Important: Ensure that all LAN-connected file system hosts confirm access to both file modules and that all SAN-attached Fibre Channel hosts have paths to both node canisters.

The second method involves restoring all software from DVD. This is disruptive and, by design, destroys the Storwize V7000 Unified system configuration, and any data that is stored on the system is lost. Unless this is a new installation or a rebuild, where no data exists on the system, *do not* perform this procedure without direct guidance from IBM Support.

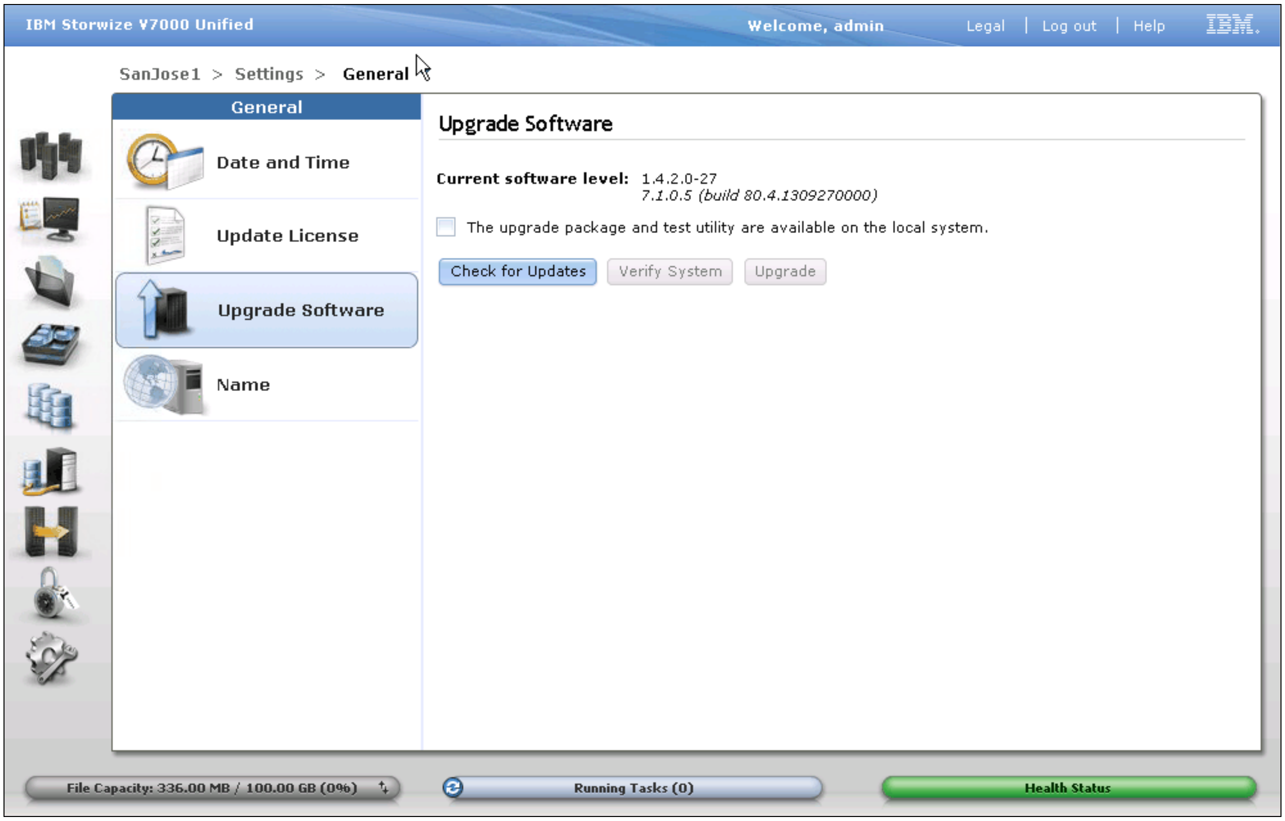
Install new software from DVD

This procedure is described in 11.5, "Install the latest software" on page 154. Use it only on a new system or under the direction of IBM Support during a recovery action.

Concurrent software upgrade

The following procedures are run from the Upgrade Software option on the management GUI of the Storwize V7000 Unified system.

Select **Settings** → **General** → **Upgrade** Software to open the window that is shown in Figure 15-23.



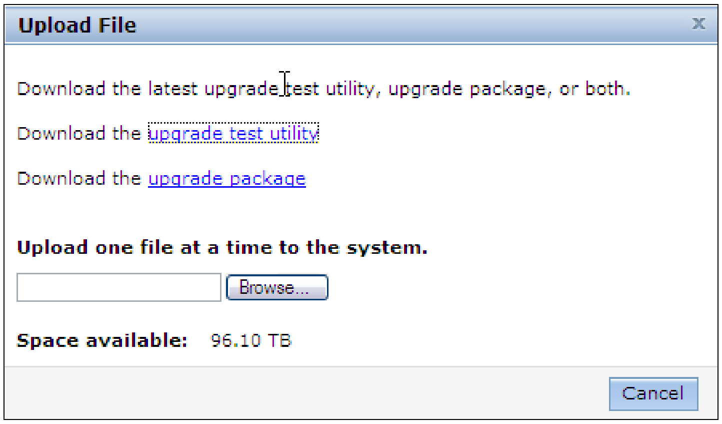
Current software level: 1.4.2.0-27
7.1.0.5 (build 80.4.1309270000)

□ The upgrade package and test utility are available on the local system.

Figure 15-23 Upgrade software: Home page

If your cluster has good web access, you can click **Check for updates** to connect to IBM and download the latest upgrade file to the cluster. This file is over 1 GB and might take time to download by using this method.

Alternatively, by using the procedure described here, connect to IBM Fix Central and download the upgrade file to your workstation. Then, you must upload the file to the cluster. Click **Upload File** on the page to open the pop-up window that is shown in Figure 15-24.



Download the latest upgrade test utility, upgrade package, or both.

Download the upgrade test utility

Download the upgrade package

Upload one file at a time to the system.

Space available: 96.10 TB

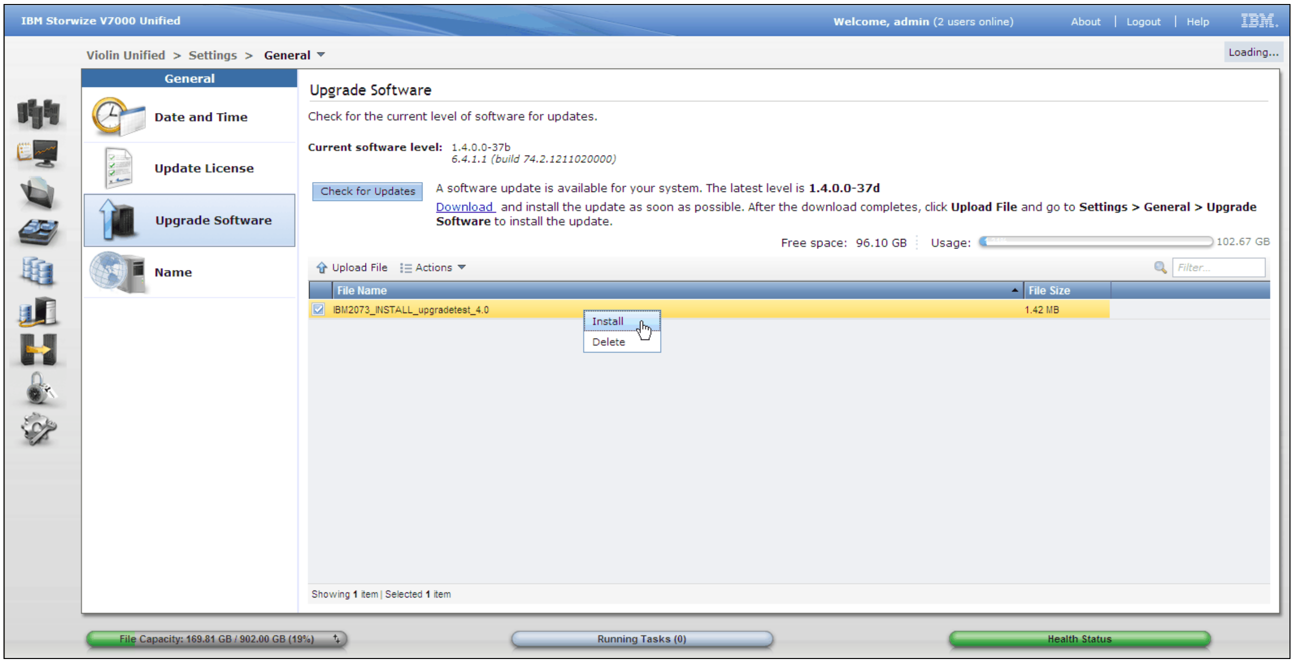
Figure 15-24 Upload File window

Tip: When the upgrade process is complete (success or fail), the file is deleted. For this reason, you might find it beneficial to download the file to your workstation and then to upload it to the cluster, rather than use the “Check for updates” button. This way, you always have a copy of the file and need to upload it again only in the event of a failure. This also saves time and bandwidth if you have multiple clusters. The likelihood of needing the file again on the same machine after the upgrade is successfully applied is very low, but this method offers peace of mind.

Browse to the upgrade file that you downloaded on your workstation, and click **OK**.

Repeat the upload step for the test utility that you also downloaded with the code file.

The new files are opened and interrogated by the cluster code. If they are acceptable, they are displayed in the list. First, run the upgrade test, always using the latest version of the utility that is available. Select the file. Then, right-click to select **Install**, or use the **Actions** pull-down menu, as shown in Figure 15-25 on page 285.



Upgrade Software

Check for the current level of software for updates.

Current software level: 1.4.0.0-37b
6.4.1.1 (build 74.2.1211020000)

A software update is available for your system. The latest level is **1.4.0.0-37d**
Download and install the update as soon as possible. After the download completes, click **Upload File** and go to **Settings > General > Upgrade Software** to install the update.

Free space: 96.10 GB Usage: 102.67 GB

File Name	File Size
☑ IBM2073_INSTALL_upgradetest_4.0	1.42 MB

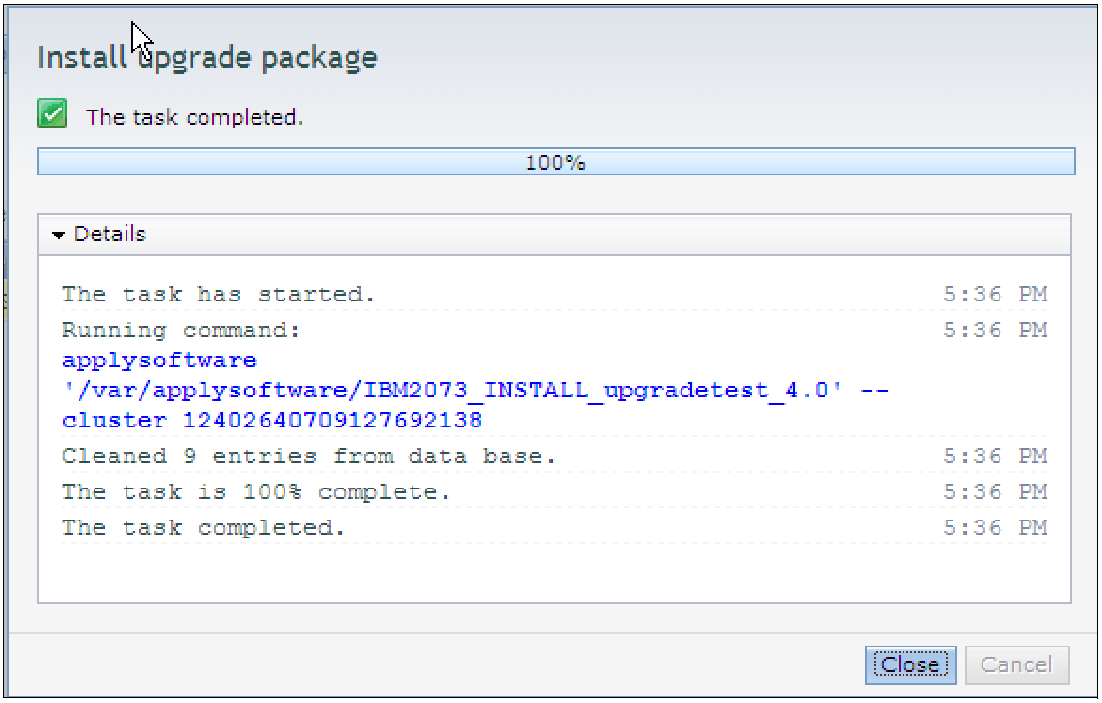
Showing **1** item | Selected **1** item

File Capacity: 169.81 GB / 902.00 GB (19%)

Figure 15-25 Upgrade software: Install test utility

Hint: You must select the file before you perform an action, which is shown as the area highlighted in yellow. Then, the actions are available.

Wait for the utility to run. When it is finished, review the result in the window before closing it. Figure 15-26 shows an example of a successful run.



☑ The task completed.

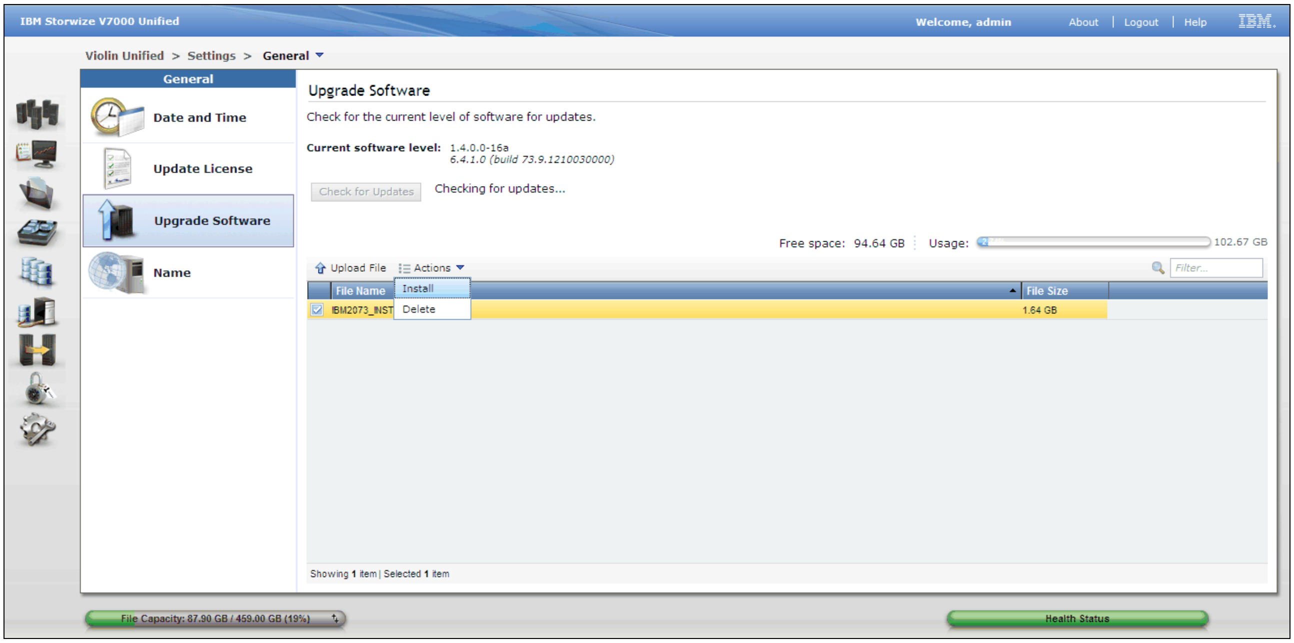
```
The task has started.                                       5:36 PM
Running command:                                            5:36 PM
applysoftware
'/var/applysoftware/IBM2073_INSTALL_upgradetest_4.0' --
cluster 12402640709127692138
Cleaned 9 entries from data base.                           5:36 PM
The task is 100% complete.                                  5:36 PM
The task completed.                                         5:36 PM
```

Figure 15-26 Upgrade software: Test results

If there were any problems reported, they must be resolved before continuing with the software upgrade.

> **Important:** While code upgrades are being applied, all data access continues. Block access continues to be provided over the SAN if hosts have multipathing correctly configured with active paths to both nodes, and hosts with file access have IP connectivity to both file modules. It supports a failover of the IP addresses between the file modules.
>
> However, do not make any configuration changes to the cluster during the software upgrade. Many tools and functions are temporarily disabled by the Apply process to prevent changes from occurring.

When you confirm that the test utility shows no problems, click the upgrade file to select it. Then, select **Install**, as shown in Figure 15-27.



Upgrade Software

Check for the current level of software for updates.

Current software level: 1.4.0.0-16a
6.4.1.0 *(build 73.9.1210030000)*

Checking for updates...

Free space: 94.64 GB Usage: 102.67 GB

File Name	File Size
IBM2073_INST	1.64 GB

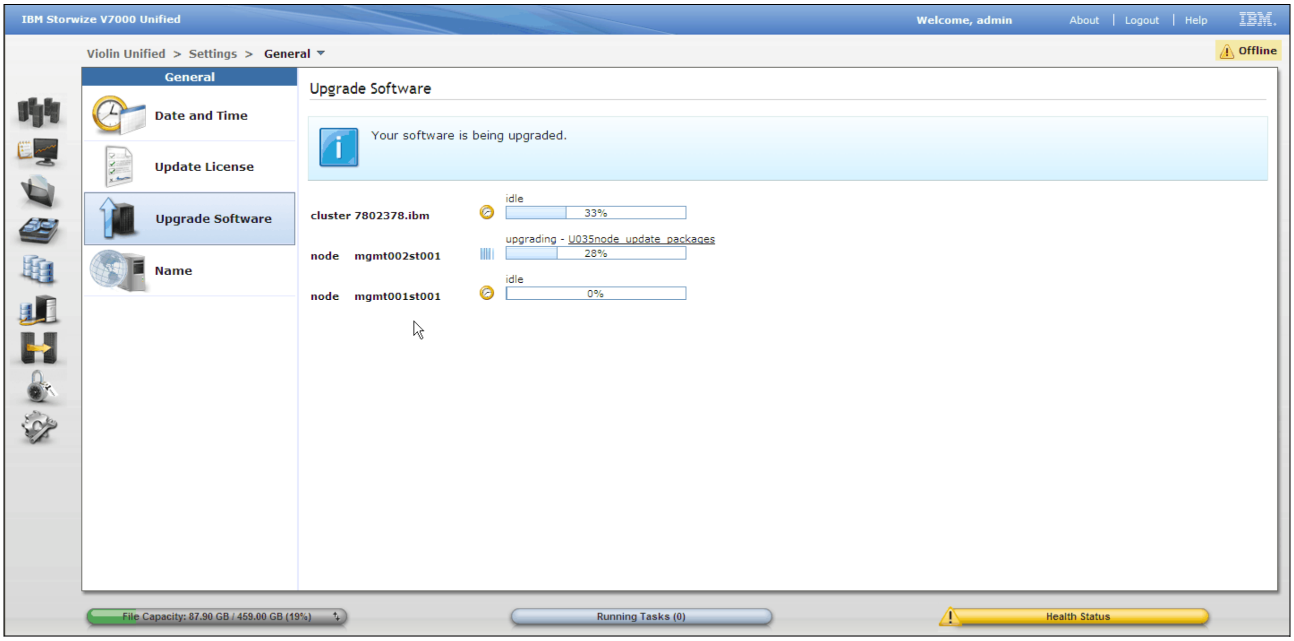
Showing **1** item | Selected **1** item

File Capacity: 87.90 GB / 459.00 GB (19%)

Figure 15-27 Upgrade software: Install

A pop-up progress window opens. Wait for the task to complete, and then close the window. This does not indicate that the upgrade is applied, but only that it has been successfully started. The upgrade software window now shows a status window with three progress bars. The top one is the overall progress, and the other two are for the file modules. If the new upgrade includes a version that is higher than what is currently on the storage control enclosure, it automatically performs an upgrade of the control enclosure first. This step takes about 90 minutes and can be monitored by the progress bar. It is complete when the bar is at 33%.

When the storage is complete, the process moves on to the first file module, as shown in Figure 15-28 on page 287.



Upgrade Software

Your software is being upgraded.

cluster 7802378.ibm — idle — 33%

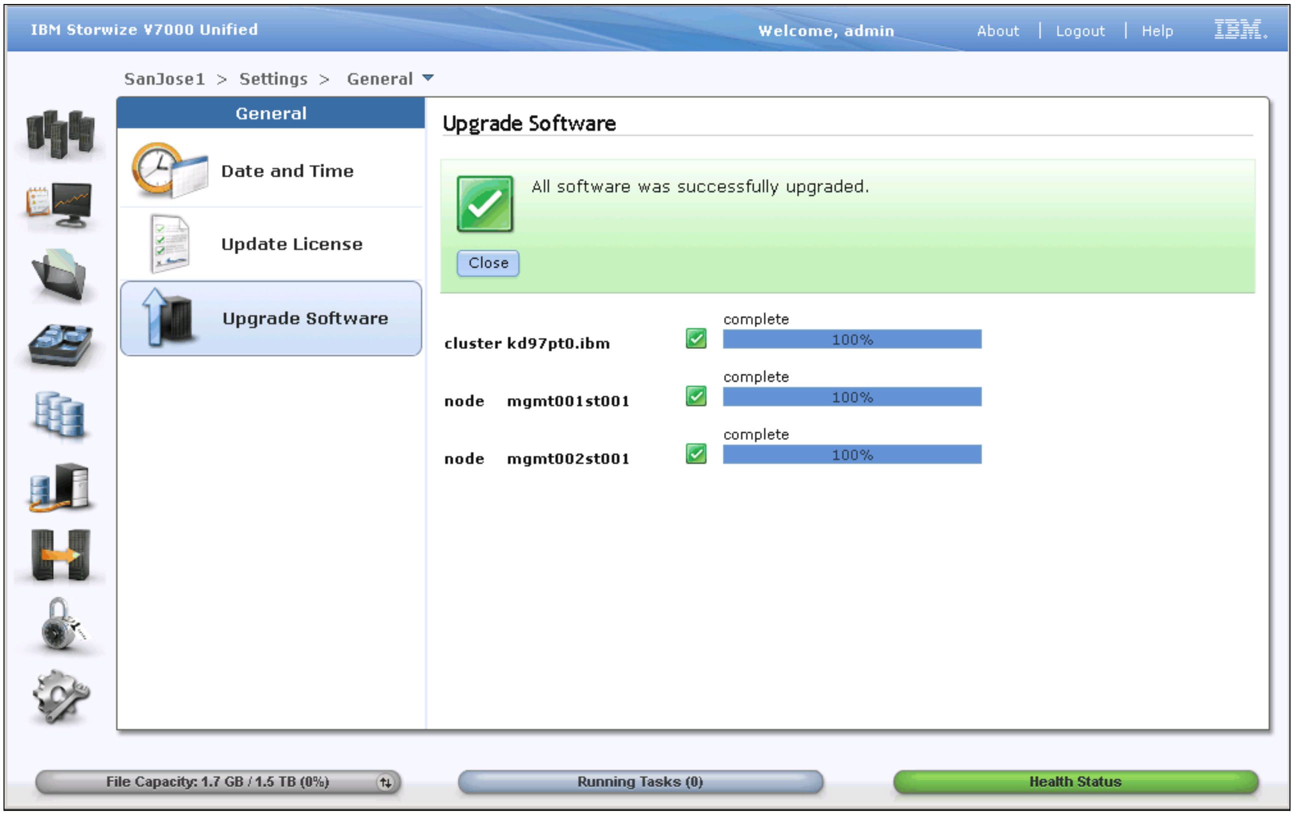
node mgmt002st001 — upgrading - U035node_update_packages — 28%

node mgmt001st001 — idle — 0%

Figure 15-28 Upgrade software: Progress

The file modules take about 30 minutes each to upgrade. When the upgrade on the first one is complete, the process moves on and upgrades the second one. At this time, the GUI connection is lost because the active management module is removed and replaced. You must log back on to continue monitoring the progress.

When all three modules are successfully upgraded, the process to upgrade cluster common modules and processes begins. This takes about 30 minutes. The upgrade is complete when this process finishes, as shown in Figure 15-29 on page 288.



All software was successfully upgraded.

```
cluster kd97pt0.ibm     ✓ complete 100%
node    mgmt001st001    ✓ complete 100%
node    mgmt002st001    ✓ complete 100%
```

Figure 15-29 Upgrade software: Complete

Your software is now upgraded.

16

IBM Real-time Compression in the Storwize V7000 Unified system

This chapter explains ways to use IBM Real-time Compression on the IBM Storwize V7000 Unified Disk System and provides an example. It also describes planning, managing, and reporting related to compression.

16.1 Compression and block volume compression use cases

Real-time Compression in SAN Volume Controller and Storwize V7000, REDP-4859, also an IBM Redbooks publication, describes the base technology for real-time compression in detail and describes the compression use cases for compressed block volumes. Support for Real-time Compression within the Storwize V7000 Unified system started with Version 1.4. It uses the Real-time Compression function in the Storwize V7000 subsystem and applies it to block volumes that are used for file systems by the V7000 Unified system.

The following sections cover specific considerations of compression in the Storwize V7000 Unified file system.

16.2 Compressed file system pool configurations

Using the feature to compress volumes with the NAS feature of the Storwize V7000 Unified system provides the following advantages over using compression with an external host that creates a file system on a compressed volume:

- The policy language capabilities of the Storwize V7000 Unified system allow a flexible way to create compression rules that are based on file types and the path where files are placed.
- The file system's graphical user interface (GUI) panel enables you to monitor the capacity of file systems, file system pools, and the related MDisk groups in a single view.
- Creation of file systems by using the Storwize V7000 Unified GUI ensures that preferred practices are followed.

Compression in the V7000 Unified storage system is defined by volume. It is not advisable to use all of the degrees of freedom that result from this implementation. This section describes configuration examples that are applicable to different use cases.

Regard the file system pools as the smallest entity where compression is enabled. Therefore, although compression is implemented at a volume level, we refer to *compressed pools* or *compressed file systems*, meaning that all data volumes of a file system or all data volumes of a file system pool are compressed.

16.3 A selectively compressed file system with two pools

This configuration is considered a preferred way to create file systems that use compression. The creation of two pools, along with a placement policy, allows placing only compressible files in a compressed pool. This approach provides high compression savings with best performance. This configuration is shown in Figure 16-1 on page 291.

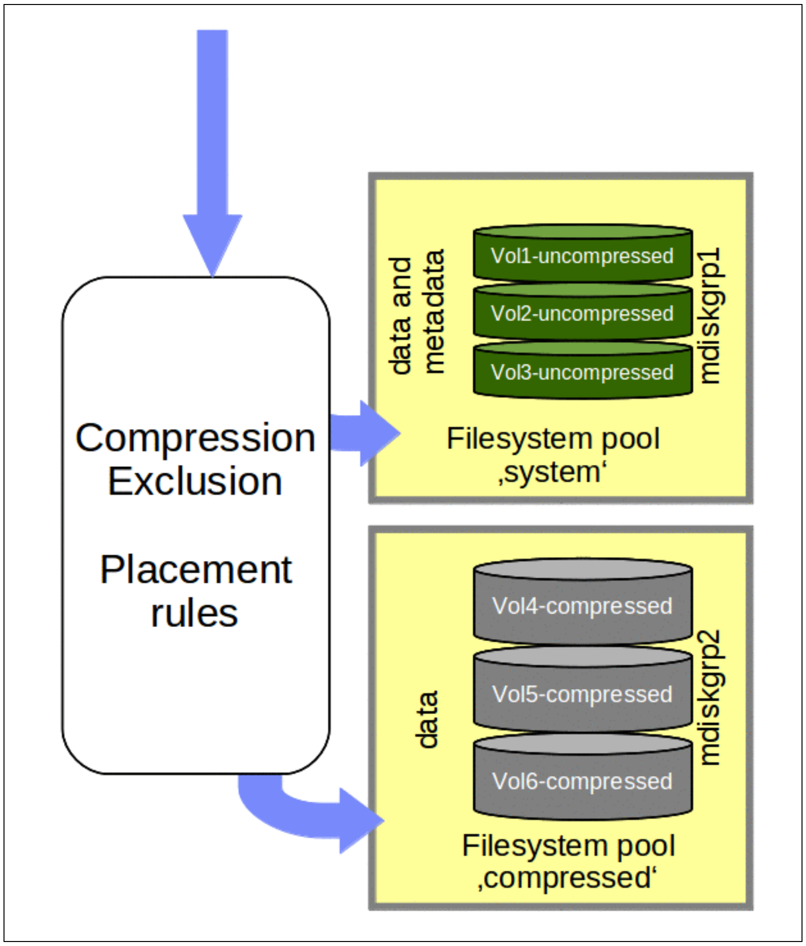


Figure 16-1 Selective compressed configuration with two file system pools

This file system consists of two file system pools:

- The *system* file system pool is based on uncompressed volumes.
- The *compressed* pool consists of compressed volumes only.

A placement policy governs where files are placed when they are created. Files that do not compress well to the *system* file system pool. All other files are placed in the *compressed* file system pool. The placement rules are evaluated only while a file is created. Updating a file does not change the placement. A change in a placement rule does not apply retrospectively. To correct a faulty placement, a migration rule can be run. Examples are included in 16.6, "Managing compressed file systems" on page 323.

Pools: File system storage pools are different from V7000 storage pools.

The Storwize V7000 *storage pools* are synonymous with *MDisk groups*. To avoid confusion, we use the term "MDisk group" in the sections that follow, rather than "storage pool."

The file system storage pools are part of a file system. The initial file system pool of a file system is called *system*. Only the *system* pool contains the metadata of the file system, but it can also contain data. All other file system pools contain data only. When a file volume is created, the use type of the volume can be data only, metadata only, or data and metadata. Compressed volumes should be data-only volumes.

16.3.1 Configuring a selectively compressed file system

This section describes the steps to create a selectively compressed file system by using the Compressed option in the GUI.

1. Open the **Files** folder that is shown in Figure 16-2, and select **File Systems**.

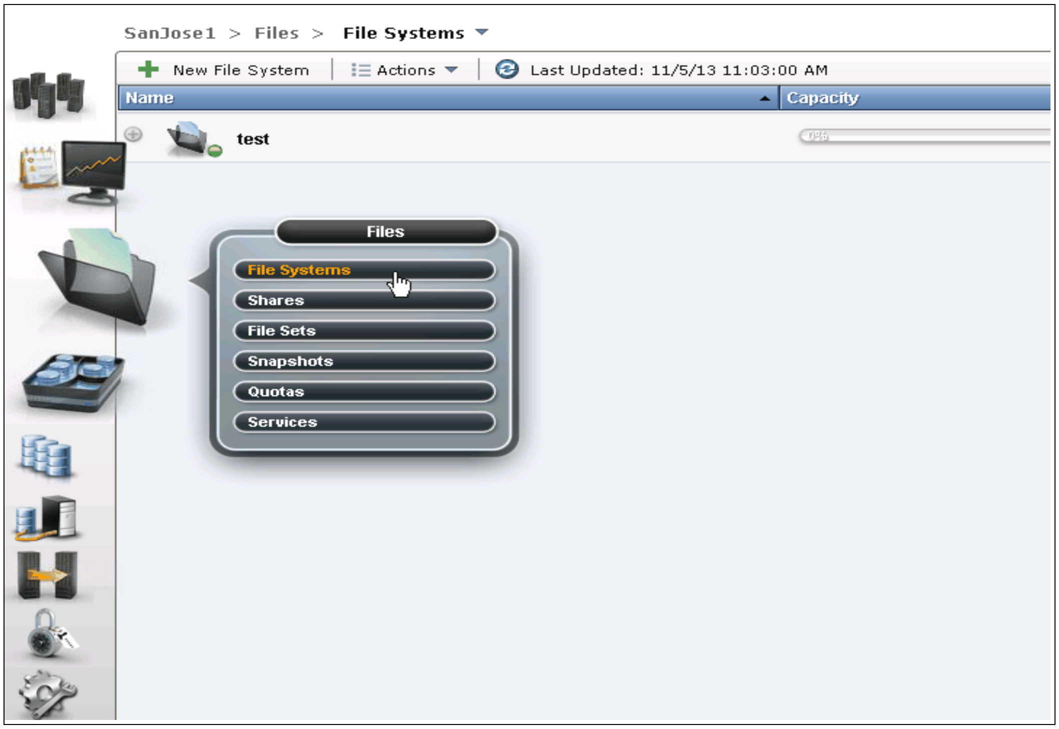
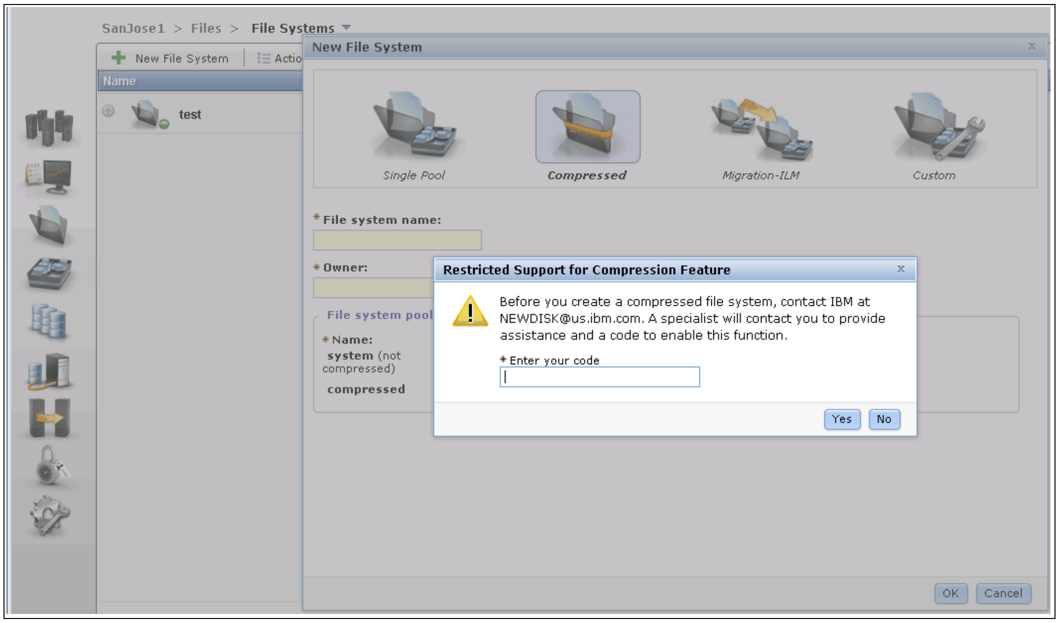


Figure 16-2 Creating a file system

2. Click **New File System** in the top toolbar, and select **Compressed**, as shown in Figure 16-3 on page 293.



*File system name:

*Owner:

File system pool

*Name:

system (not compressed)

compressed

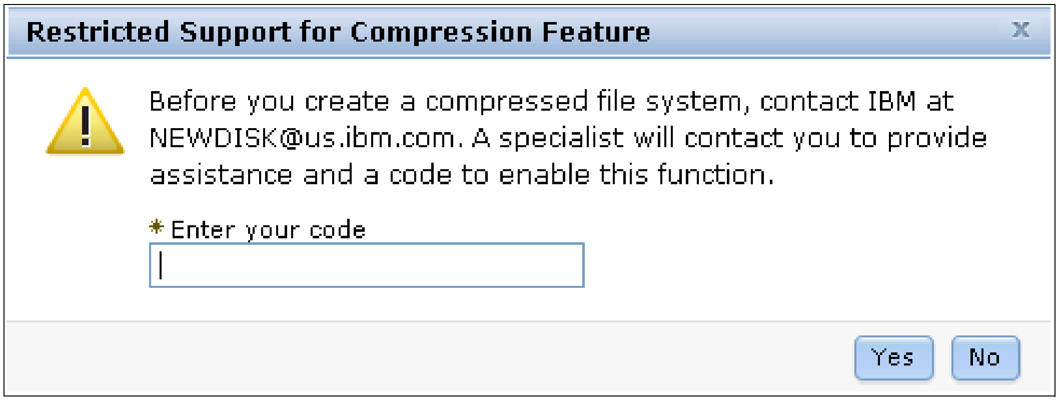
Restricted Support for Compression Feature

Before you create a compressed file system, contact IBM at NEWDISK@us.ibm.com. A specialist will contact you to provide assistance and a code to enable this function.

*Enter your code

Figure 16-3 Compressed file system option

3. The "Restricted Support for Compression Feature" dialog window shown in Figure 16-4 asks for a code to enable compression. This code is provided by NEWDISK@us.ibm.com.



Restricted Support for Compression Feature

Before you create a compressed file system, contact IBM at NEWDISK@us.ibm.com. A specialist will contact you to provide assistance and a code to enable this function.

*Enter your code

Figure 16-4 Compression function code entry

4. Complete the screen information as shown in Figure 16-5 on page 294 with the following information:

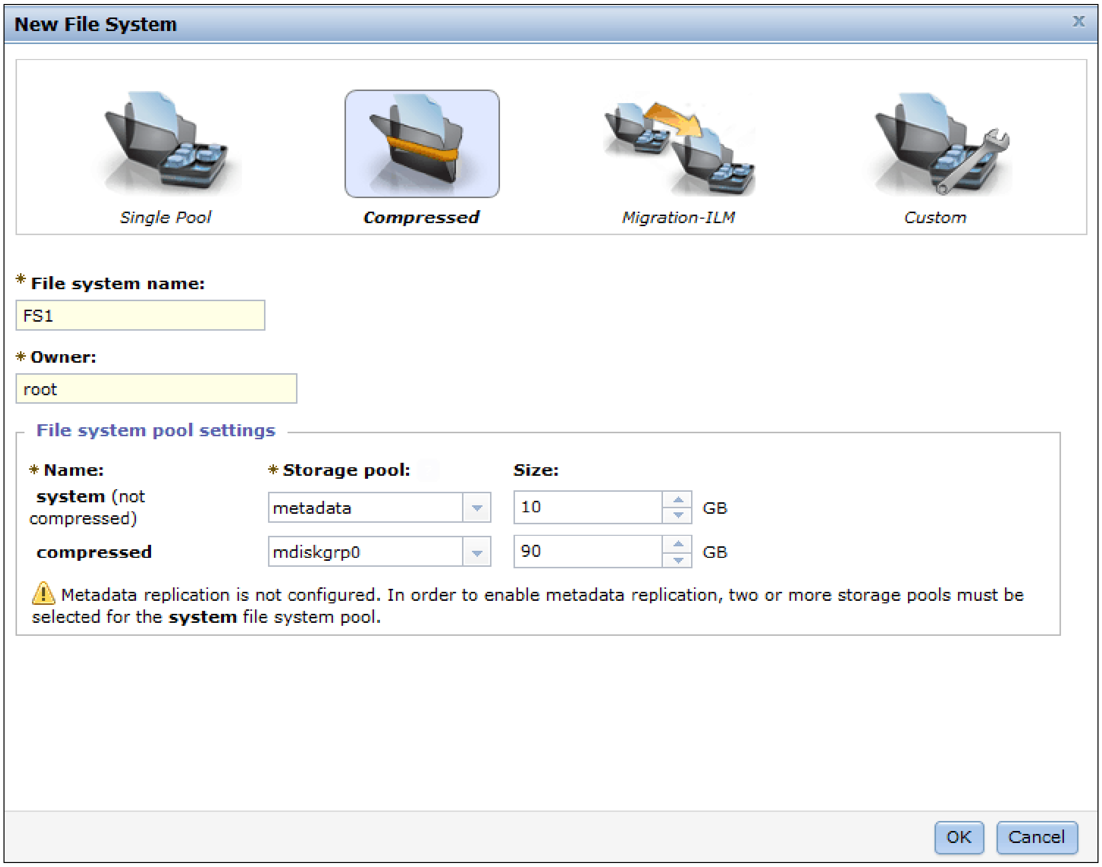
   – File system name: *name* (FS1 in this example)

   – Owner: root

   – Enter names for the storage pools.

     - The *system* storage pool is used for storing the metadata and uncompressed files.
     - The *compressed* pool is used for storing the compressed files.

   – Set the storage pool sizes. There is a minimum size requirement for system plus compressed pools of 100 GB which is enforced



* File system name: FS1

* Owner: root

File system pool settings

* Name:	* Storage pool:	Size:
system (not compressed)	metadata	10 GB
compressed	mdiskgrp0	90 GB

⚠ Metadata replication is not configured. In order to enable metadata replication, two or more storage pools must be selected for the **system** file system pool.

Figure 16-5 Compressed file system option

We suggest that each file system pool with compressed and uncompressed files be in separate MDisk groups. However, this is not enforced, because you can select the same storage pool for both file system pools. Also, for improved performance, it is best to put the system pool that contains metadata in a separate MDisk group. In systems that contain solid-state drives (SSDs), the MDisk groups that contain the SSDs can then be used to provide metadata-only volumes.

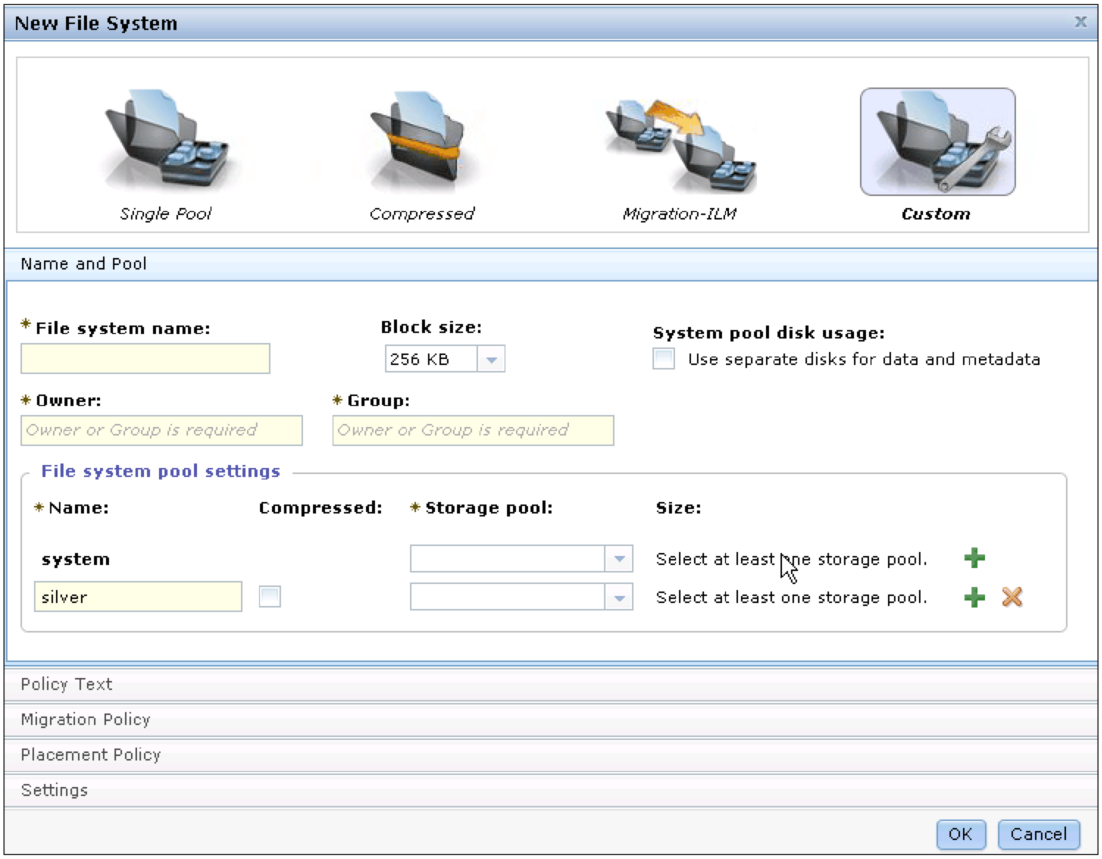
When only one MDisk group is chosen for the *system* file system pool, the metadata replication feature is not configured. If sufficient MDisk groups are available, it is best to define two MDisk groups for the system pool and enable metadata replication.

5. Finish creating the file system. The file system creation process is triggered by clicking **OK** and confirming the summary message. The file system creation dialog window provides the command-line interface (CLI) commands, which are called while the file system is being created.

Custom configuration of a selectively compressed file system

> **Note:** Use the following steps only if you want to use a **Custom** option to create a compressed file system.

The **Custom** preset can also be used to configure a selectively compressed file system, but you must change some of the settings by using the steps that follow (also see Figure 16-6):



* File system name:

Block size: 256 KB

System pool disk usage:
□ Use separate disks for data and metadata

* Owner: Owner or Group is required

* Group: Owner or Group is required

File system pool settings

* Name:	Compressed:	* Storage pool:	Size:
system			Select at least one storage pool.
silver	□		Select at least one storage pool.

*Figure 16-6 Create a file system by using the **Custom** option*

1. A file system pool named “Compressed” is automatically added and enabled for compression. When you are adding a second pool by using the Custom option, the default name is “Silver.” You can change the name, but you must select the **Compressed** option.
2. You must enable the use of separate disks for data and metadata for the system pool used with the Custom preset. Also, specify the block size, which has a default of 256 KB but can be set to 1 MB or 4 MB. The default GPFS file system block size is set to 256 KB with the Compressed preset.
3. A placement policy is included with the Compressed preset, but you need to add one manually with a Custom preset (see the information that follows).

   The placement policy is at the heart of how the Storwize V7000 Unified system implements compression. The default placement policy contains a list of file extensions that are known to be highly compressed by nature. Do not place them in a compressed pool, because the data is already compressed and that does not yield any savings. The Storwize V7000 Unified system uses this method to automatically place files that are compressible into the *compressed* pool.

Files that are not compressible are placed in the *system* pool. This placement occurs when the file is written to disk.

If you are using a Custom option, you must define a placement policy to exclude uncompressible files from the compressed pool. Example 16-1 lists file types that, at the time of writing, are known to not compress well because they contain either compressed or encrypted data. You can copy and paste the list of extensions in this example into the GUI to define exclusion rules for a placement policy.

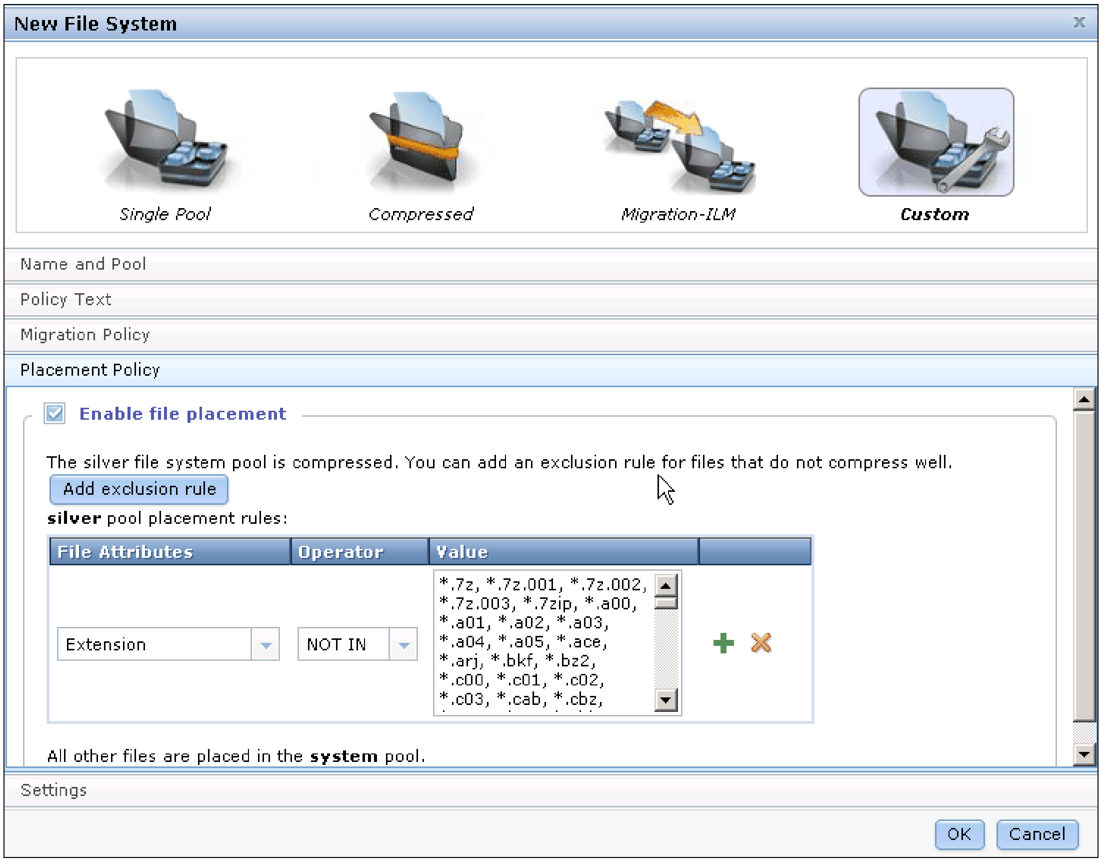
Example 16-1 Extensions of file types that are known to not compress well

```
*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04,
*.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz,
*.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05, *.r06, *.r07,
*.r08, *.r09, *.r10, *.rar, *.sisx, *.sit, *.sitx, *.tar.gz, *.tgz, *.wba,
*.z01, *.z02, *.z03, *.z04, *.z05, *.zip, *.zix, *.aac, *.cda, *.dvf, *.flac,
*.gp5, *.gpx, *.logic, *.m4a, *.m4b, *.m4p, *.mp3, *.mts, *.ogg, *.wma, *.wv,
*.bin, *.img, *.iso, *.docm, *.pps, *.pptx, *.acsm, *.menc, *.emz, *.gif,
*.jpeg, *.jpg, *.png, *.htm, *.swf, *.application, *.exe, *.ipa, *.part1.exe,
*.crw, *.cso, *.mdi, *.odg, *.rpm, *.dcr, *.jad, *.pak, *.rem, *.3g2, *.3gp,
*.asx, *.flv, *.m2t, *.m2ts, *.m4v, *.mkv, *.mov, *.mp4, *.mpg, *.tod, *.ts,
*.vob, *.wmv, *.hqx, *.docx, *.ppt, *.pptm, *.thmx, *.djvu, *.dt2, *.mrw,
*.wbmp, *.abr, *.ai, *.icon, *.ofx, *.pzl, *.tif, *.u3d, *.msi, *.xlsm, *.scr,
*.wav, *.idx, *.abw, *.azw, *.contact, *.dot, *.dotm, *.dotx, *.epub,
*.keynote, *.mobi, *.mswmm, *.odt, *.one, *.otf, *.pages, *.pdf, *.ppsx,
*.prproj, *.pwi, *.onepkg, *.potx, *.tiff, *.!ut, *.atom, *.bc!, *.opml,
*.torrent, *.xhtml, *.jar, *.xlsx, *.fnt, *.sc2replay, *.1st, *.air, *.apk,
*.cbr, *.daa, *.isz, *.m3u8, *.rmvb, *.sxw, *.tga, *.uax, *.crx, *.safariextz,
*.xpi, *.theme, *.themepack, *.3dr, *.dic, *.dlc, *.lng, *.ncs, *.pcd, *.pmd,
*.rss, *.sng, *.svp, *.swp, *.thm, *.uif, *.upg, *.avi, *.fla, *.pcm, *.bbb,
*.bik, *.nba, *.nbu, *.nco, *.wbcat, *.dao, *.dmg, *.tao, *.toast, *.pub,
*.fpx, *.prg, *.cpt, *.eml, *.nvram, *.vmsd, *.vmxf, *.vswp
```

The GUI provides an editor for the placement policy to define the exclusion rules for file types that should not be compressed, as shown in Figure 16-7 on page 297. The placement rule definitions in the GUI editor are *not case-sensitive*. You must choose these options for the respective fields under Placement Policy for the exclusions:

– File Attributes: **Extension**
– Operator: **NOT IN**
– Value: Copy and paste the list from Example 16-1.

The GUI entry field accepts extensions with various delimiters and formats it automatically, as Figure 16-7 on page 297 shows.



☑ **Enable file placement**

The silver file system pool is compressed. You can add an exclusion rule for files that do not compress well.

silver pool placement rules:

File Attributes	Operator	Value	
Extension	NOT IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz,	✚ ✖

All other files are placed in the **system** pool.

Figure 16-7 Defining an exclusion list of files that should not be placed on the compressed pool

4. Adapt the placement policy to the individual configuration.

   The list of file extensions to exclude from compression might need adapting for different reasons:

   – Some file types are not compressible, but they are not listed in the compression exclusion list.

     If you know that your environment contains files that are not compressible, because they are already compressed or they are encrypted, add them to the compression exclusion list. Select **Files** → **File System** → **Edit** to add the file types to the exclusion list. Adding the missing file types prevents the system from trying to compress file types that cannot be compressed further. Compressing files that are already compressed can be detrimental to performance.

   – Some files are compressible, but they are listed in the compression exclusion list.

     Some extensions might be used by different applications with different content. If there are compressible files with extensions that are part of the compression exclusion list, edit the list to remove the extension. Select **Files** → **File System** → **Edit**.

   Saving the adapted placement policy does not move files to other pools. Moving files that were previously misplaced is possible by adding a migration policy.

16.3.2 Compression rules by file set

Placement rules can be applied to an entire file system or to a file set. A *file set* is a subtree of a file system that, in many respects, behaves as a separate file system. File sets give you the ability to partition a file system to allow administrative operations at a finer granularity than the entire file system.

Placement rules for each file set allow different rules for different parts of the file system. Shares that are created in the directory tree below these file sets are subject to those rules.

In the following complex example, you must create placement rules by directly entering placement rule text. To define placement rules, which are specific to a file set, the file sets can be dependent or independent.

Example scenario

A placement rule for the whole file system, such as in the selected compressed configuration, ensures that uncompressible files are excluded from the compressed pool.

The shares in the file set do not use compression at all.

The shares in the `special_application` file set are used by an application that writes only two types of files.

- The `*.cfg` files are control files that are placed in the system pool.
- The `*.pak` files are highly compressible and placed in the compressed pool.

The file extensions used by this application are different from the ones in the built-in file placement policy in the Compressed option. Special rules are necessary to place the `.cfg` files in the compressed pool and `.pak` files in the noncompressed pool.

Steps to create this example scenario:

1. Create a compressed file system with two pools.

   Use the configuration described in 16.3, “A selectively compressed file system with two pools” on page 290 to create a file system and the default placement policy.

2. Create two file sets.

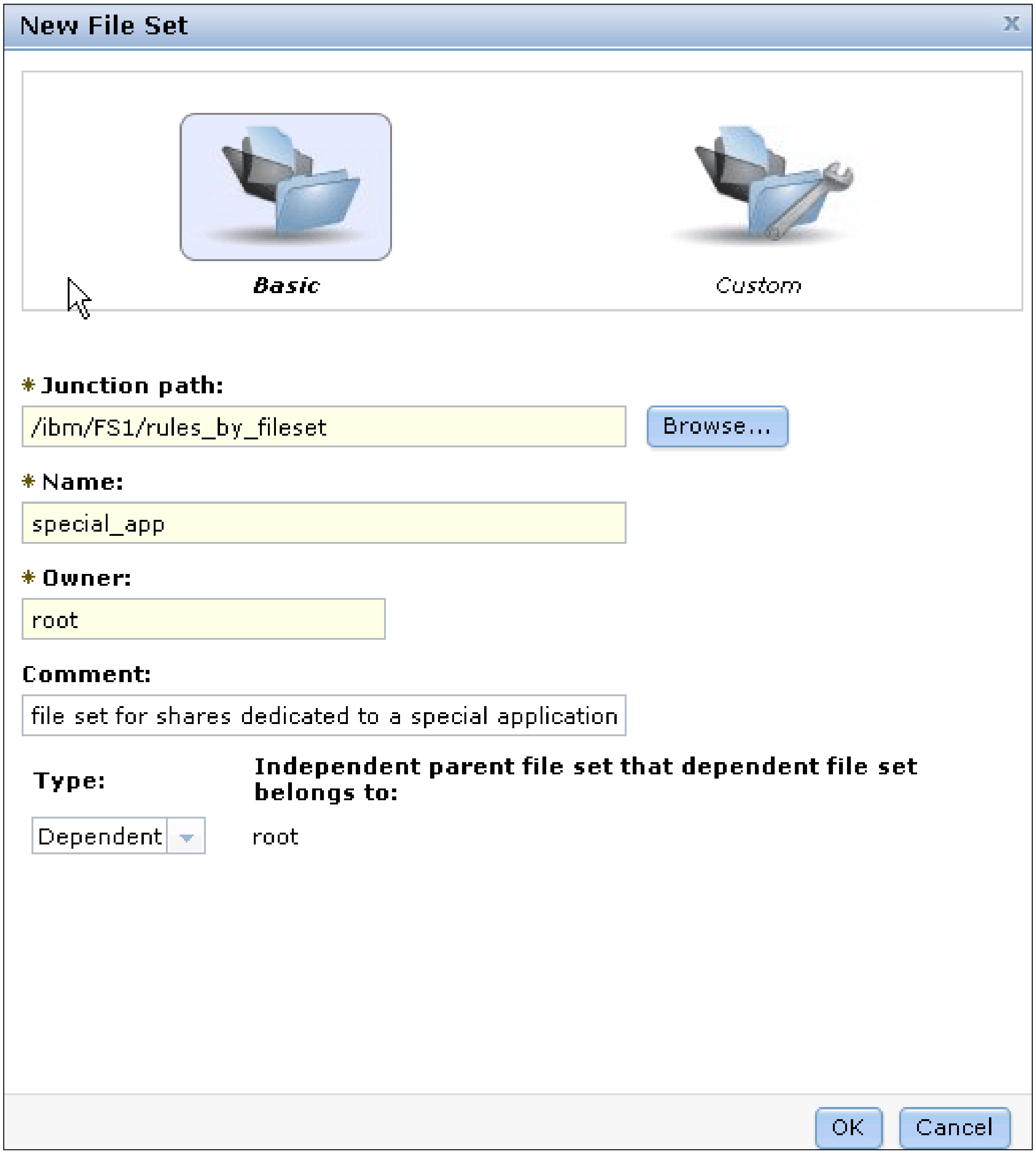
   Select **Files** → **File Sets** and then click **+ New File Set** to open the dialog window that is shown in Figure 16-8 on page 299.

   Use the Junction path **Browse** button to choose the file system mount point.

   In the Name field, provide a subdirectory name, which will be automatically created.

   The Owner field needs to specify a user who owns the path where the file set is linked in the file system tree. For Common Internet File System (CIFS) shares, this can be the domain administrator. The file set name must be unique for each file system.

   Placement rules can reference both dependent and independent file sets. In this example, a dependent file set was chosen in the Type field.



* Junction path: /ibm/FS1/rules_by_fileset

* Name: special_app

* Owner: root

Comment: file set for shares dedicated to a special application

Type: Dependent

Independent parent file set that dependent file set belongs to: root

Figure 16-8   Create file sets in the Files → File Sets → New File Set dialog window

3. Define placement rules.

   Select **Files** → **File System** and then click **Edit** to edit the placement rules.

   The custom rules that you need to add cannot be configured by using the Placement Policy tab. We use the Policy Text editor in the GUI to define the additional rules.

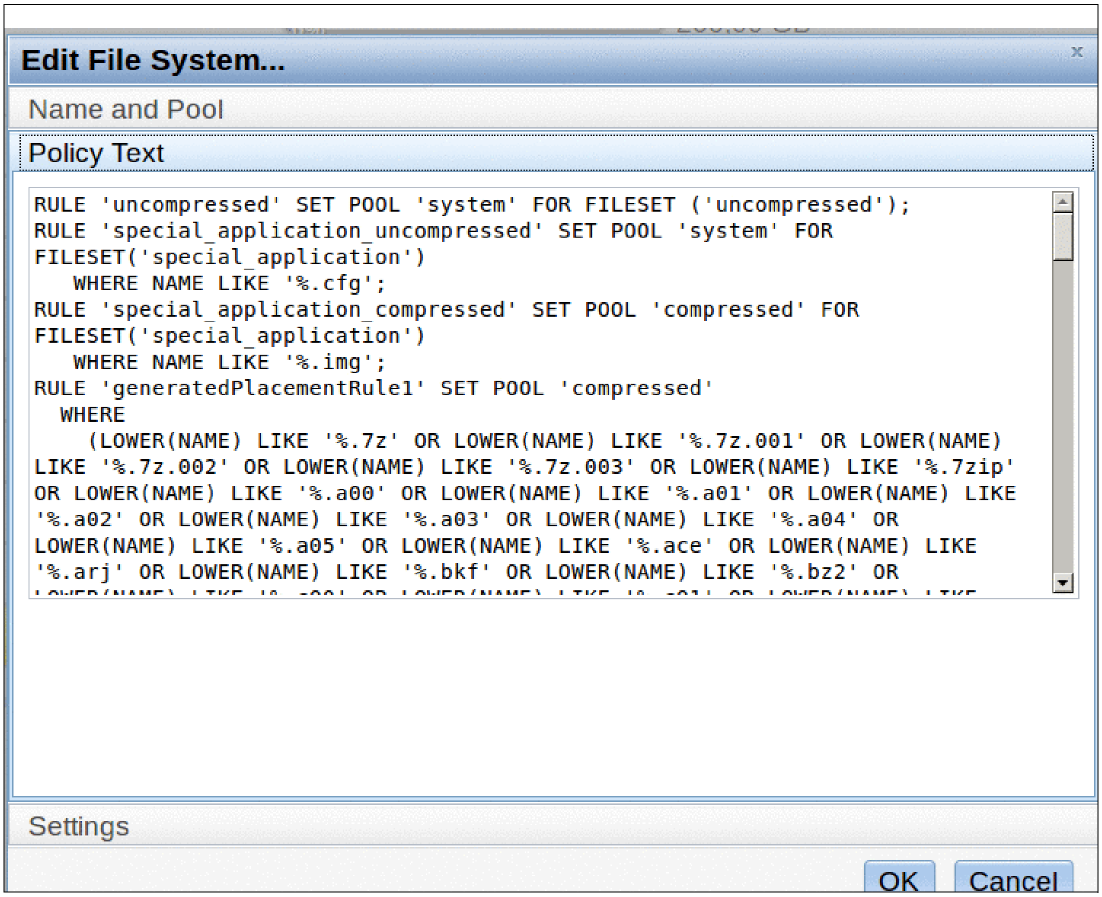
   The three rules must be added *before* the default placement rule text defined in Step 1. If custom rules were created by using the policy text editor in the GUI, the placement policy editor cannot be used.

   In this case, we refer to the system pool as a *target pool* for a special rule, which is not supported in the placement policy editor. However, we created the default placement policy by using the editor first. This is useful because it processes the list of known file type extensions that do not compress well into a correct placement policy string. See Example 16-2 on page 300.

Example 16-2 Additional rules that must be added before the default compression exclusion rule

```
RULE 'uncompressed' SET POOL 'system' FOR FILESET ('uncompressed');
RULE 'special_application_uncompressed' SET POOL 'system'
FOR FILESET('special_application')
   WHERE NAME LIKE '%.cfg';
RULE 'special_application_compressed' SET POOL 'compressed'
FOR FILESET('special_application')
   WHERE NAME LIKE '%.img';
```

Figure 16-9 shows the Edit File System window with the policy text added to define more rules for the newly created file sets.



```
RULE 'uncompressed' SET POOL 'system' FOR FILESET ('uncompressed');
RULE 'special_application_uncompressed' SET POOL 'system' FOR
FILESET('special_application')
   WHERE NAME LIKE '%.cfg';
RULE 'special_application_compressed' SET POOL 'compressed' FOR
FILESET('special_application')
   WHERE NAME LIKE '%.img';
RULE 'generatedPlacementRule1' SET POOL 'compressed'
  WHERE
    (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME)
LIKE '%.7z.002' OR LOWER(NAME) LIKE '%.7z.003' OR LOWER(NAME) LIKE '%.7zip'
OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME) LIKE '%.a01' OR LOWER(NAME) LIKE
'%.a02' OR LOWER(NAME) LIKE '%.a03' OR LOWER(NAME) LIKE '%.a04' OR
LOWER(NAME) LIKE '%.a05' OR LOWER(NAME) LIKE '%.ace' OR LOWER(NAME) LIKE
'%.arj' OR LOWER(NAME) LIKE '%.bkf' OR LOWER(NAME) LIKE '%.bz2' OR
```

Figure 16-9 Edit window with policy text added

The rules are processed in the order that they are defined, so order is important. The rules that we defined for this example work in the following ways:

- All files that are placed in the uncompressed file set are placed in the system pool.
- All files with the `.cfg` extension that are written to the `special_application` file set are written to the system pool.
- All files with the `.img` extension that are written to the `special_application` file set are written to the compressed pool.
- All files that belong to the exclusion list as specified in the generatedPlacementRule1 are written to the system pool. This means that the `.img` files, which are not written to the file set special application, will not be compressed.

Note: The *generatedPlacementRule1* is the policy name generated automatically in the GUI. You can give a policy a unique name when you use the CLI but not when you use the GUI. Example 16-5 on page 309 is a listing of all policies that are in effect.

- All other files are written to the compressed pool because of the `NOT` expression in the generatedPlacementRule1.
- The default placement rule for the system pool is never evaluated. All remaining files are already covered by the generatedPlacementRule1.

Note: A high number of file set-specific placement rules can affect performance. Although it is possible to create thousands of file sets for each file system, the management of separate placement rules for each of them can become problematic. The first placement rule that matches the file that is written is applied, which means that all rules must be traversed to find files that match only the default rule.

If you want to start over with a new set of exclusion rules and need the placement policy editor again, you can apply a default placement policy to the system pool in the Policy Text editor, as in Example 16-3, and the placement policy editor returns.

Example 16-3 Default placement policy

```
RULE 'default' SET POOL 'system';
```

16.3.3 Use case for compressing and decompressing existing data

This procedure goes through the detailed steps to migrate data from or to a compressed file system. The procedures guide you through the process of how to add a compressed file system pool to an existing uncompressed file system and concurrently compress existing data. The reverse procedure is also used to decompress existing data. The later procedure might be needed to disable compression. Compression is enabled when the first NSD or volume is created and disabled only when the last NSD or volume is uncompressed or deleted.

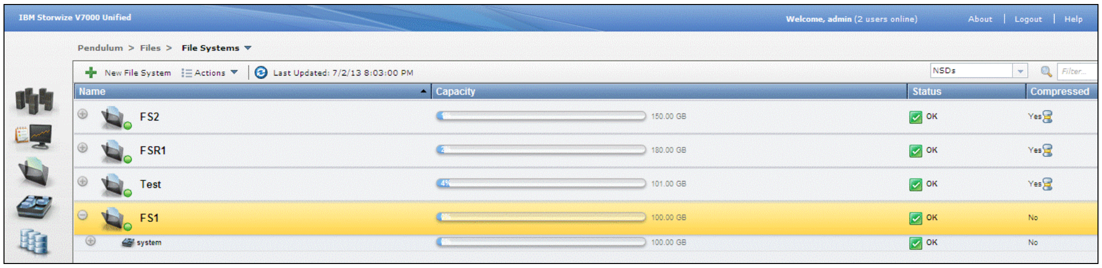
Compressing an existing uncompressed file system

When testing compression and the performance effect on any data set or workload, it is a good idea to create a baseline first with a noncompressed file system. Or, you might already have an existing file system with uncompressed data that needs to be compressed.

When a file system that is not using compression is to be converted to a file system that uses compression, several steps are necessary after analysis of the data in the noncompressed file system.

Adding a compressed pool by using the GUI

1. Figure 16-10 on page 302 shows a file system (FS1) that is uncompressed and has only one file system pool, called “system.” FS1 has a default placement policy rule that places all data into the *file system* system pool.


IBM Storwize V7000 Unified
Welcome, admin (2 users online)
About | Logout | Help
Pendulum > Files > File Systems
New File System
Actions
Last Updated: 7/2/13 8:03:00 PM
NSDs
Filter...

Name	Capacity	Status	Compressed
FS2	150.00 GB	OK	Yes
FSR1	180.00 GB	OK	Yes
Test	101.00 GB	OK	Yes
FS1	100.00 GB	OK	No
system	100.00 GB	OK	No

Figure 16-10 Uncompressed file system

2. Right-click **FS1** and select **Edit File System**, as shown in Figure 16-11.

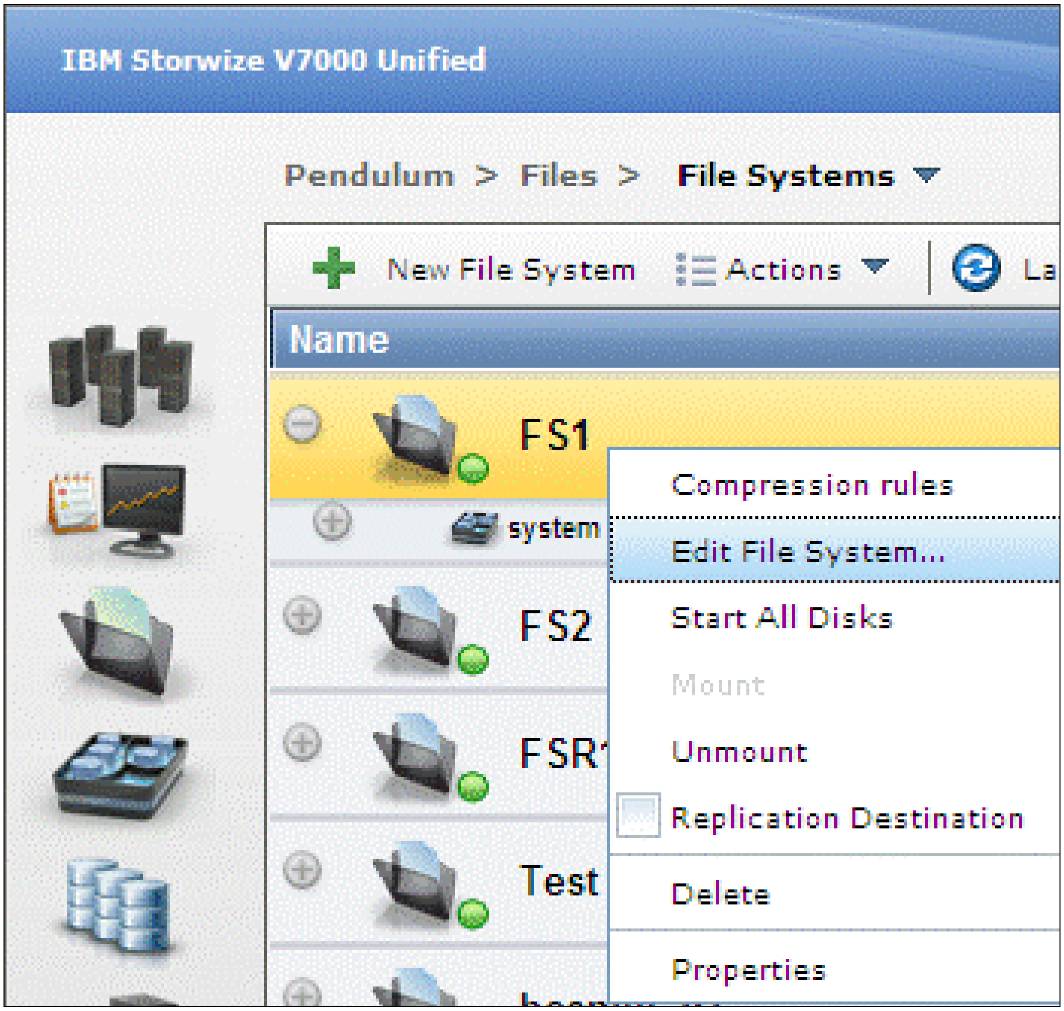
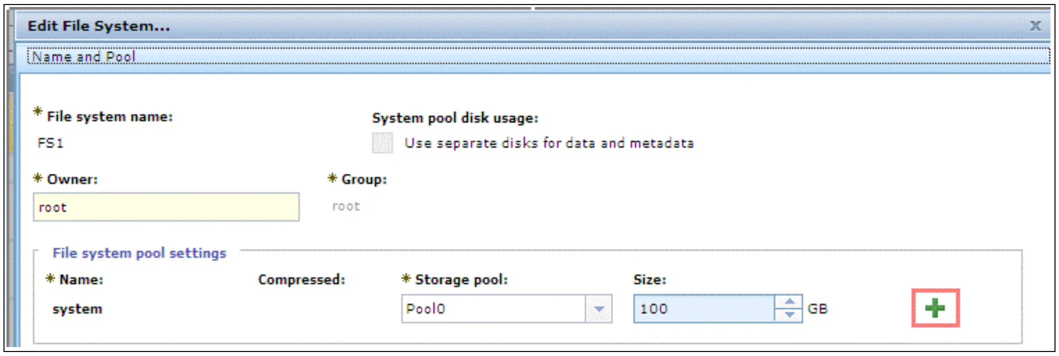


Figure 16-11 Edit file system

3. Select the **+** symbol to add another file system pool (seen Figure 16-12 on page 303).



Edit File System...

Name and Pool

* File system name: FS1

System pool disk usage:

□ Use separate disks for data and metadata

* Owner: root

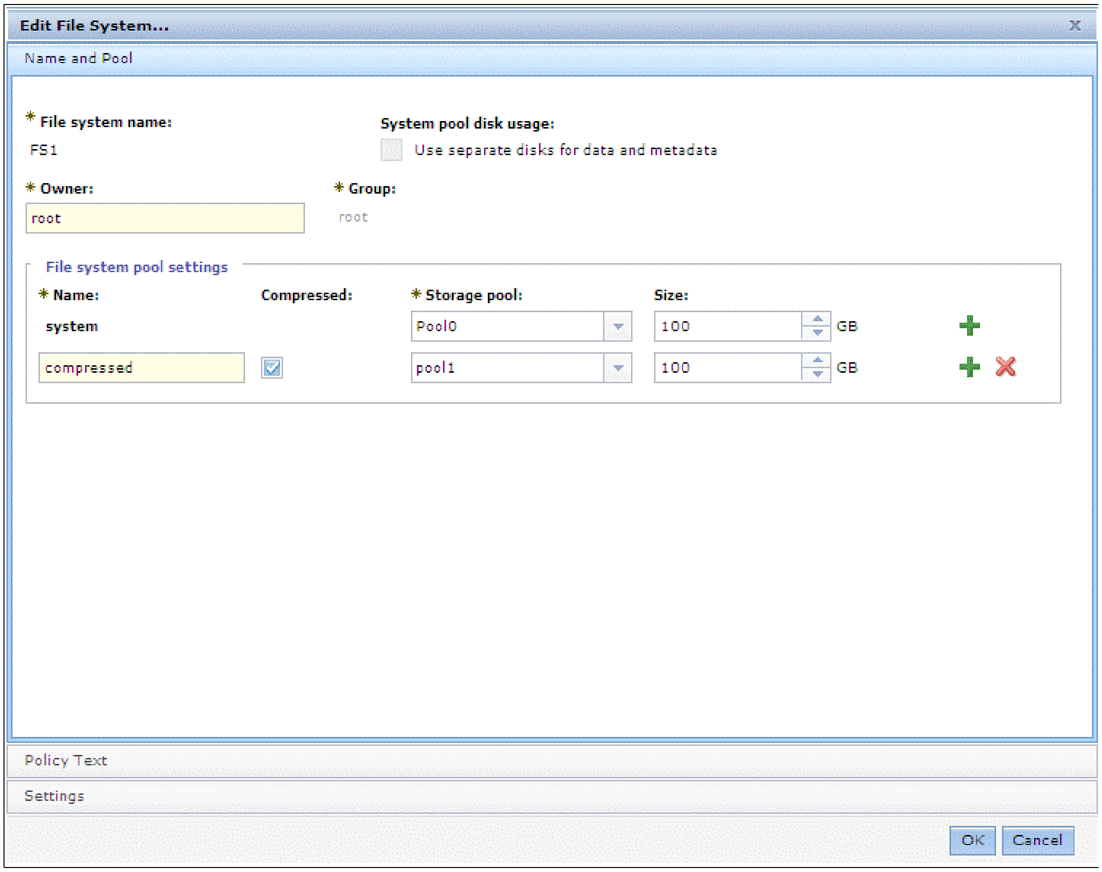
* Group: root

File system pool settings

* Name:	Compressed:	* Storage pool:	Size:
system		Pool0	100 GB

Figure 16-12   Add a file system pool

4. Enter the file system name, such as `compressed`, select a storage pool, specify the size for the pool, and click the **Compressed** check box, as shown in Figure 16-13.



Edit File System...

Name and Pool

* File system name: FS1

System pool disk usage:

□ Use separate disks for data and metadata

* Owner: root

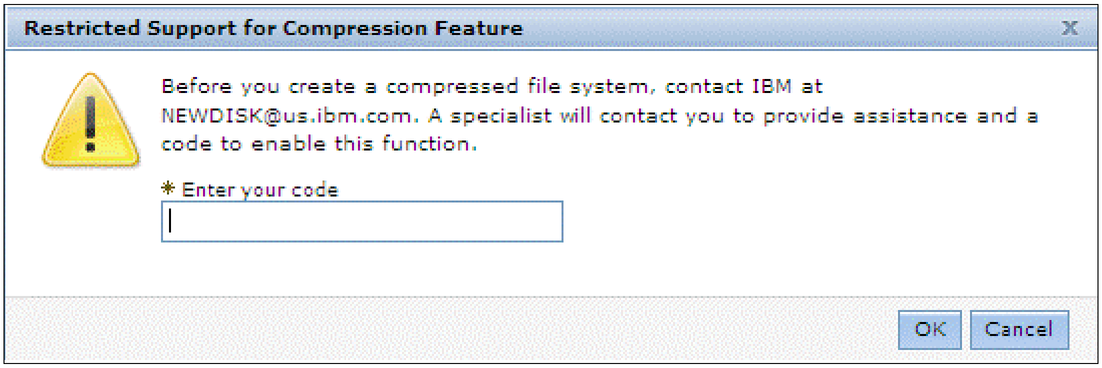
* Group: root

File system pool settings

* Name:	Compressed:	* Storage pool:	Size:
system		Pool0	100 GB
compressed	☑	pool1	100 GB

Figure 16-13   Adding owner and system names

5. When a pop-up window opens, you are asked for a code to enable the compression feature, as shown in Figure 16-14 on page 304. You must get the code from NEWDISK@us.ibm.com. Enter the code, and click **OK** to create the compressed pool.



Restricted Support for Compression Feature

Before you create a compressed file system, contact IBM at NEWDISK@us.ibm.com. A specialist will contact you to provide assistance and a code to enable this function.

* Enter your code

Figure 16-14 Compression feature, code entry field

At this point, all data remains in the system pool. Run a migration policy to move data that is compressible to the compressed pool.

After the migration finishes, disable the migration policy. Then a placement policy needs to be added so that when new files are created, they are directed automatically to the proper pool, depending on whether the files are compressible or not.

The migration policy runs once and moves the compressible data to the compressed pool. The migration can be performed by using either the GUI or the CLI. When you use the GUI, the migration starts as soon as you click **OK**. The migration is concurrent, but it might cause a temporary performance reduction. It that happens, it might be better to use the CLI, because you can control when the migration starts, such as during low I/O activity during the night or on a weekend.

The following substeps explain how to use the GUI or CLI to migrate data and add a placement policy to replace the current default policy.

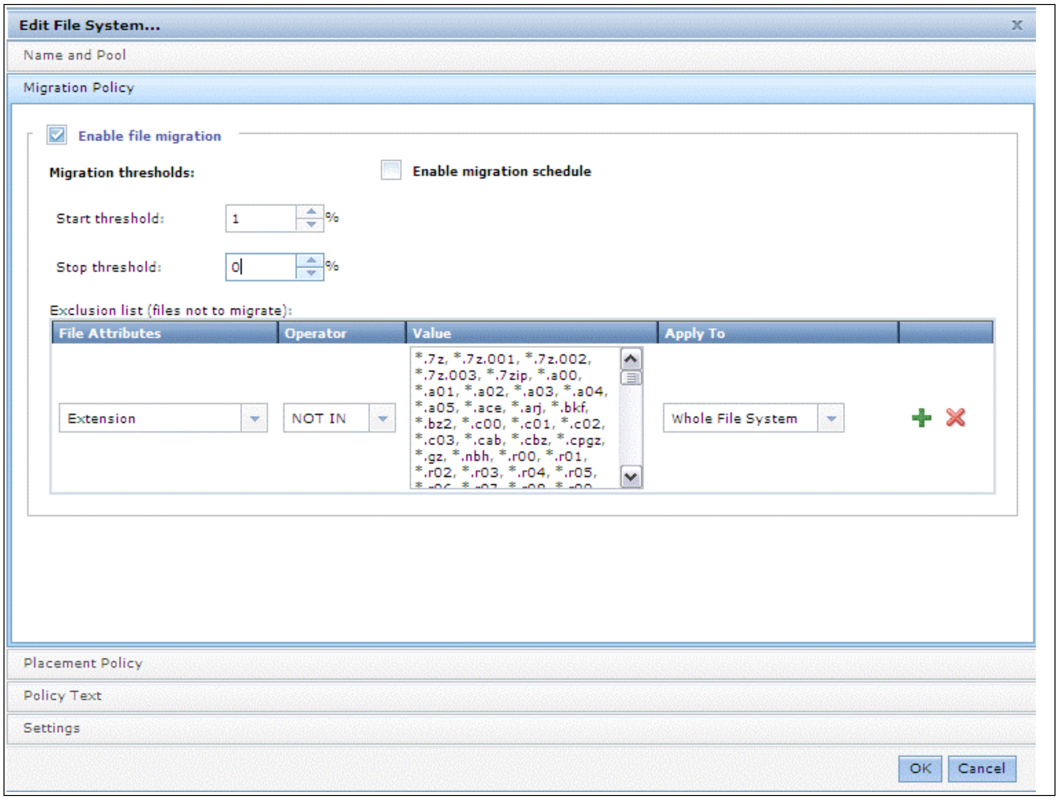
Using the GUI to migrate data from the system pool to a compressed pool

1. Right-click the file system, select **Edit File System**, and select the **Migration Policy** tab, as shown in Figure 16-15 on page 305.
2. Check **Enable file migration**.

   The threshold to start the migration at 1% and stop at 0% indicates that the migration moves the entire analyzed contents from the system pool to the compressed pool. The threshold can be used to control the amount of data that is migrated. This can be important if you want to move only a limited amount of data each time.
3. Under Exclusion list, select these options:

   File Attributes: **Extension**
   Operator: **NOT IN**

   Copy the list of extensions shown in Example 16-1 on page 296, and paste the list in the Value section. Under Apply to, select either **Whole File System** or individual File Sets.



Edit File System...

Name and Pool

Migration Policy

☑ Enable file migration

Migration thresholds: □ Enable migration schedule

Start threshold: 1 %

Stop threshold: 0 %

Exclusion list (files not to migrate):

File Attributes	Operator	Value	Apply To
Extension	NOT IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz, *.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05,	Whole File System

Placement Policy

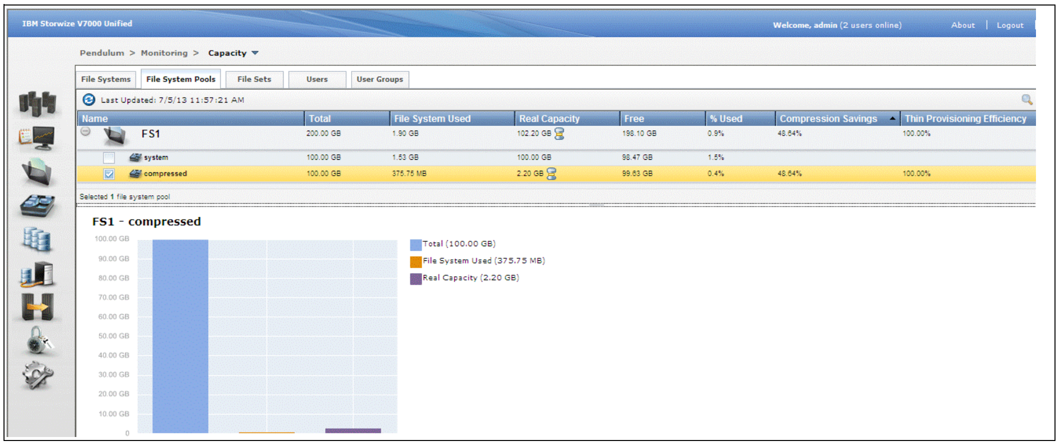
Policy Text

Settings

Figure 16-15 Migration Policy tab

4. When you click **OK**, the migration starts immediately. You can get better control of when to start the migration by using the CLI. Then GUI control resumes after the migration is complete. The policy rule indicates that all files that do not contain the list of extensions are to be moved from the system pool to the file system pool called “compressed.”

5. Using the GUI, select **Monitoring** → **Capacity** → **File System Pools** as shown in Figure 16-16. Verify the compression rate and the amount of data that was migrated between pools.



IBM Storwize V7000 Unified
Welcome, admin (2 users online)
About
Logout
Pendulum > Monitoring > Capacity
File Systems
File System Pools
File Sets
Users
User Groups

Last Updated: 7/5/13 11:57:21 AM

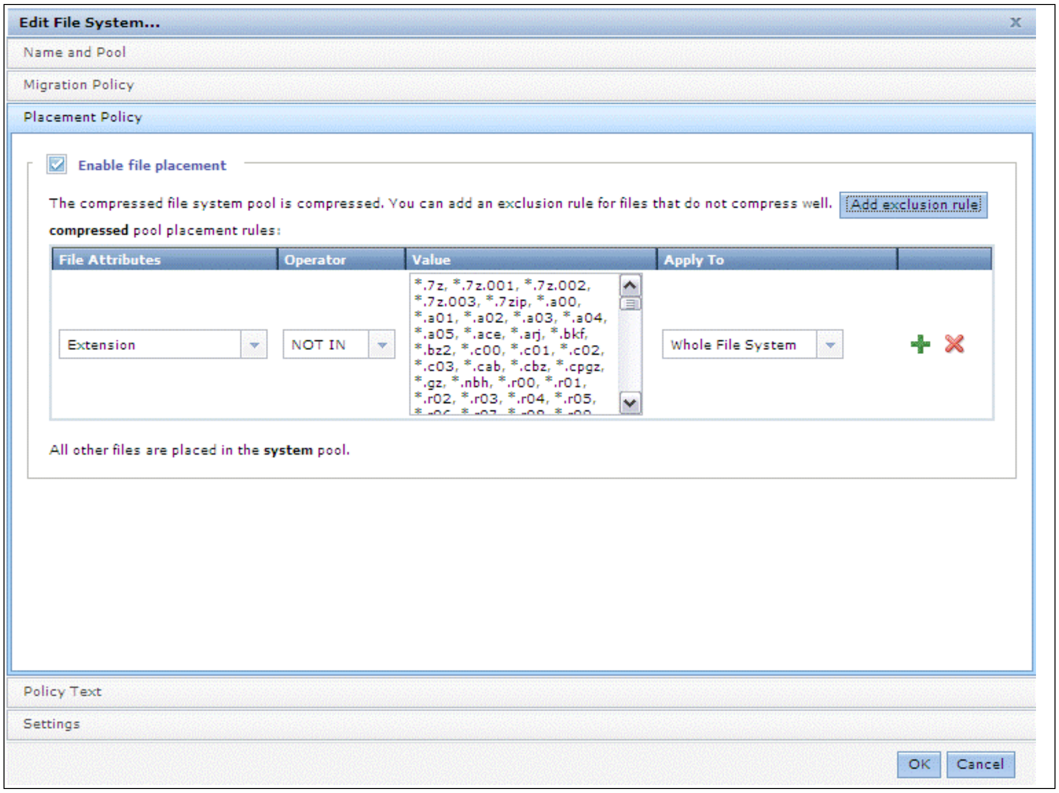
Name	Total	File System Used	Real Capacity	Free	% Used	Compression Savings	Thin Provisioning Efficiency
FS1	200.00 GB	1.90 GB	102.20 GB	198.10 GB	0.9%	48.64%	100.00%
□ system	100.00 GB	1.53 GB	100.00 GB	98.47 GB	1.5%		
☑ compressed	100.00 GB	375.75 MB	2.20 GB	99.63 GB	0.4%	48.64%	100.00%

Selected 1 file system pool

FS1 - compressed

Figure 16-16 Monitoring capacity

6. Immediately after the migration completes, disable the migration by clearing **Enable file migration** on the **Migration Policy** tab. Select the **Placement Policy** tab, as shown in Figure 16-17, and then select **Enable file placement**. Select the **Add exclusion rule** to automatically add the placement rule that places any new files that are created into either the compressed or uncompressed file system pool. Select **OK** to save your choices.



Edit File System...

Name and Pool

Migration Policy

Placement Policy

☑ Enable file placement

The compressed file system pool is compressed. You can add an exclusion rule for files that do not compress well. [Add exclusion rule]

compressed pool placement rules:

File Attributes	Operator	Value	Apply To
Extension	NOT IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz, *.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05,	Whole File System

All other files are placed in the **system** pool.

Policy Text

Settings

Figure 16-17 Placement policy

Using the CLI to migrate data from the system pool to a compressed pool

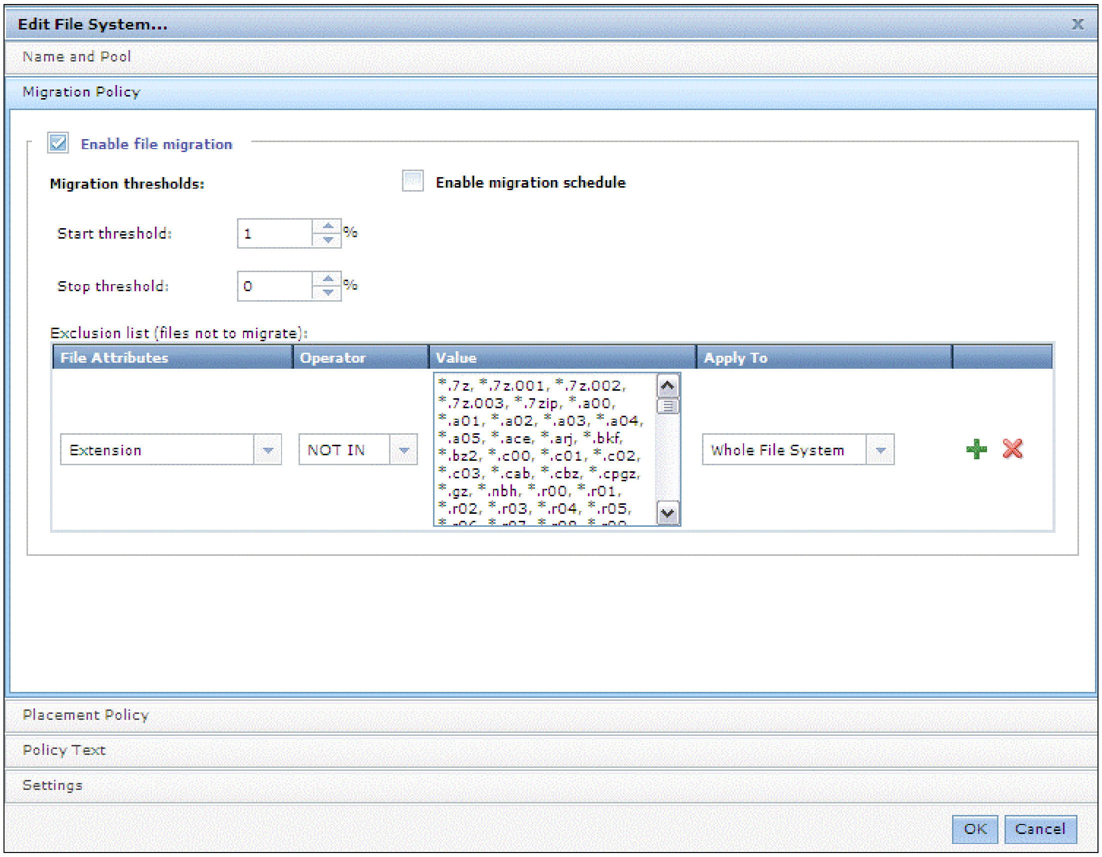
As explained, when using the GUI to migrate data, the migration begins as soon as you click OK. This might not be desirable at all times. You can use the GUI and CLI in combination to accomplish the migration, but the migration starts only when you use the CLI to run the **`runpolicy`** command.

The GUI policy editor can be used to create the **`mkpolicy`** migration policy command, because it nicely formats the long list of extensions to exclude from migration into the policy language statements, as shown in Figure 16-18 on page 307. If you prefer, you can use the CLI to create such a policy, instead. That is cumbersome, but it makes it possible to run the policy by using the **`runpolicy`** command, which executes the migration policy only when the command is issued.

The entire command needs to be in one line, and there are two file extensions (`!ut` and `bc!`) that need to be modified. The `!` is a special UNIX character that needs to be modified to (`\!ut` and `bc\!`). You can use a text editor, such as Notepad, to create the CLI command in combination with the policy editor, as the following example shows.

1. Using the Migration Policy tab shown in Figure 16-18 on page 307, create the policy.

> **Note:** You cancel this after the migration policy is created. *Do not* select OK to run this from the GUI. Instead, use the CLI **`runpolicy`** command.



Edit File System...

Name and Pool

Migration Policy

☑ Enable file migration

Migration thresholds:

□ Enable migration schedule

Start threshold: 1 %

Stop threshold: 0 %

Exclusion list (files not to migrate):

File Attributes	Operator	Value	Apply To	
Extension	NOT IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz, *.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05, *.r06, *.r07, *.r08, *.r09,	Whole File System	

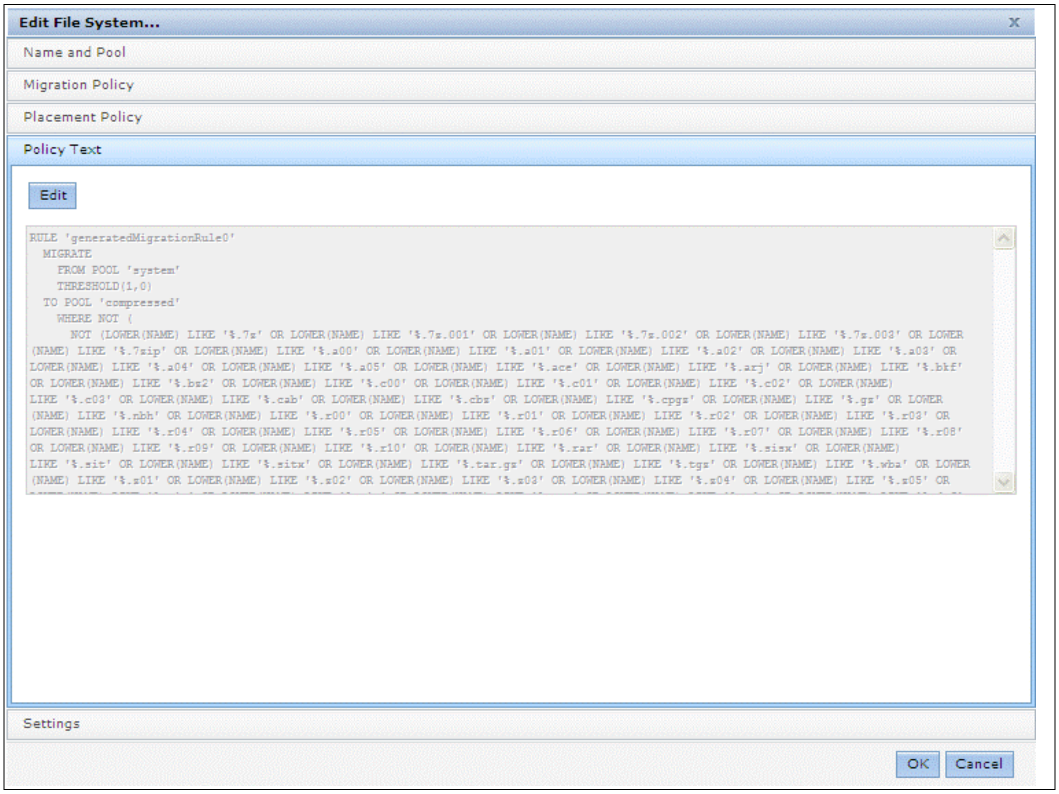
Placement Policy

Policy Text

Settings

Figure 16-18 Migration policy values

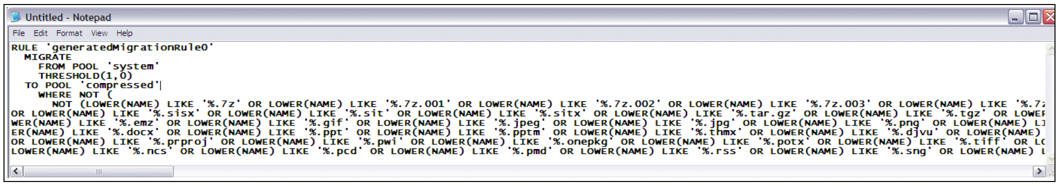
2. Select the **Policy Text** tab, and then select **Edit** to open the Edit File System window shown in Figure 16-19 on page 308. Select and copy the policy text, and use it in an editor such as Notepad to format the CLI command. If you prefer, you can use the text that follows for the `mkpolicy` command instead, and edit it to suit your requirements.



```
RULE 'generatedMigrationRule0'
  MIGRATE
    FROM POOL 'system'
    THRESHOLD(1,0)
  TO POOL 'compressed'
    WHERE NOT (
      NOT (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE '%.7z.002' OR LOWER(NAME) LIKE '%.7z.003' OR LOWER
(NAME) LIKE '%.7zip' OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME) LIKE '%.a01' OR LOWER(NAME) LIKE '%.a02' OR LOWER(NAME) LIKE '%.a03' OR
LOWER(NAME) LIKE '%.a04' OR LOWER(NAME) LIKE '%.a05' OR LOWER(NAME) LIKE '%.ace' OR LOWER(NAME) LIKE '%.arj' OR LOWER(NAME) LIKE '%.bkf'
OR LOWER(NAME) LIKE '%.bz2' OR LOWER(NAME) LIKE '%.c00' OR LOWER(NAME) LIKE '%.c01' OR LOWER(NAME) LIKE '%.c02' OR LOWER(NAME)
LIKE '%.c03' OR LOWER(NAME) LIKE '%.cab' OR LOWER(NAME) LIKE '%.cbz' OR LOWER(NAME) LIKE '%.cpgz' OR LOWER(NAME) LIKE '%.gz' OR LOWER
(NAME) LIKE '%.nbh' OR LOWER(NAME) LIKE '%.r00' OR LOWER(NAME) LIKE '%.r01' OR LOWER(NAME) LIKE '%.r02' OR LOWER(NAME) LIKE '%.r03' OR
LOWER(NAME) LIKE '%.r04' OR LOWER(NAME) LIKE '%.r05' OR LOWER(NAME) LIKE '%.r06' OR LOWER(NAME) LIKE '%.r07' OR LOWER(NAME) LIKE '%.r08'
OR LOWER(NAME) LIKE '%.r09' OR LOWER(NAME) LIKE '%.r10' OR LOWER(NAME) LIKE '%.rar' OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME)
LIKE '%.sit' OR LOWER(NAME) LIKE '%.sitx' OR LOWER(NAME) LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER(NAME) LIKE '%.wba' OR LOWER
(NAME) LIKE '%.z01' OR LOWER(NAME) LIKE '%.z02' OR LOWER(NAME) LIKE '%.z03' OR LOWER(NAME) LIKE '%.z04' OR LOWER(NAME) LIKE '%.z05' OR
```

Figure 16-19 Policy Text tab

3. Make sure that you click **Cancel** to exit the GUI, not OK, before you run the command.
4. Figure 16-20 shows a Notepad sample of building the command. Notice that it must be in a single line. Also notice the proper use of quotation marks.



```
RULE 'generatedMigrationRule0'
  MIGRATE
    FROM POOL 'system'
    THRESHOLD(1,0)
  TO POOL 'compressed'
    WHERE NOT (
      NOT (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE '%.7z.002' OR LOWER(NAME) LIKE '%.7z.003' OR LOWER(NAME) LIKE '%.7z
OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME) LIKE '%.sit' OR LOWER(NAME) LIKE '%.sitx' OR LOWER(NAME) LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER
WER(NAME) LIKE '%.emz' OR LOWER(NAME) LIKE '%.gif' OR LOWER(NAME) LIKE '%.jpeg' OR LOWER(NAME) LIKE '%.jpg' OR LOWER(NAME) LIKE '%.png' OR LOWER(NAME) LI
ER(NAME) LIKE '%.docx' OR LOWER(NAME) LIKE '%.ppt' OR LOWER(NAME) LIKE '%.pptm' OR LOWER(NAME) LIKE '%.thmx' OR LOWER(NAME) LIKE '%.djvu' OR LOWER(NAME)
OR LOWER(NAME) LIKE '%.prproj' OR LOWER(NAME) LIKE '%.pwi' OR LOWER(NAME) LIKE '%.onepkg' OR LOWER(NAME) LIKE '%.potx' OR LOWER(NAME) LIKE '%.tiff' OR LO
LOWER(NAME) LIKE '%.ncs' OR LOWER(NAME) LIKE '%.pcd' OR LOWER(NAME) LIKE '%.pmd' OR LOWER(NAME) LIKE '%.rss' OR LOWER(NAME) LIKE '%.sng' OR LOWER(NAME) L
```

Figure 16-20 The CLI command in Notepad

5. Example 16-4 shows execution of the **`mkpolicy`** CLI command. The name of the policy that is being created is `migratetocompress`. This policy migrates all of the files that do not have one of the listed do-not-migrate extensions from the system pool to the compress pool. Because the threshold for the migration was set low, the system starts the migration job automatically. You can modify the Start and Stop % fields for thresholds to limit the amount of data that is migrated. That can be important if there is a large amount of data to be transferred within a certain time limit. The **`mkpolicy`** command does not initiate a migration. The migration starts only when the **`runpolicy`** command is used.

Example 16-4 CLI mkpolicy command

```
[7803088.ibm]$
[7803088.ibm]$ mkpolicy migratetocompress -R "RULE 'migratetocompress' migrate from pool 'system' threshold(1,0)
to pool 'compressed' where NOT (NOT (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE
'%.7z.002' OR LOWER(NAME) LIKE '%.7z.003' OR LOWER(NAME) LIKE '%.7zip' OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME)
LIKE '%.a01' OR LOWER(NAME) LIKE '%.a02' OR LOWER(NAME) LIKE '%.a03' OR LOWER(NAME) LIKE '%.a04' OR LOWER(NAME)
LIKE '%.a05' OR LOWER(NAME) LIKE '%.ace' OR LOWER(NAME) LIKE '%.arj' OR LOWER(NAME) LIKE '%.bkf' OR LOWER(NAME)
LIKE '%.bz2' OR LOWER(NAME) LIKE '%.c00' OR LOWER(NAME) LIKE '%.c01' OR LOWER(NAME) LIKE '%.c02' OR LOWER(NAME)
LIKE '%.c03' OR LOWER(NAME) LIKE '%.cab' OR LOWER(NAME) LIKE '%.cbz' OR LOWER(NAME) LIKE '%.cpgz' OR LOWER(NAME)
LIKE '%.gz' OR LOWER(NAME) LIKE '%.nbh' OR LOWER(NAME) LIKE '%.r00' OR LOWER(NAME) LIKE '%.r01' OR LOWER(NAME)
LIKE '%.r02' OR LOWER(NAME) LIKE '%.r03' OR LOWER(NAME) LIKE '%.r04' OR LOWER(NAME) LIKE '%.r05' OR LOWER(NAME)
LIKE '%.r06' OR LOWER(NAME) LIKE '%.r07' OR LOWER(NAME) LIKE '%.r08' OR LOWER(NAME) LIKE '%.r09' OR LOWER(NAME)
LIKE '%.r10' OR LOWER(NAME) LIKE '%.rar' OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME) LIKE '%.sit' OR LOWER(NAME)
LIKE '%.sitx' OR LOWER(NAME) LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER(NAME) LIKE '%.wba' OR
LOWER(NAME) LIKE '%.z01' OR LOWER(NAME) LIKE '%.z02' OR LOWER(NAME) LIKE '%.z03' OR LOWER(NAME) LIKE '%.z04' OR
LOWER(NAME) LIKE '%.z05' OR LOWER(NAME) LIKE '%.zip' OR LOWER(NAME) LIKE '%.zix' OR LOWER(NAME) LIKE '%.aac' OR
LOWER(NAME) LIKE '%.cda' OR LOWER(NAME) LIKE '%.dvf' OR LOWER(NAME) LIKE '%.flac' OR LOWER(NAME) LIKE '%.gp5' OR
LOWER(NAME) LIKE '%.gpx' OR LOWER(NAME) LIKE '%.logic' OR LOWER(NAME) LIKE '%.m4a' OR LOWER(NAME) LIKE '%.m4b' OR
LOWER(NAME) LIKE '%.m4p' OR LOWER(NAME) LIKE '%.mp3' OR LOWER(NAME) LIKE '%.mts' OR LOWER(NAME) LIKE '%.ogg' OR
LOWER(NAME) LIKE '%.wma' OR LOWER(NAME) LIKE '%.wv' OR LOWER(NAME) LIKE '%.bin' OR LOWER(NAME) LIKE '%.img' OR
LOWER(NAME) LIKE '%.iso' OR LOWER(NAME) LIKE '%.docm' OR LOWER(NAME) LIKE '%.pps' OR LOWER(NAME) LIKE '%.pptx' OR
LOWER(NAME) LIKE '%.acsm' OR LOWER(NAME) LIKE '%.menc' OR LOWER(NAME) LIKE '%.emz' OR LOWER(NAME) LIKE '%.gif' OR
LOWER(NAME) LIKE '%.jpeg' OR LOWER(NAME) LIKE '%.jpg' OR LOWER(NAME) LIKE '%.png' OR LOWER(NAME) LIKE '%.htm' OR
LOWER(NAME) LIKE '%.swf' OR LOWER(NAME) LIKE '%.application' OR LOWER(NAME) LIKE '%.exe' OR LOWER(NAME) LIKE
'%.ipa' OR LOWER(NAME) LIKE '%.part1.exe' OR LOWER(NAME) LIKE '%.crw' OR LOWER(NAME) LIKE '%.cso' OR LOWER(NAME)
LIKE '%.mdi' OR LOWER(NAME) LIKE '%.odg' OR LOWER(NAME) LIKE '%.rpm' OR LOWER(NAME) LIKE '%.dcr' OR LOWER(NAME)
LIKE '%.jad' OR LOWER(NAME) LIKE '%.pak' OR LOWER(NAME) LIKE '%.rem' OR LOWER(NAME) LIKE '%.3g2' OR LOWER(NAME)
LIKE '%.3gp' OR LOWER(NAME) LIKE '%.asx' OR LOWER(NAME) LIKE '%.flv' OR LOWER(NAME) LIKE '%.m2t' OR LOWER(NAME)
LIKE '%.m2ts' OR LOWER(NAME) LIKE '%.m4v' OR LOWER(NAME) LIKE '%.mkv' OR LOWER(NAME) LIKE '%.mov' OR LOWER(NAME)
LIKE '%.mp4' OR LOWER(NAME) LIKE '%.mpg' OR LOWER(NAME) LIKE '%.tod' OR LOWER(NAME) LIKE '%.ts' OR LOWER(NAME)
LIKE '%.vob' OR LOWER(NAME) LIKE '%.wmv' OR LOWER(NAME) LIKE '%.hqx' OR LOWER(NAME) LIKE '%.docx' OR LOWER(NAME)
LIKE '%.ppt' OR LOWER(NAME) LIKE '%.pptm' OR LOWER(NAME) LIKE '%.thmx' OR LOWER(NAME) LIKE '%.djvu' OR
LOWER(NAME) LIKE '%.dt2' OR LOWER(NAME) LIKE '%.mrw' OR LOWER(NAME) LIKE '%.wbmp' OR LOWER(NAME) LIKE '%.abr' OR
LOWER(NAME) LIKE '%.ai' OR LOWER(NAME) LIKE '%.icon' OR LOWER(NAME) LIKE '%.ofx' OR LOWER(NAME) LIKE '%.pzl' OR
LOWER(NAME) LIKE '%.tif' OR LOWER(NAME) LIKE '%.u3d' OR LOWER(NAME) LIKE '%.msi' OR LOWER(NAME) LIKE '%.xlsm' OR
LOWER(NAME) LIKE '%.scr' OR LOWER(NAME) LIKE '%.wav' OR LOWER(NAME) LIKE '%.idx' OR LOWER(NAME) LIKE '%.abw' OR
LOWER(NAME) LIKE '%.azw' OR LOWER(NAME) LIKE '%.contact' OR LOWER(NAME) LIKE '%.dot' OR LOWER(NAME) LIKE '%.dotm'
OR LOWER(NAME) LIKE '%.dotx' OR LOWER(NAME) LIKE '%.epub' OR LOWER(NAME) LIKE '%.keynote' OR LOWER(NAME) LIKE
'%.mobi' OR LOWER(NAME) LIKE '%.mswmm' OR LOWER(NAME) LIKE '%.odt' OR LOWER(NAME) LIKE '%.one' OR LOWER(NAME)
LIKE '%.otf' OR LOWER(NAME) LIKE '%.pages' OR LOWER(NAME) LIKE '%.pdf' OR LOWER(NAME) LIKE '%.ppsx' OR
LOWER(NAME) LIKE '%.prproj' OR LOWER(NAME) LIKE '%.pwi' OR LOWER(NAME) LIKE '%.onepkg' OR LOWER(NAME) LIKE
'%.potx' OR LOWER(NAME) LIKE '%.tiff' OR LOWER(NAME) LIKE '%.\!ut' OR LOWER(NAME) LIKE '%.atom' OR LOWER(NAME)
LIKE '%.bc\!' OR LOWER(NAME) LIKE '%.opml' OR LOWER(NAME) LIKE '%.torrent' OR LOWER(NAME) LIKE '%.xhtml' OR
LOWER(NAME) LIKE '%.jar' OR LOWER(NAME) LIKE '%.xlsx' OR LOWER(NAME) LIKE '%.fnt' OR LOWER(NAME) LIKE
'%.sc2replay' OR LOWER(NAME) LIKE '%.1st' OR LOWER(NAME) LIKE '%.air' OR LOWER(NAME) LIKE '%.apk' OR LOWER(NAME)
LIKE '%.cbr' OR LOWER(NAME) LIKE '%.daa' OR LOWER(NAME) LIKE '%.isz' OR LOWER(NAME) LIKE '%.m3u8' OR LOWER(NAME)
LIKE '%.rmvb' OR LOWER(NAME) LIKE '%.sxw' OR LOWER(NAME) LIKE '%.tga' OR LOWER(NAME) LIKE '%.uax' OR LOWER(NAME)
LIKE '%.crx' OR LOWER(NAME) LIKE '%.safariextz' OR LOWER(NAME) LIKE '%.xpi' OR LOWER(NAME) LIKE '%.theme' OR
LOWER(NAME) LIKE '%.themepack' OR LOWER(NAME) LIKE '%.3dr' OR LOWER(NAME) LIKE '%.dic' OR LOWER(NAME) LIKE
'%.dlc' OR LOWER(NAME) LIKE '%.lng' OR LOWER(NAME) LIKE '%.ncs' OR LOWER(NAME) LIKE '%.pcd' OR LOWER(NAME) LIKE
'%.pmd' OR LOWER(NAME) LIKE '%.rss' OR LOWER(NAME) LIKE '%.sng' OR LOWER(NAME) LIKE '%.svp' OR LOWER(NAME) LIKE
'%.swp' OR LOWER(NAME) LIKE '%.thm' OR LOWER(NAME) LIKE '%.uif' OR LOWER(NAME) LIKE '%.upg' OR LOWER(NAME) LIKE
'%.avi' OR LOWER(NAME) LIKE '%.fla' OR LOWER(NAME) LIKE '%.pcm' OR LOWER(NAME) LIKE '%.bbb' OR LOWER(NAME) LIKE
'%.bik' OR LOWER(NAME) LIKE '%.nba' OR LOWER(NAME) LIKE '%.nbu' OR LOWER(NAME) LIKE '%.nco' OR LOWER(NAME) LIKE
'%.wbcat' OR LOWER(NAME) LIKE '%.dao' OR LOWER(NAME) LIKE '%.dmg' OR LOWER(NAME) LIKE '%.tao' OR LOWER(NAME) LIKE
'%.toast' OR LOWER(NAME) LIKE '%.pub' OR LOWER(NAME) LIKE '%.fpx' OR LOWER(NAME) LIKE '%.prg' OR LOWER(NAME) LIKE
'%.cpt' OR LOWER(NAME) LIKE '%.eml' OR LOWER(NAME) LIKE '%.nvram' OR LOWER(NAME) LIKE '%.vmsd' OR LOWER(NAME)
LIKE '%.vmxf' OR LOWER(NAME) LIKE '%.vswp'));"
EFSSG1000I The command completed successfully.
[7803088.ibm]$
```

6. The **lspolicy** command can be used to verify that the policy exists and what the policy contains. The **lspolicy** command displays all existing policies, as shown in Example 16-5.

Example 16-5 lspolicy command

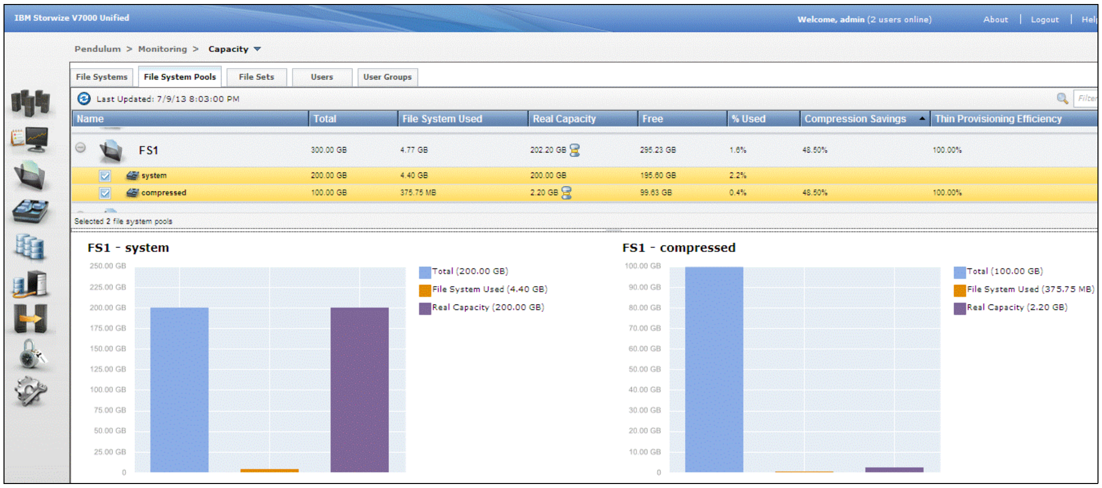
```
[7803088.ibm]$ lspolicy
Policy Name                                     Declarations (define/RULE)
default                                         default
FS1_generatedPolicy_2013_07_05_19_32_40generatedPlacementRule1
FS1_generatedPolicy_2013_07_08_19_44_51default
migratetocompress                               migratetocompress
migratetosystem                                 migratetosystem
EFSSG1000I The command completed successfully.
[7803088.ibm]$
```

7. The **`lspolicy -P`** ***`policy name`*** command displays the contents of the entire policy, as shown in Example 16-6.

Example 16-6 lspolicy -P command

```
[7803088.ibm]$ lspolicy -P migratetocompress
Policy Name       Declaration Name  Default Declarations
migratetocompress migratetocompress N       RULE 'migratetocompress' MIGRATE FROM POOL 'system' THRESHOLD(1,0) TO
POOL 'compressed' WHERE NOT  (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE
'%.7z.002' OR LOWER(NAME) LIKE '%.7z.003' OR LOWER(NAME) LIKE '%.7zip' OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME)
LIKE '%.a01' OR LOWER(NAME) LIKE '%.a02' OR LOWER(NAME) LIKE '%.a03' OR LOWER(NAME) LIKE '%.a04' OR LOWER(NAME)
LIKE '%.a05' OR LOWER(NAME) LIKE '%.ace' OR LOWER(NAME) LIKE '%.arj' OR LOWER(NAME) LIKE '%.bkf' OR LOWER(NAME)
LIKE '%.bz2' OR LOWER(NAME) LIKE '%.c00' OR LOWER(NAME) LIKE '%.c01' OR LOWER(NAME) LIKE '%.c02' OR LOWER(NAME)
LIKE '%.c03' OR LOWER(NAME) LIKE '%.cab' OR LOWER(NAME) LIKE '%.cbz' OR LOWER(NAME) LIKE '%.cpgz' OR LOWER(NAME)
LIKE '%.gz' OR LOWER(NAME) LIKE '%.nbh' OR LOWER(NAME) LIKE '%.r00' OR LOWER(NAME) LIKE '%.r01' OR LOWER(NAME)
LIKE '%.r02' OR LOWER(NAME) LIKE '%.r03' OR LOWER(NAME) LIKE '%.r04' OR LOWER(NAME) LIKE '%.r05' OR LOWER(NAME)
LIKE '%.r06' OR LOWER(NAME) LIKE '%.r07' OR LOWER(NAME) LIKE '%.r08' OR LOWER(NAME) LIKE '%.r09' OR LOWER(NAME)
LIKE '%.r10' OR LOWER(NAME) LIKE '%.rar' OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME) LIKE '%.sit' OR LOWER(NAME)
LIKE '%.sitx' OR LOWER(NAME) LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER(NAME) LIKE '%.wba' OR
LOWER(NAME) LIKE '%.z01' OR LOWER(NAME) LIKE '%.z02' OR LOWER(NAME) LIKE '%.z03' OR LOWER(NAME) LIKE '%.z04' OR
LOWER(NAME) LIKE '%.z05' OR LOWER(NAME) LIKE '%.zip' OR LOWER(NAME) LIKE '%.zix' OR LOWER(NAME) LIKE '%.aac' OR
LOWER(NAME) LIKE '%.cda' OR LOWER(NAME) LIKE '%.dvf' OR LOWER(NAME) LIKE '%.flac' OR LOWER(NAME) LIKE '%.gp5' OR
LOWER(NAME) LIKE '%.gpx' OR LOWER(NAME) LIKE '%.logic' OR LOWER(NAME) LIKE '%.m4a' OR LOWER(NAME) LIKE '%.m4b' OR
LOWER(NAME) LIKE '%.m4p' OR LOWER(NAME) LIKE '%.mp3' OR LOWER(NAME) LIKE '%.mts' OR LOWER(NAME) LIKE '%.ogg' OR
LOWER(NAME) LIKE '%.wma' OR LOWER(NAME) LIKE '%.wv' OR LOWER(NAME) LIKE '%.bin' OR LOWER(NAME) LIKE '%.img' OR
LOWER(NAME) LIKE '%.iso' OR LOWER(NAME) LIKE '%.docm' OR LOWER(NAME) LIKE '%.pps' OR LOWER(NAME) LIKE '%.pptx' OR
LOWER(NAME) LIKE '%.acsm' OR LOWER(NAME) LIKE '%.menc' OR LOWER(NAME) LIKE '%.emz' OR LOWER(NAME) LIKE '%.gif' OR
LOWER(NAME) LIKE '%.jpeg' OR LOWER(NAME) LIKE '%.jpg' OR LOWER(NAME) LIKE '%.png' OR LOWER(NAME) LIKE '%.htm' OR
LOWER(NAME) LIKE '%.swf' OR LOWER(NAME) LIKE '%.application' OR LOWER(NAME) LIKE '%.exe' OR LOWER(NAME) LIKE
'%.ipa' OR LOWER(NAME) LIKE '%.part1.exe' OR LOWER(NAME) LIKE '%.crw' OR LOWER(NAME) LIKE '%.cso' OR LOWER(NAME)
LIKE '%.mdi' OR LOWER(NAME) LIKE '%.odg' OR LOWER(NAME) LIKE '%.rpm' OR LOWER(NAME) LIKE '%.dcr' OR LOWER(NAME)
LIKE '%.jad' OR LOWER(NAME) LIKE '%.pak' OR LOWER(NAME) LIKE '%.rem' OR LOWER(NAME) LIKE '%.3g2' OR LOWER(NAME)
LIKE '%.3gp' OR LOWER(NAME) LIKE '%.asx' OR LOWER(NAME) LIKE '%.flv' OR LOWER(NAME) LIKE '%.m2t' OR LOWER(NAME)
LIKE '%.m2ts' OR LOWER(NAME) LIKE '%.m4v' OR LOWER(NAME) LIKE '%.mkv' OR LOWER(NAME) LIKE '%.mov' OR LOWER(NAME)
LIKE '%.mp4' OR LOWER(NAME) LIKE '%.mpg' OR LOWER(NAME) LIKE '%.tod' OR LOWER(NAME) LIKE '%.ts' OR LOWER(NAME)
LIKE '%.vob' OR LOWER(NAME) LIKE '%.wmv' OR LOWER(NAME) LIKE '%.hqx' OR LOWER(NAME) LIKE '%.docx' OR LOWER(NAME)
LIKE '%.ppt' OR LOWER(NAME) LIKE '%.pptm' OR LOWER(NAME) LIKE '%.thmx' OR LOWER(NAME) LIKE '%.djvu' OR
LOWER(NAME) LIKE '%.dt2' OR LOWER(NAME) LIKE '%.mrw' OR LOWER(NAME) LIKE '%.wbmp' OR LOWER(NAME) LIKE '%.abr' OR
LOWER(NAME) LIKE '%.ai' OR LOWER(NAME) LIKE '%.icon' OR LOWER(NAME) LIKE '%.ofx' OR LOWER(NAME) LIKE '%.pzl' OR
LOWER(NAME) LIKE '%.tif' OR LOWER(NAME) LIKE '%.u3d' OR LOWER(NAME) LIKE '%.msi' OR LOWER(NAME) LIKE '%.xlsm' OR
LOWER(NAME) LIKE '%.scr' OR LOWER(NAME) LIKE '%.wav' OR LOWER(NAME) LIKE '%.idx' OR LOWER(NAME) LIKE '%.abw' OR
LOWER(NAME) LIKE '%.azw' OR LOWER(NAME) LIKE '%.contact' OR LOWER(NAME) LIKE '%.dot' OR LOWER(NAME) LIKE '%.dotm'
OR LOWER(NAME) LIKE '%.dotx' OR LOWER(NAME) LIKE '%.epub' OR LOWER(NAME) LIKE '%.keynote' OR LOWER(NAME) LIKE
'%.mobi' OR LOWER(NAME) LIKE '%.mswmm' OR LOWER(NAME) LIKE '%.odt' OR LOWER(NAME) LIKE '%.one' OR LOWER(NAME)
LIKE '%.otf' OR LOWER(NAME) LIKE '%.pages' OR LOWER(NAME) LIKE '%.pdf' OR LOWER(NAME) LIKE '%.ppsx' OR
LOWER(NAME) LIKE '%.prproj' OR LOWER(NAME) LIKE '%.pwi' OR LOWER(NAME) LIKE '%.onepkg' OR LOWER(NAME) LIKE
'%.potx' OR LOWER(NAME) LIKE '%.tiff' OR LOWER(NAME) LIKE %.!ut OR LOWER(NAME) LIKE '%.atom' OR LOWER(NAME) LIKE
%.bc! OR LOWER(NAME) LIKE '%.opml' OR LOWER(NAME) LIKE '%.torrent' OR LOWER(NAME) LIKE '%.xhtml' OR LOWER(NAME)
LIKE '%.jar' OR LOWER(NAME) LIKE '%.xlsx' OR LOWER(NAME) LIKE '%.fnt' OR LOWER(NAME) LIKE '%.sc2replay' OR
LOWER(NAME) LIKE '%.1st' OR LOWER(NAME) LIKE '%.air' OR LOWER(NAME) LIKE '%.apk' OR LOWER(NAME) LIKE '%.cbr' OR
LOWER(NAME) LIKE '%.daa' OR LOWER(NAME) LIKE '%.isz' OR LOWER(NAME) LIKE '%.m3u8' OR LOWER(NAME) LIKE '%.rmvb' OR
LOWER(NAME) LIKE '%.sxw' OR LOWER(NAME) LIKE '%.tga' OR LOWER(NAME) LIKE '%.uax' OR LOWER(NAME) LIKE '%.crx' OR
LOWER(NAME) LIKE '%.safariextz' OR LOWER(NAME) LIKE '%.xpi' OR LOWER(NAME) LIKE '%.theme' OR LOWER(NAME) LIKE
'%.themepack' OR LOWER(NAME) LIKE '%.3dr' OR LOWER(NAME) LIKE '%.dic' OR LOWER(NAME) LIKE '%.dlc' OR LOWER(NAME)
LIKE '%.lng' OR LOWER(NAME) LIKE '%.ncs' OR LOWER(NAME) LIKE '%.pcd' OR LOWER(NAME) LIKE '%.pmd' OR LOWER(NAME)
LIKE '%.rss' OR LOWER(NAME) LIKE '%.sng' OR LOWER(NAME) LIKE '%.svp' OR LOWER(NAME) LIKE '%.swp' OR LOWER(NAME)
LIKE '%.thm' OR LOWER(NAME) LIKE '%.uif' OR LOWER(NAME) LIKE '%.upg' OR LOWER(NAME) LIKE '%.avi' OR LOWER(NAME)
LIKE '%.fla' OR LOWER(NAME) LIKE '%.pcm' OR LOWER(NAME) LIKE '%.bbb' OR LOWER(NAME) LIKE '%.bik' OR LOWER(NAME)
LIKE '%.nba' OR LOWER(NAME) LIKE '%.nbu' OR LOWER(NAME) LIKE '%.nco' OR LOWER(NAME) LIKE '%.wbcat' OR LOWER(NAME)
LIKE '%.dao' OR LOWER(NAME) LIKE '%.dmg' OR LOWER(NAME) LIKE '%.tao' OR LOWER(NAME) LIKE '%.toast' OR LOWER(NAME)
LIKE '%.pub' OR LOWER(NAME) LIKE '%.fpx' OR LOWER(NAME) LIKE '%.prg' OR LOWER(NAME) LIKE '%.cpt' OR LOWER(NAME)
LIKE '%.eml' OR LOWER(NAME) LIKE '%.nvram' OR LOWER(NAME) LIKE '%.vmsd' OR LOWER(NAME) LIKE '%.vmxf' OR
LOWER(NAME) LIKE '%.vswp'))
EFSSG1000I The command completed successfully.
[7803088.ibm]$
```

8. Before running the **runpolicy** command, you can use the GUI to verify current information for all of the pools, as shown in Figure 16-21. The system pool contains 4.40 GB of both compressed and noncompressible data, and the compressed pool contains 375.75 MB of used data.


IBM Storwize V7000 Unified
Welcome, admin (2 users online)
About | Logout | Help
Pendulum > Monitoring > Capacity
File Systems
File System Pools
File Sets
Users
User Groups
Filter

Last Updated: 7/9/13 8:03:00 PM

Name	Total	File System Used	Real Capacity	Free	% Used	Compression Savings	Thin Provisioning Efficiency
FS1	300.00 GB	4.77 GB	202.20 GB	295.23 GB	1.6%	48.50%	100.00%
☑ system	200.00 GB	4.40 GB	200.00 GB	195.60 GB	2.2%		
☑ compressed	100.00 GB	375.75 MB	2.20 GB	99.63 GB	0.4%	48.50%	100.00%

Selected 2 file system pools

Figure 16-21 Capacity

9. You can view the active jobs by using the **lsjobstatus** CLI command. To see the migration job results, use the **showlog** command. Example 16-7 shows the sequence of using the CLI commands to initiate the migration by using the **runpolicy** command and then monitoring progress.

Example 16-7 Job progress

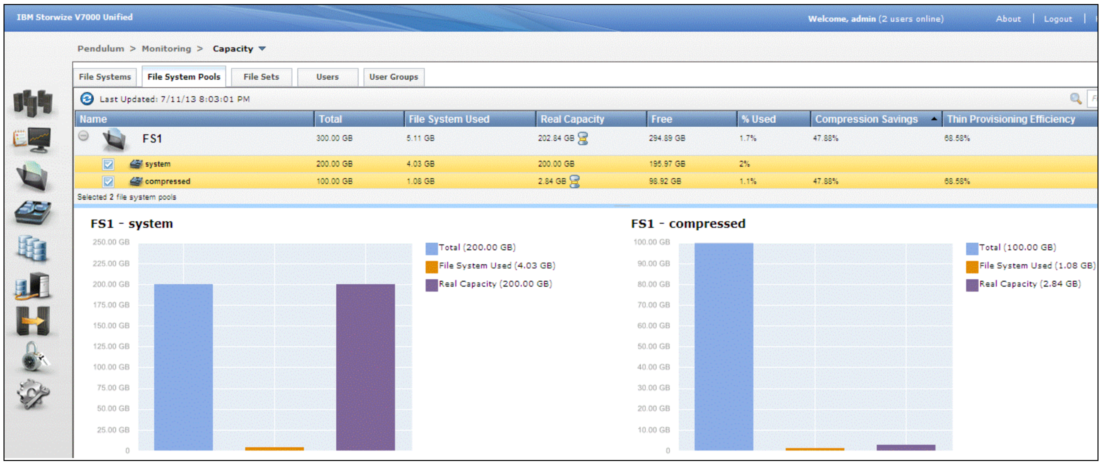
```
[7803088.ibm]$
[7803088.ibm]$ runpolicy FS1 -P migratetocompress
EFSSA0184I The policy is started on FS1 with JobID 58.
[7803088.ibm]$
[7803088.ibm]$ lsjobstatus
File system Job       Job id Status  Start time             End time/Progress RC Message
FS1         runpolicy 58     running 7/12/13 11:07:07 PM IDT                     Job has started.
EFSSG1000I The command completed successfully.
[7803088.ibm]$
[7803088.ibm]$ showlog 58
Primary node: mgmt001st001
Job ID : 58
[I] GPFS Current Data Pool Utilization in KB and %
compressed      6912     104857600       0.006592%
system  5104896 209715200       2.434204%
[I] 156427 of 1000448 inodes used: 15.635695%.
[I] Loaded policy rules from /var/opt/IBM/sofs/PolicyFiles/policy4599440165344313344.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP = 2013-07-12@20:07:08 UTC
parsed 0 Placement Rules, 0 Restore Rules, 1 Migrate/Delete/Exclude Rules,
        0 List Rules, 0 External Pool/List Rules
RULE 'migratetocompress' migrate from pool 'system' threshold(1,0) to pool 'compressed' where NOT (NOT
(LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE '%.7z.002' OR LOWER(NAME) LIKE
'%.7z.003' OR LOWER(NAME) LIKE '%.7zip' OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME) LIKE '%.a01' OR LOWER(NAME)
LIKE '%.a02' OR LOWER(NAME) LIKE '%.a03' OR LOWER(NAME) LIKE '%.a04' OR LOWER(NAME) LIKE '%.a05' OR LOWER(NAME)
LIKE '%.ace' OR LOWER(NAME) LIKE '%.arj' OR LOWER(NAME) LIKE '%.bkf' OR LOWER(NAME) LIKE '%.bz2' OR LOWER(NAME)
LIKE '%.c00' OR LOWER(NAME) LIKE '%.c01' OR LOWER(NAME) LIKE '%.c02' OR LOWER(NAME) LIKE '%.c03' OR LOWER(NAME)
LIKE '%.cab' OR LOWER(NAME) LIKE '%.cbz' OR LOWER(NAME) LIKE '%.cpgz' OR LOWER(NAME) LIKE '%.gz' OR LOWER(NAME)
LIKE '%.nbh' OR LOWER(NAME) LIKE '%.r00' OR LOWER(NAME) LIKE '%.r01' OR LOWER(NAME) LIKE '%.r02' OR LOWER(NAME)
LIKE '%.r03' OR LOWER(NAME) LIKE '%.r04' OR LOWER(NAME) LIKE '%.r05' OR LOWER(NAME) LIKE '%.r06' OR LOWER(NAME)
LIKE '%.r07' OR LOWER(NAME) LIKE '%.r08' OR LOWER(NAME) LIKE '%.r09' OR LOWER(NAME) LIKE '%.r10' OR LOWER(NAME)
LIKE '%.rar' OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME) LIKE '%.sit' OR LOWER(NAME) LIKE '%.sitx' OR LOWER(NAME)
LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER(NAME) LIKE '%.wba' OR LOWER(NAME) LIKE '%.z01' OR
LOWER(NAME) LIKE '%.z02' OR LOWER(NAME) LIKE '%.z03' OR LOWER(NAME) LIKE '%.z04' OR LOWER(NAME) LIKE '%.z05' OR
LOWER(NAME) LIKE '%.zip' OR LOWER(NAME) LIKE '%.zix' OR LOWER(NAME) LIKE '%.aac' OR LOWER(NAME) LIKE '%.cda' OR
LOWER(NAME) LIKE '%.dvf' OR LOWER(NAME) LIKE '%.flac' OR LOWER(NAME) LIKE '%.gp5' OR LOWER(NAME) LIKE '%.gpx' OR
LOWER(NAME) LIKE '%.logic' OR LOWER(NAME) LIKE '%.m4a' OR LOWER(NAME) LIKE '%.m4b' OR LOWER(NAME) LIKE '%.m4p' OR
LOWER(NAME) LIKE '%.mp3' OR LOWER(NAME) LIKE '%.mts' OR LOWER(NAME) LIKE '%.ogg' OR LOWER(NAME) LIKE '%.wma' OR
LOWER(NAME) LIKE '%.wv' OR LOWER(NAME) LIKE '%.bin' OR LOWER(NAME) LIKE '%.img' OR LOWER(NAME) LIKE '%.iso' OR
LOWER(NAME) LIKE '%.docm' OR LOWER(NAME) LIKE '%.pps' OR LOWER(NAME) LIKE '%.pptx' OR LOWER(NAME) LIKE '%.acsm'
OR LOWER(NAME) LIKE '%.menc' OR LOWER(NAME) LIKE '%.emz' OR LOWER(NAME) LIKE '%.gif' OR LOWER(NAME) LIKE '%.jpeg'
OR LOWER(NAME) LIKE '%.jpg' OR LOWER(NAME) LIKE '%.png' OR LOWER(NAME) LIKE '%.htm' OR LOWER(NAME) LIKE '%.swf'
OR LOWER(NAME) LIKE '%.application' OR LOWER(NAME) LIKE '%.exe' OR LOWER(NAME) LIKE '%.ipa' OR LOWER(NAME) LIKE
'%.part1.exe' OR LOWER(NAME) LIKE '%.crw' OR LOWER(NAME) LIKE '%.cso' OR LOWER(NAME) LIKE '%.mdi' OR LOWER(NAME)
LIKE '%.odg' OR LOWER(NAME) LIKE '%.rpm' OR LOWER(NAME) LIKE '%.dcr' OR LOWER(NAME) LIKE '%.jad' OR LOWER(NAME)
LIKE '%.pak' OR LOWER(NAME) LIKE '%.rem' OR LOWER(NAME) LIKE '%.3g2' OR LOWER(NAME) LIKE '%.3gp' OR LOWER(NAME)
LIKE '%.asx' OR LOWER(NAME) LIKE '%.flv' OR LOWER(NAME) LIKE '%.m2t' OR LOWER(NAME) LIKE '%.m2ts' OR LOWER(NAME)
LIKE '%.m4v' OR LOWER(NAME) LIKE '%.mkv' OR LOWER(NAME) LIKE '%.mov' OR LOWER(NAME) LIKE '%.mp4' OR LOWER(NAME)
LIKE '%.mpg' OR LOWER(NAME) LIKE '%.tod' OR LOWER(NAME) LIKE '%.ts' OR LOWER(NAME) LIKE '%.vob' OR LOWER(NAME)
LIKE '%.wmv' OR LOWER(NAME) LIKE '%.hqx' OR LOWER(NAME) LIKE '%.docx' OR LOWER(NAME) LIKE '%.ppt' OR LOWER(NAME)
LIKE '%.pptm' OR LOWER(NAME) LIKE '%.thmx' OR LOWER(NAME) LIKE '%.djvu' OR LOWER(NAME) LIKE '%.dt2' OR
LOWER(NAME) LIKE '%.mrw' OR LOWER(NAME) LIKE '%.wbmp' OR LOWER(NAME) LIKE '%.abr' OR LOWER(NAME) LIKE '%.ai' OR
LOWER(NAME) LIKE '%.icon' OR LOWER(NAME) LIKE '%.ofx' OR LOWER(NAME) LIKE '%.pzl' OR LOWER(NAME) LIKE '%.tif' OR
LOWER(NAME) LIKE '%.u3d' OR LOWER(NAME) LIKE '%.msi' OR LOWER(NAME) LIKE '%.xlsm' OR LOWER(NAME) LIKE '%.scr' OR
LOWER(NAME) LIKE '%.wav' OR LOWER(NAME) LIKE '%.idx' OR LOWER(NAME) LIKE '%.abw' OR LOWER(NAME) LIKE '%.azw' OR
LOWER(NAME) LIKE '%.contact' OR LOWER(NAME) LIKE '%.dot' OR LOWER(NAME) LIKE '%.dotm' OR LOWER(NAME) LIKE
'%.dotx' OR LOWER(NAME) LIKE '%.epub' OR LOWER(NAME) LIKE '%.keynote' OR LOWER(NAME) LIKE '%.mobi' OR LOWER(NAME)
LIKE '%.mswmm' OR LOWER(NAME) LIKE '%.odt' OR LOWER(NAME) LIKE '%.one' OR LOWER(NAME) LIKE '%.otf' OR LOWER(NAME)
LIKE '%.pages' OR LOWER(NAME) LIKE '%.pdf' OR LOWER(NAME) LIKE '%.ppsx' OR LOWER(NAME) LIKE '%.prproj' OR
LOWER(NAME) LIKE '%.pwi' OR LOWER(NAME) LIKE '%.onepkg' OR LOWER(NAME) LIKE '%.potx' OR LOWER(NAME) LIKE '%.tiff'
OR LOWER(NAME) LIKE '%.\!ut' OR LOWER(NAME) LIKE '%.atom' OR LOWER(NAME) LIKE '%.bc\!' OR LOWER(NAME) LIKE
'%.opml' OR LOWER(NAME) LIKE '%.torrent' OR LOWER(NAME) LIKE '%.xhtml' OR LOWER(NAME) LIKE '%.jar' OR LOWER(NAME)
LIKE '%.xlsx' OR LOWER(NAME) LIKE '%.fnt' OR LOWER(NAME) LIKE '%.sc2replay' OR LOWER(NAME) LIKE '%.1st' OR
LOWER(NAME) LIKE '%.air' OR LOWER(NAME) LIKE '%.apk' OR LOWER(NAME) LIKE '%.cbr' OR LOWER(NAME) LIKE '%.daa' OR
LOWER(NAME) LIKE '%.isz' OR LOWER(NAME) LIKE '%.m3u8' OR LOWER(NAME) LIKE '%.rmvb' OR LOWER(NAME) LIKE '%.sxw' OR
LOWER(NAME) LIKE '%.tga' OR LOWER(NAME) LIKE '%.uax' OR LOWER(NAME) LIKE '%.crx' OR LOWER(NAME) LIKE
'%.safariextz' OR LOWER(NAME) LIKE '%.xpi' OR LOWER(NAME) LIKE '%.theme' OR LOWER(NAME) LIKE '%.themepack' OR
LOWER(NAME) LIKE '%.3dr' OR LOWER(NAME) LIKE '%.dic' OR LOWER(NAME) LIKE '%.dlc' OR LOWER(NAME) LIKE '%.lng' OR
LOWER(NAME) LIKE '%.ncs' OR LOWER(NAME) LIKE '%.pcd' OR LOWER(NAME) LIKE '%.pmd' OR LOWER(NAME) LIKE '%.rss' OR
LOWER(NAME) LIKE '%.sng' OR LOWER(NAME) LIKE '%.svp' OR LOWER(NAME) LIKE '%.swp' OR LOWER(NAME) LIKE '%.thm' OR
LOWER(NAME) LIKE '%.uif' OR LOWER(NAME) LIKE '%.upg' OR LOWER(NAME) LIKE '%.avi' OR LOWER(NAME) LIKE '%.fla' OR
LOWER(NAME) LIKE '%.pcm' OR LOWER(NAME) LIKE '%.bbb' OR LOWER(NAME) LIKE '%.bik' OR LOWER(NAME) LIKE '%.nba' OR
LOWER(NAME) LIKE '%.nbu' OR LOWER(NAME) LIKE '%.nco' OR LOWER(NAME) LIKE '%.wbcat' OR LOWER(NAME) LIKE '%.dao' OR
LOWER(NAME) LIKE '%.dmg' OR LOWER(NAME) LIKE '%.tao' OR LOWER(NAME) LIKE '%.toast' OR LOWER(NAME) LIKE '%.pub' OR
LOWER(NAME) LIKE '%.fpx' OR LOWER(NAME) LIKE '%.prg' OR LOWER(NAME) LIKE '%.cpt' OR LOWER(NAME) LIKE '%.eml' OR
LOWER(NAME) LIKE '%.nvram' OR LOWER(NAME) LIKE '%.vmsd' OR LOWER(NAME) LIKE '%.vmxf' OR LOWER(NAME) LIKE
'%.vswp'))
[I] Directories scan: 144224 files, 8224 directories, 0 other objects, 0 'skipped' files and/or errors.
[I] Summary of Rule Applicability and File Choices:
 Rule#  Hit_Cnt KB_Hit  Chosen  KB_Chosen       KB_Ill  Rule
  0     57894   1128288 57894   1128288 16      RULE 'migratetocompress' MIGRATE FROM POOL 'system'
THRESHOLD(1,0) TO POOL 'compressed' WHERE(.)
[I] Filesystem objects with no applicable rules: 94540.
[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 1128288KB: 57894 of 57894 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 0KB: 0 of 0 candidates;
 Chose to list 0KB: 0 of 0 candidates;
 16KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
compressed      1135200 104857600       1.082611%
system  3994784 209715200       1.904861%
[I] Because some data is illplaced or illreplicated, predicted pool utilization may be negative and/or
misleading!
----------------------------------------------------------
End of log - runpolicy still running
----------------------------------------------------------

EFSSG1000I The command completed successfully.
[7803088.ibm]$
[7803088.ibm]$ showlog 58
Primary node: mgmt001st001
Job ID : 58
[I] GPFS Current Data Pool Utilization in KB and %
compressed      6912    104857600       0.006592%
system  5104896 209715200       2.434204%
[I] 156427 of 1000448 inodes used: 15.635695%.
[I] Loaded policy rules from /var/opt/IBM/sofs/PolicyFiles/policy4599440165344313344.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP = 2013-07-12@20:07:08 UTC
parsed 0 Placement Rules, 0 Restore Rules, 1 Migrate/Delete/Exclude Rules,
        0 List Rules, 0 External Pool/List Rules
RULE 'migratetocompress' migrate from pool 'system' threshold(1,0) to pool 'compressed' where NOT (NOT
(LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE '%.7z.002' OR LOWER(NAME) LIKE
'%.7z.003' OR LOWER(NAME) LIKE '%.7zip' OR LOWER(NAME) LIKE '%.a00' OR LOWER(NAME) LIKE '%.a01' OR LOWER(NAME)
LIKE '%.a02' OR LOWER(NAME) LIKE '%.a03' OR LOWER(NAME) LIKE '%.a04' OR LOWER(NAME) LIKE '%.a05' OR LOWER(NAME)
LIKE '%.ace' OR LOWER(NAME) LIKE '%.arj' OR LOWER(NAME) LIKE '%.bkf' OR LOWER(NAME) LIKE '%.bz2' OR LOWER(NAME)
LIKE '%.c00' OR LOWER(NAME) LIKE '%.c01' OR LOWER(NAME) LIKE '%.c02' OR LOWER(NAME) LIKE '%.c03' OR LOWER(NAME)
LIKE '%.cab' OR LOWER(NAME) LIKE '%.cbz' OR LOWER(NAME) LIKE '%.cpgz' OR LOWER(NAME) LIKE '%.gz' OR LOWER(NAME)
LIKE '%.nbh' OR LOWER(NAME) LIKE '%.r00' OR LOWER(NAME) LIKE '%.r01' OR LOWER(NAME) LIKE '%.r02' OR LOWER(NAME)
LIKE '%.r03' OR LOWER(NAME) LIKE '%.r04' OR LOWER(NAME) LIKE '%.r05' OR LOWER(NAME) LIKE '%.r06' OR LOWER(NAME)
LIKE '%.r07' OR LOWER(NAME) LIKE '%.r08' OR LOWER(NAME) LIKE '%.r09' OR LOWER(NAME) LIKE '%.r10' OR LOWER(NAME)
LIKE '%.rar' OR LOWER(NAME) LIKE '%.sisx' OR LOWER(NAME) LIKE '%.sit' OR LOWER(NAME) LIKE '%.sitx' OR LOWER(NAME)
LIKE '%.tar.gz' OR LOWER(NAME) LIKE '%.tgz' OR LOWER(NAME) LIKE '%.wba' OR LOWER(NAME) LIKE '%.z01' OR
```

```
LOWER(NAME) LIKE '%.z02' OR LOWER(NAME) LIKE '%.z03' OR LOWER(NAME) LIKE '%.z04' OR LOWER(NAME) LIKE '%.z05' OR
LOWER(NAME) LIKE '%.zip' OR LOWER(NAME) LIKE '%.zix' OR LOWER(NAME) LIKE '%.aac' OR LOWER(NAME) LIKE '%.cda' OR
LOWER(NAME) LIKE '%.dvf' OR LOWER(NAME) LIKE '%.flac' OR LOWER(NAME) LIKE '%.gp5' OR LOWER(NAME) LIKE '%.gpx' OR
LOWER(NAME) LIKE '%.logic' OR LOWER(NAME) LIKE '%.m4a' OR LOWER(NAME) LIKE '%.m4b' OR LOWER(NAME) LIKE '%.m4p' OR
LOWER(NAME) LIKE '%.mp3' OR LOWER(NAME) LIKE '%.mts' OR LOWER(NAME) LIKE '%.ogg' OR LOWER(NAME) LIKE '%.wma' OR
LOWER(NAME) LIKE '%.wv' OR LOWER(NAME) LIKE '%.bin' OR LOWER(NAME) LIKE '%.img' OR LOWER(NAME) LIKE '%.iso' OR
LOWER(NAME) LIKE '%.docm' OR LOWER(NAME) LIKE '%.pps' OR LOWER(NAME) LIKE '%.pptx' OR LOWER(NAME) LIKE '%.acsm'
OR LOWER(NAME) LIKE '%.menc' OR LOWER(NAME) LIKE '%.emz' OR LOWER(NAME) LIKE '%.gif' OR LOWER(NAME) LIKE '%.jpeg'
OR LOWER(NAME) LIKE '%.jpg' OR LOWER(NAME) LIKE '%.png' OR LOWER(NAME) LIKE '%.htm' OR LOWER(NAME) LIKE '%.swf'
OR LOWER(NAME) LIKE '%.application' OR LOWER(NAME) LIKE '%.exe' OR LOWER(NAME) LIKE '%.ipa' OR LOWER(NAME) LIKE
'%.part1.exe' OR LOWER(NAME) LIKE '%.crw' OR LOWER(NAME) LIKE '%.cso' OR LOWER(NAME) LIKE '%.mdi' OR LOWER(NAME)
LIKE '%.odg' OR LOWER(NAME) LIKE '%.rpm' OR LOWER(NAME) LIKE '%.dcr' OR LOWER(NAME) LIKE '%.jad' OR LOWER(NAME)
LIKE '%.pak' OR LOWER(NAME) LIKE '%.rem' OR LOWER(NAME) LIKE '%.3g2' OR LOWER(NAME) LIKE '%.3gp' OR LOWER(NAME)
LIKE '%.asx' OR LOWER(NAME) LIKE '%.flv' OR LOWER(NAME) LIKE '%.m2t' OR LOWER(NAME) LIKE '%.m2ts' OR LOWER(NAME)
LIKE '%.m4v' OR LOWER(NAME) LIKE '%.mkv' OR LOWER(NAME) LIKE '%.mov' OR LOWER(NAME) LIKE '%.mp4' OR LOWER(NAME)
LIKE '%.mpg' OR LOWER(NAME) LIKE '%.tod' OR LOWER(NAME) LIKE '%.ts' OR LOWER(NAME) LIKE '%.vob' OR LOWER(NAME)
LIKE '%.wmv' OR LOWER(NAME) LIKE '%.hqx' OR LOWER(NAME) LIKE '%.docx' OR LOWER(NAME) LIKE '%.ppt' OR LOWER(NAME)
LIKE '%.pptm' OR LOWER(NAME) LIKE '%.thmx' OR LOWER(NAME) LIKE '%.djvu' OR LOWER(NAME) LIKE '%.dt2' OR
LOWER(NAME) LIKE '%.mrw' OR LOWER(NAME) LIKE '%.wbmp' OR LOWER(NAME) LIKE '%.abr' OR LOWER(NAME) LIKE '%.ai' OR
LOWER(NAME) LIKE '%.icon' OR LOWER(NAME) LIKE '%.ofx' OR LOWER(NAME) LIKE '%.pzl' OR LOWER(NAME) LIKE '%.tif' OR
LOWER(NAME) LIKE '%.u3d' OR LOWER(NAME) LIKE '%.msi' OR LOWER(NAME) LIKE '%.xlsm' OR LOWER(NAME) LIKE '%.scr' OR
LOWER(NAME) LIKE '%.wav' OR LOWER(NAME) LIKE '%.idx' OR LOWER(NAME) LIKE '%.abw' OR LOWER(NAME) LIKE '%.azw' OR
LOWER(NAME) LIKE '%.contact' OR LOWER(NAME) LIKE '%.dot' OR LOWER(NAME) LIKE '%.dotm' OR LOWER(NAME) LIKE
'%.dotx' OR LOWER(NAME) LIKE '%.epub' OR LOWER(NAME) LIKE '%.keynote' OR LOWER(NAME) LIKE '%.mobi' OR LOWER(NAME)
LIKE '%.mswmm' OR LOWER(NAME) LIKE '%.odt' OR LOWER(NAME) LIKE '%.one' OR LOWER(NAME) LIKE '%.otf' OR LOWER(NAME)
LIKE '%.pages' OR LOWER(NAME) LIKE '%.pdf' OR LOWER(NAME) LIKE '%.ppsx' OR LOWER(NAME) LIKE '%.prproj' OR
LOWER(NAME) LIKE '%.pwi' OR LOWER(NAME) LIKE '%.onepkg' OR LOWER(NAME) LIKE '%.potx' OR LOWER(NAME) LIKE '%.tiff'
OR LOWER(NAME) LIKE '%.\!ut' OR LOWER(NAME) LIKE '%.atom' OR LOWER(NAME) LIKE '%.bc\!' OR LOWER(NAME) LIKE
'%.opml' OR LOWER(NAME) LIKE '%.torrent' OR LOWER(NAME) LIKE '%.xhtml' OR LOWER(NAME) LIKE '%.jar' OR LOWER(NAME)
LIKE '%.xlsx' OR LOWER(NAME) LIKE '%.fnt' OR LOWER(NAME) LIKE '%.sc2replay' OR LOWER(NAME) LIKE '%.1st' OR
LOWER(NAME) LIKE '%.air' OR LOWER(NAME) LIKE '%.apk' OR LOWER(NAME) LIKE '%.cbr' OR LOWER(NAME) LIKE '%.daa' OR
LOWER(NAME) LIKE '%.isz' OR LOWER(NAME) LIKE '%.m3u8' OR LOWER(NAME) LIKE '%.rmvb' OR LOWER(NAME) LIKE '%.sxw' OR
LOWER(NAME) LIKE '%.tga' OR LOWER(NAME) LIKE '%.uax' OR LOWER(NAME) LIKE '%.crx' OR LOWER(NAME) LIKE
'%.safariextz' OR LOWER(NAME) LIKE '%.xpi' OR LOWER(NAME) LIKE '%.theme' OR LOWER(NAME) LIKE '%.themepack' OR
LOWER(NAME) LIKE '%.3dr' OR LOWER(NAME) LIKE '%.dic' OR LOWER(NAME) LIKE '%.dlc' OR LOWER(NAME) LIKE '%.lng' OR
LOWER(NAME) LIKE '%.ncs' OR LOWER(NAME) LIKE '%.pcd' OR LOWER(NAME) LIKE '%.pmd' OR LOWER(NAME) LIKE '%.rss' OR
LOWER(NAME) LIKE '%.sng' OR LOWER(NAME) LIKE '%.svp' OR LOWER(NAME) LIKE '%.swp' OR LOWER(NAME) LIKE '%.thm' OR
LOWER(NAME) LIKE '%.uif' OR LOWER(NAME) LIKE '%.upg' OR LOWER(NAME) LIKE '%.avi' OR LOWER(NAME) LIKE '%.fla' OR
LOWER(NAME) LIKE '%.pcm' OR LOWER(NAME) LIKE '%.bbb' OR LOWER(NAME) LIKE '%.bik' OR LOWER(NAME) LIKE '%.nba' OR
LOWER(NAME) LIKE '%.nbu' OR LOWER(NAME) LIKE '%.nco' OR LOWER(NAME) LIKE '%.wbcat' OR LOWER(NAME) LIKE '%.dao' OR
LOWER(NAME) LIKE '%.dmg' OR LOWER(NAME) LIKE '%.tao' OR LOWER(NAME) LIKE '%.toast' OR LOWER(NAME) LIKE '%.pub' OR
LOWER(NAME) LIKE '%.fpx' OR LOWER(NAME) LIKE '%.prg' OR LOWER(NAME) LIKE '%.cpt' OR LOWER(NAME) LIKE '%.eml' OR
LOWER(NAME) LIKE '%.nvram' OR LOWER(NAME) LIKE '%.vmsd' OR LOWER(NAME) LIKE '%.vmxf' OR LOWER(NAME) LIKE
'%.vswp'))
[I] Directories scan: 144224 files, 8224 directories, 0 other objects, 0 'skipped' files and/or errors.
[I] Summary of Rule Applicability and File Choices:
 Rule#  Hit_Cnt KB_Hit  Chosen  KB_Chosen       KB_Ill  Rule
  0     57894   1128288 57894   1128288 16      RULE 'migratetocompress' MIGRATE FROM POOL 'system'
THRESHOLD(1,0) TO POOL 'compressed' WHERE(.)
[I] Filesystem objects with no applicable rules: 94540.
[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 1128288KB: 57894 of 57894 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 0KB: 0 of 0 candidates;
 Chose to list 0KB: 0 of 0 candidates;
 16KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
compressed      1135200 104857600       1.082611%
system  3994784 209715200       1.904861%
[I] Because some data is illplaced or illreplicated, predicted pool utilization may be negative and/or
misleading!
[I] A total of 57894 files have been migrated, deleted or processed by an EXTERNAL EXEC/script;
        0 'skipped' files and/or errors.
----------------------------------------------------------
End of log - runpolicy job ended
----------------------------------------------------------

EFSSG1000I The command completed successfully.
[7803088.ibm]$
```

10. After running the **runpolicy** command, you can use the GUI to verify current information for all of the pools, as shown Figure 16-22 on page 314. Make sure that data moved between pools as expected.


IBM Storwize V7000 Unified
Welcome, admin (2 users online)
About
Logout
Pendulum > Monitoring > Capacity
File Systems
File System Pools
File Sets
Users
User Groups

Last Updated: 7/11/13 8:03:01 PM

Name	Total	File System Used	Real Capacity	Free	% Used	Compression Savings	Thin Provisioning Efficiency
FS1	300.00 GB	5.11 GB	202.84 GB	294.89 GB	1.7%	47.88%	68.58%
☑ system	200.00 GB	4.03 GB	200.00 GB	195.97 GB	2%		
☑ compressed	100.00 GB	1.08 GB	2.84 GB	98.92 GB	1.1%	47.88%	68.58%

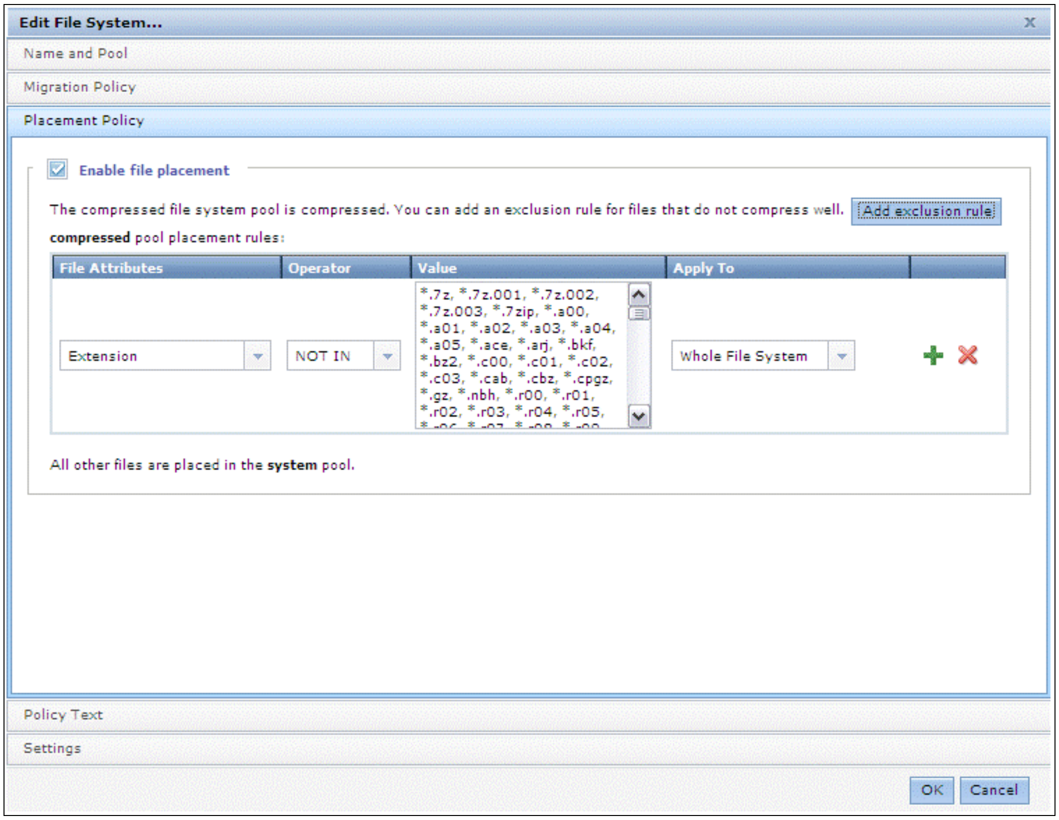
Selected 2 file system pools

Figure 16-22 Capacity update

Enabling a placement policy after the migration

After the migration completes, add a placement policy so that when new files are created, they are directed automatically to the proper pool, depending on whether the files are compressible or not.

1. You can do this through the GUI in the Edit File System window shown in Figure 16-23. Select the **Placement Policy** tab, click the **Enable file placement** check box, and click **Add exclusion rule**. This automatically adds the rule that places any new files that are created into either the compressed or uncompressed file system pool. Click **OK** to finish.



☑ Enable file placement

The compressed file system pool is compressed. You can add an exclusion rule for files that do not compress well. Add exclusion rule

compressed pool placement rules:

File Attributes	Operator	Value	Apply To
Extension	NOT IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz, *.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05, *.r06, *.r07, *.r08, *.r09	Whole File System

All other files are placed in the **system** pool.

Figure 16-23 Placement policy

Migrating data from the compressed pool to the system pool

You can to revert a selectively compressed file system to uncompressed while there is full data access. Use the following steps with the CLI to migrate a file system that has an uncompressed system pool and a second compressed file system pool to a fully uncompressed file system pool:

1. Analyze how much physical space is needed to move all data to the system pool. It is easiest to do the conversion when the uncompressed space that you need is available as free disk space, in addition to the compressed pool. If this is not the case, you must take a step-by-step approach and remove volumes from the compressed pool that is emptied.
2. Create a default placement rule for the file system by using this rule:

   ```
   RULE 'default' SET POOL 'system'
   ```

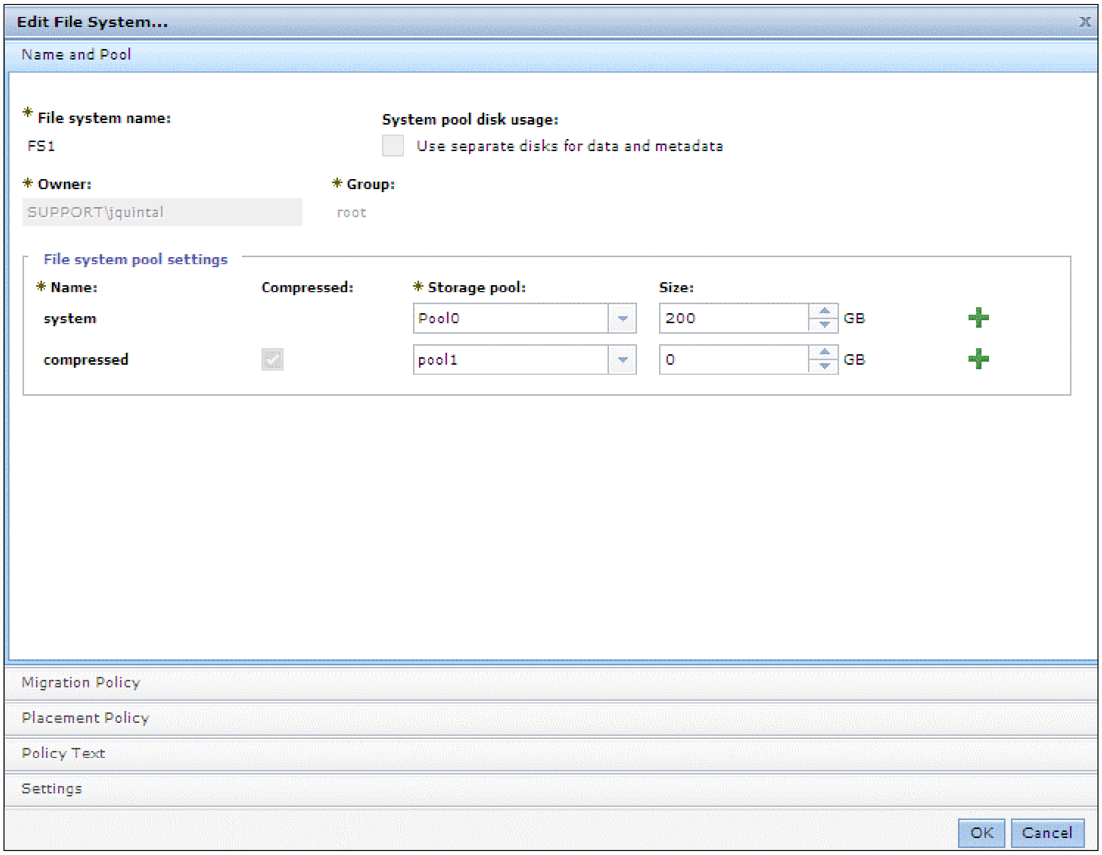
3. Create and run a migration rule that moves data to the system pool. To migrate data back to the system pool, a simple policy rule that moves all data to the system pool from the compressed pool is created. A migration policy can be applied to a file system. This ensures that the migration starts automatically and is executed regularly. This happens if the policy is created in the GUI policy text editor. Because the migration rule must run only once, another approach is chosen. The CLI is used to create a policy, start a policy run, and monitor the migration of data. The migration starts only when the **`runpolicy`** command is executed by using the CLI. This procedure is shown in example Example 16-8.

Example 16-8 Uncompressing data

```
# Create a policy rule - this policy is not applied to any file system
[7802378.ibm]# mkpolicy migratetosystem -R "RULE 'migratetosystem' migrate from
pool 'compressed' to pool 'system';"
EFSSG1000I The command completed successfully.
# The runpolicy command will only run the policy once on our file system
[7802378.ibm]$ runpolicy FS1 -P migratetosystem
EFSSA0184I The policy is started on FS1_test with JobID 806.
# The showlog command provides the output of policy runs
[7802378.ibm]$ showlog 806
Primary node: mgmt001st001
Job ID : 806
[I] GPFS Current Data Pool Utilization in KB and %
compressed 6531584 39845888 16.392116%
system 20494848 49283072 41.585979%
[I] 1000507 of 1200640 inodes used: 83.331140%.
[I] Loaded policy rules from
/var/opt/IBM/sofs/PolicyFiles/policy4252244384126468096.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP =
2012-11-15@07:39:05 UTC
parsed 0 Placement Rules, 0 Restore Rules, 1 Migrate/Delete/Exclude Rules,
0 List Rules, 0 External Pool/List Rules
RULE 'migratetosystem' migrate from pool 'compressed' to pool 'system'
[I] Directories scan: 923603 files, 72924 directories, 0 other objects, 0
'skipped' files and/or errors.
[I] Summary of Rule Applicability and File Choices:
Rule# Hit_Cnt KB_Hit ChosenKB_Chosen nkB_Ill Rule
0574246 49452407 424649452400RULE 'migratetosystem' MIGRATE FROM POOL
'compressed' TO POOL 'system'
[I] Filesystem objects with no applicable rules: 422280.
[I] GPFS Policy Decisions and File Choice Totals:
Chose to migrate 4945240KB: 574246 of 574246 candidates;
Chose to premigrate 0KB: 0 candidates;
Already co-managed 0KB: 0 candidates;
Chose to delete 0KB: 0 of 0 candidates;
Chose to list 0KB: 0 of 0 candidates;
0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
compressed1586344398458883.981199%
system254728564928307251.686827%
[I] A total of 574246 files have been migrated, deleted or processed by an
EXTERNAL EXEC/script;
0 'skipped' files and/or errors.
---------------------------------------------------------
End of log - runpolicy job ended
---------------------------------------------------------
EFSSG1000I The command completed successfully.
```

4. Remove all volumes of the compressed pool.

After the policy run has finished and all data was moved to the system pool, use the Edit File System dialog window in the GUI to remove the volumes of the compressed pool. Setting the capacity of the compressed storage pool size to **0** GB, as shown in Figure 16-24, removes all compressed volumes (NSDs). Also remove the compressed file system pool.



* File system name: FS1

System pool disk usage: □ Use separate disks for data and metadata

* Owner: SUPPORT\jquintal

* Group: root

File system pool settings

* Name:	Compressed:	* Storage pool:	Size:
system		Pool0	200 GB
compressed	☑	pool1	0 GB

Figure 16-24 Emptying the compressed pool

16.4 Capacity planning

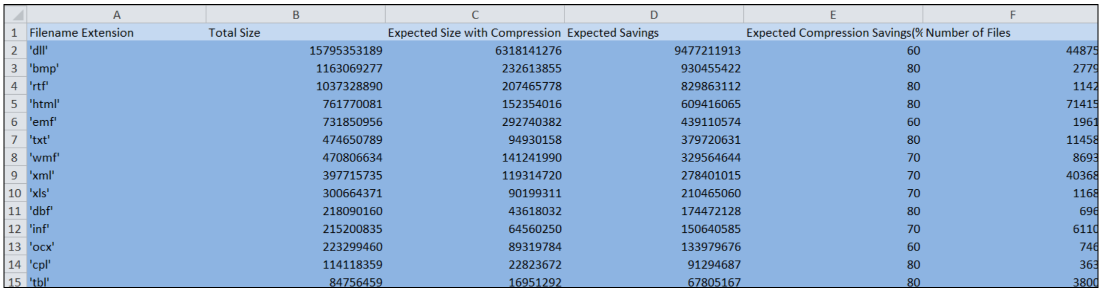
The capacity planning approach is determined by the file system storage pool architecture that was chosen. The approaches that are described here are a fully compressed file system and a selectively compressed file system.

16.4.1 Planning capacity with the NAS Compression Estimation Utility

The NAS Compression Estimation Utility is a command-line host-based utility that can be used to estimate the expected compression ratio in a network-attached storage (NAS) environment and can also be used to plan the size of a file system. You can download it from this web page:

http://www14.software.ibm.com/webapp/set2/sas/f/comprestimator/NAS_Compression_estimation_utility.html

The utility uses a file type extension list (located in the installation directory) with the corresponding compression ratio. The utility creates a Microsoft Excel spreadsheet (Figure 16-25) that lists the expected compression rate and the total size before and after compression, plus expected storage space savings. The information is presented as a total for each file extension. You can use this information to size the file system compression pool.



	A	B	C	D	E	F
1	Filename Extension	Total Size	Expected Size with Compression	Expected Savings	Expected Compression Savings(%	Number of Files
2	'dll'	15795353189	6318141276	9477211913	60	44875
3	'bmp'	1163069277	232613855	930455422	80	2779
4	'rtf'	1037328890	207465778	829863112	80	1142
5	'html'	761770081	152354016	609416065	80	71415
6	'emf'	731850956	292740382	439110574	60	1961
7	'txt'	474650789	94930158	379720631	80	11458
8	'wmf'	470806634	141241990	329564644	70	8693
9	'xml'	397715735	119314720	278401015	70	40368
10	'xls'	300664371	90199311	210465060	70	1168
11	'dbf'	218090160	43618032	174472128	80	696
12	'inf'	215200835	64560250	150640585	70	6110
13	'ocx'	223299460	89319784	133979676	60	746
14	'cpl'	114118359	22823672	91294687	80	363
15	'tbl'	84756459	16951292	67805167	80	3800

Figure 16-25 NAS Compression Utility output

The utility runs on a Microsoft Windows host that has access to the target filer and to the N series, NetApp filer, or CIFS share that is to be analyzed. It performs only read operations, so it has no effect on the data that is stored in the NAS. This information provides insight into how much physical space is needed to store compressed data. The average compressibility of the data provides the information that helps you size and configure a fully compressed file system.

Note: The utility updates the access time of scanned directories.

The utility is supported on these platforms:

- Windows file servers running Windows Server 2003, 2008, 2012 and later
- NetApp or N series filers running ONTAP 8.0.x and later

Note: The utility might run for a long time when it is scanning an entire NAS or a share with a large amount of data.

16.4.2 Installing and using the NAS Compression Estimation Utility

If you installed the *NAS Compression Estimation Utility* by using a Windows installation file, the utility files are available in the folder that you selected during the installation.

By default, the files are copied to the following directories:

In a Windows 64-bit system:

```
C:\Program Files (x86)\ibm\NAS Compression Estimation Utility
```

In a Windows 32-bit system:

```
C:\Program Files\ibm\ NAS Compression Estimation Utility
```

To install NAS Compression Estimation Utility on Windows, run this executable file:

```
NAS_Compression_estimation_utility.exe
```

Then, follow the prompts in the installation wizard.

16.4.3 Using NAS Compression Estimation Utility

To use the NAS Compression Estimation Utility on a Windows server, follow these steps:

1. Log in to the host by using an account that has domain administrator privileges. If it is a CIFS mount on another domain, it is advisable to authenticate it with domain admin user credentials from that domain to traverse all of the files.
2. Open an elevated command prompt with administrator rights (Run as Administrator). Navigate to the installation folder.
3. Run the NAS Compression Estimation Utility by using the syntax in Example 16-9 and the flags that are listed in Table 16-1.

Example 16-9 Syntax using Microsoft Windows

```
NAS_RtC_Estimate.exe -target unc_path | -filer hostname|ip [-batch file [-detail]]
[-casesense] [-category] [-heartbeats] [-ignoreDirs dir1,dir2...] [-logFile file]
[-loglevel 1-6] [-threads 1-50] [-loglevel 1-6]
```

Table 16-1 Syntax flags

`--target`	Specifies shares to analyze. You can also can specify a local or mapped drive. Example 1: `--target \\192.168.1.2\sharename` Example 2: `--target M:`
`--filer`	Specifies the whole filer by IP address or a host name. The utility prepares the list of all shares and scans them one by one. Example: `--filer 192.168.1.2`
`--batch`	Configuration file that contains a list of shares to analyze. This file must contain a list of shares in a specific format, with one share in each line. Example 1: `\\192.168.1.2\sharename` Example 2: `\\192.168.1.2\sharename\directory` Example 3: `\\hostname01\sharename`
`--casesense`	By default, the file extensions are not case-sensitive. Example: `.ISO` and `.iso` are considered the to be the same file extension. Set `--casesense` to modify the utility to be case-sensitive.
`--category`	Group file types by category in the final report (as listed in the .csv extensions file)
`--detail`	Used only with --batch. Creates a separate .csv for each share that is listed in the file.
`--heartbeats`	Specifies the frequency of the progress information update during the scan. Displays how many files and folders are scanned every n seconds. Default value is 5. Example: `--heartbeats 2` displays progress information every 2 seconds.
`--ignoreDirs`	Used to specify an exclude list for certain directories. By default, directories named `.snapshot` and so on are skipped. Hidden folders are scanned.
`--logFile`	Specify the log file path and name or only the name (installation folder) for the log file. By default, `utility.log` is used.
`--loglevel`	Specifies the log level. Default value is 2, range is 1 - 6. Levels from high to low: TRACE, DEBUG, INFO, WARN, ERROR, FATAL
`--threads`	Specifies the number of concurrent running threads. Default value is 10, valid range is 1 - 50. Example: `--threads 15`

Example 16-10 shows a typical use case for the utility:

Example 16-10 Running the utility (use case)

```
c:\Program Files (x86)\IBM\NAS Compression Estimation Utility>NAS_RtC_Estimate
--target \\9.11.235.74\oracle
Scan initiated - Version: 1.1.001.025
Username: jquintal
Command line: NAS_RtC_Estimate --target \\9.11.235.74\oracle
Processing share #1 - \\9.11.235.74\oracle
Start time: Fri Nov 08 16:15:16 2013
Total data: 19.99 GB Savings: 9.99 GB Number of files: 2 Number of folders: 1
Compression Ratio: 50 Number of errors: 0
End time: Fri Nov 08 16:15:17 2013
```

16.4.4 Capacity planning for selectively compressed file systems

The sizing for the overall physical disk capacity that is needed for a selectively compressed file system can also be done with the output of the *NAS Compression Estimation Utility*. However, to configure a selectively compressed file system, more analysis steps might be needed.

To configure a selectively compressed file system, you must know the average compression rate and the distribution of files into the compressed and uncompressed pools. For best performance, files that yield compression storage space savings above 40% are best for compression, and files that yield less than 40% are better to evaluate for performance. This rule is also applied by the exclusion list described in 16.3, “A selectively compressed file system with two pools” on page 290.

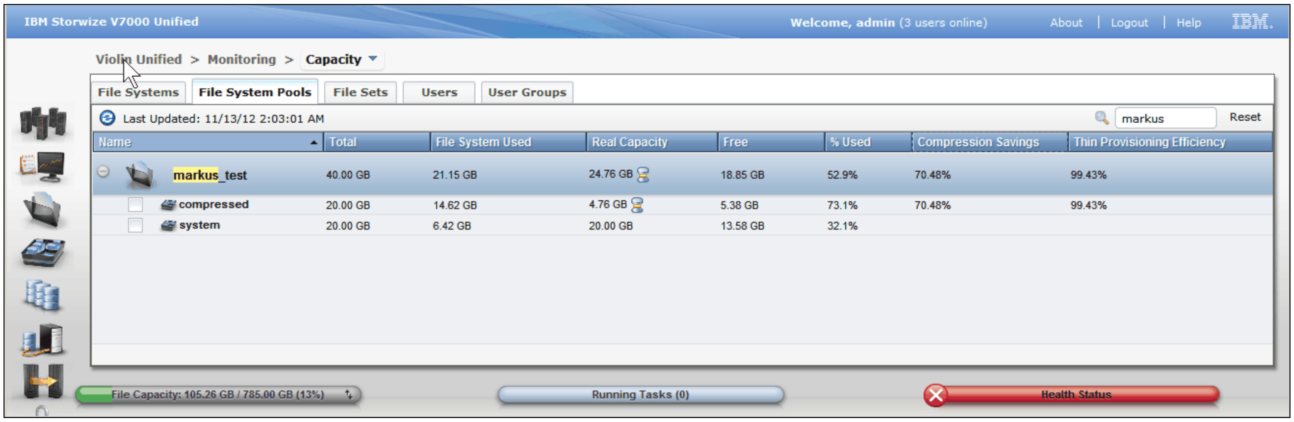
The average compression rate can be gathered by analyzing external storage systems with the NAS Compression Estimation Utility, as previously explained. The final step to predict the correct file system pool sizes is to create a small file system with two pools and the appropriate placement rules. This file system can then be populated with sample data, and the Storwize V7000 Unified system reporting provides the nformation that is required to plan for the final file system.

Note: When you follow this procedure, do not delete files from the sample file system before the estimation is complete.

The File Systems Pools window, shown in Figure 16-26 on page 320 (**Monitoring → Capacity → File System Pools**), provides the metrics that are necessary for planning:

- The *File System Used* capacity provides the information about how much uncompressed data was placed into the two pools by the placement policy. Use the ratio of the two values for the compressed and system pool to scale the file system pool sizes in the New File System creation dialog window.
- The *Real Capacity* value provides the information about how much capacity on real disk was allocated. For the uncompressed pool, which is fully provisioned, this value is identical to the file system pool size. However, for the compressed pool, this value, along with the File System Used value, indicates how much real data on the disk was used for the compressed pool. The compression savings metric ignores zeros, which are not counted. Therefore, the ratio of real capacity to file system that is used provides a value that determines how much space is required relative to the file system size.

- The *Used* block volume capacity accounts only for disk capacity that is used by data already. The Real Capacity block volume value accounts for the already allocated blocks on disk, which belong to the volume. The Used capacity is displayed only per volume, not per file system pool, and can be seen in the **File** → **File Systems** grid by viewing the Network Shared Disk (NSD) properties. Because the *Used* capacity is smaller than the *Real* Capacity, it provides more accurate and even better reduction rates. The `rsize` value, which determines the real capacity that is allocated for a given used capacity, is set to 2% for volumes that are created in the GUI. Therefore, the difference is small. For practical sizing purposes, the Real Capacity value is good enough. It provides for a small contingency, unless the `rsize` was updated through the CLI and to a much higher value.


IBM Storwize V7000 Unified
Welcome, admin (3 users online)
About | Logout | Help
IBM.
Violin Unified > Monitoring > Capacity
File Systems
File System Pools
File Sets
Users
User Groups
Reset
File Capacity: 105.26 GB / 785.00 GB (13%)
Running Tasks (0)
Health Status

Last Updated: 11/13/12 2:03:01 AM

markus

Name	Total	File System Used	Real Capacity	Free	% Used	Compression Savings	Thin Provisioning Efficiency
markus_test	40.00 GB	21.15 GB	24.76 GB	18.85 GB	52.9%	70.48%	99.43%
compressed	20.00 GB	14.62 GB	4.76 GB	5.38 GB	73.1%	70.48%	99.43%
system	20.00 GB	6.42 GB	20.00 GB	13.58 GB	32.1%		

Figure 16-26 The File System Pools view helps with capacity planning, based on a data sample

Example 16-11 shows the calculation that is based on Figure 16-26.

Example 16-11 Calculation

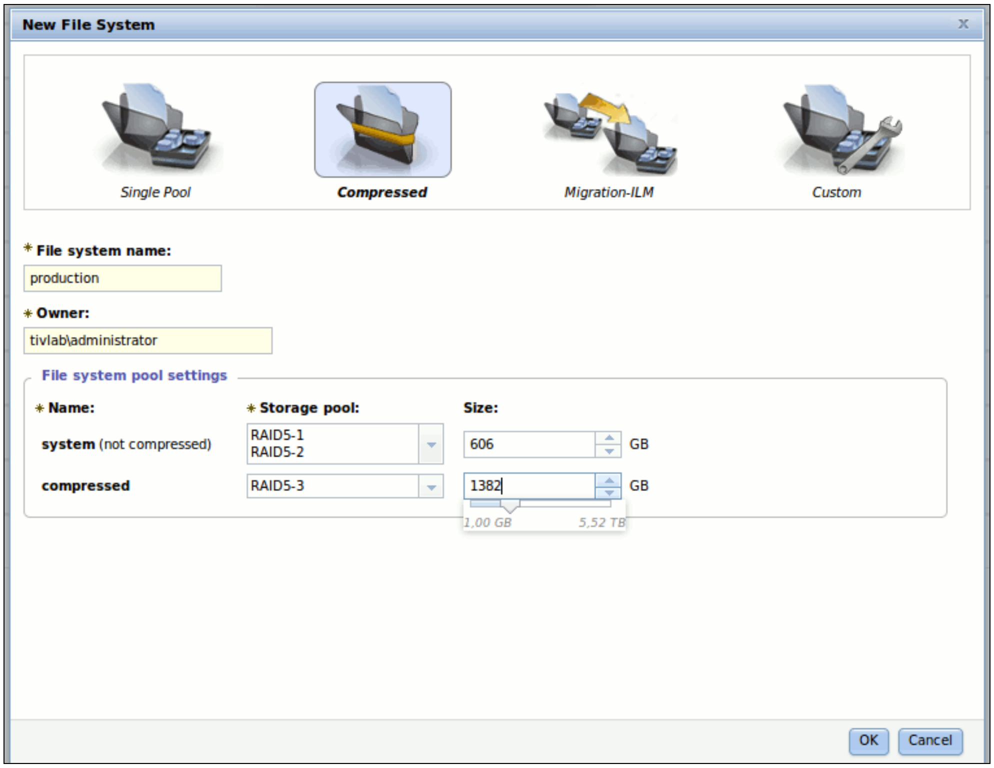
```
Planned file system size: 2000 GB
Sample file system overall used capacity: 21.15 GB

'System' file system pool:
Sample 'system' pool used capacity: 6.42 GB
File system pool capacity fraction: 6.42 GB/21.15 GB = 30.3 %
Planned 'system' file system pool capacity: 2 TB * 30.3 % = 606 GB
Planned Physical disk for the system pool: 606 GB

'Compressed' file system pool:
Sample file system compressed pool used capacity: 14.62 GB
File System pool capacity fraction = 14.62 GB/21.15 GB = 69.1 %
Planned 'compressed' file system pool capacity: 2 TB * 69.1 % = 1382 GB

Sample compressed pool Real Capacity: 4.76 GB
Reduction ratio (compression and thin provisioning): 4.76 GB/14.62 GB = 32.5 %
Physical disk for the compressed pool: 2000 GB * 69.1 % * 32.5 % = 449 GB
```

With the data as calculated in the example, the file system can be configured as shown in Figure 16-27 on page 321. The expected uncompressed file system pool capacities are entered in the pool size entry fields. For the fully provisioned MDisk group, the GUI checks whether the capacity requirements are sufficient. For the compressed pool, the GUI allows you to create the compressed file system pool if at least 25% of the entered uncompressed file size is available. This assumes a compression savings of 75%, which is too optimistic for most cases. Therefore, it is best to ensure that available space in the MDisk group that hosts the compressed pool is at least as large as the estimated physical disk capacity of the compressed pool. In this example, the estimate is 449 GB.



* File system name: production

* Owner: tivlab\administrator

File system pool settings

* Name:	* Storage pool:	Size:
system (not compressed)	RAID5-1 RAID5-2	606 GB
compressed	RAID5-3	1382 GB

1,00 GB 5,52 TB

Figure 16-27 The evaluated file system pool sizes are entered to create a file system

Notice the slider below the compressed file system pool size entry box. The slider maximum value allows you to configure four times the available physical size of the chosen storage pool.

16.5 Compression metrics for file systems

The Storwize V7000 Unified system provides several metrics that are specific to file systems and compression. The compression feature introduces the *thin provisioning* concept, so viewing these metrics provides insight into both the compression aspects and thin provisioning that is related to file system use data.

Figure 16-28 on page 322 compares the perspectives of the file system, the block device thin provisioning, and the compression engine on a file system with a few files.

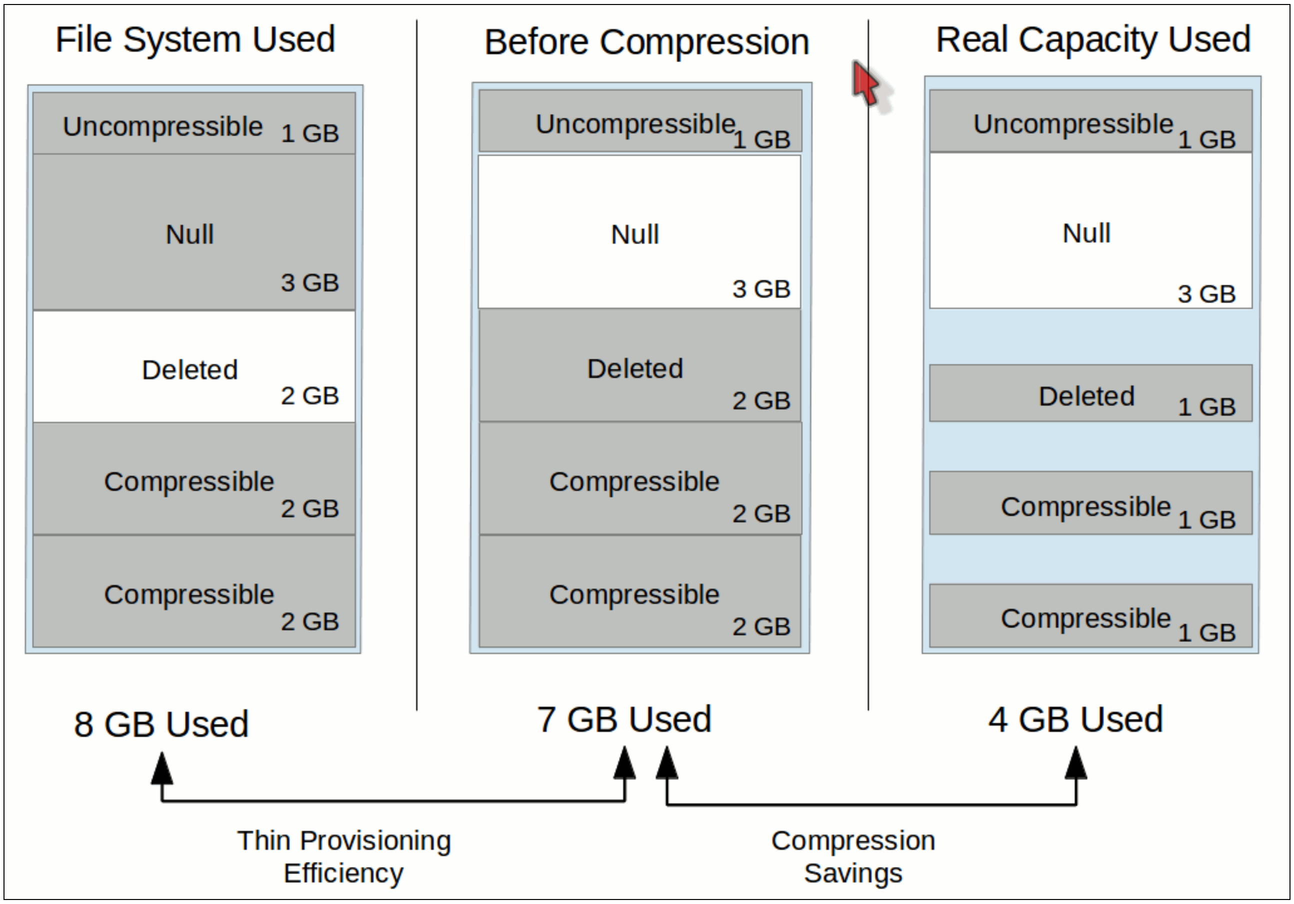


Figure 16-28 Capacity views from the file system, block device, and compression engine perspective

The file system calculations consider the uncompressed files and snapshots, clones, and metadata that it handles. The compressibility of files and the fact that files might contain zeros is not accounted for by the file system.

The block storage device layer handles zeros inside of files in a space-efficient way, because zeros are not written to disk and not accounted for, otherwise. The thin provisioning approach assumes that all blocks contain zeros if they are not specifically written with non-zero data. Therefore, files that contain zeros, such as virtual machine images or empty database containers, are not counted on the block device.

The Thin Provisioning Efficiency option that is shown on the Storwize V7000 Unified GUI in the **Monitor** → **Capacity** → **File Systems by Pools** view shows, by file system pool, the ratio of the file system used capacity, divided by the block device uncompressed used capacity. The following examples provide scenarios with expected ratios and potential user actions:

- A file system that is used as an archive, where no files were deleted yet and no files contain zeros, shows a thin provisioning efficiency of 100%.
- A file system that has just been created and filled with newly formatted database container files shows a thin provisioning efficiency of 100%. The file system counts the file size, but the block devices save space on the zeros, which are also not counted. Theoretically, the ratio between file system-used capacity and before-compression capacity is even higher than 100%, but the efficiency value is limited to 100%.

- A file system much larger than its current file system used capacity is chosen. Files are regularly created and deleted. The thin provisioning capacity is much lower than 100% because the block device still accounts for the files that were deleted. Potentially, the file system chosen is too large and the file system size must be reduced to save space. Similar to a fully provisioned file system, to avoid wasted space, the compressed pools of a file system must not be oversized.

16.6 Managing compressed file systems

In the following sections, we describe managing compressed file systems.

16.6.1 Adding a compressed pool to a file system

When a file system that is not using compression is to be converted to a file system that uses compression, several steps are necessary after analysis of the data in the not-compressed file system. This provides insight regarding whether the file system is a good candidate for compression. See *16.4.1, "Planning capacity with the NAS Compression Estimation Utility" on page 316* in this chapter for approaches to analyze data.

Create a compressed pool

Another pool that has the compression flag that is activated is created by using the GUI edit dialog (**Files** → **File Systems**). The pool sizes need to match the expected uncompressed capacity. It is not yet necessary to define a placement policy

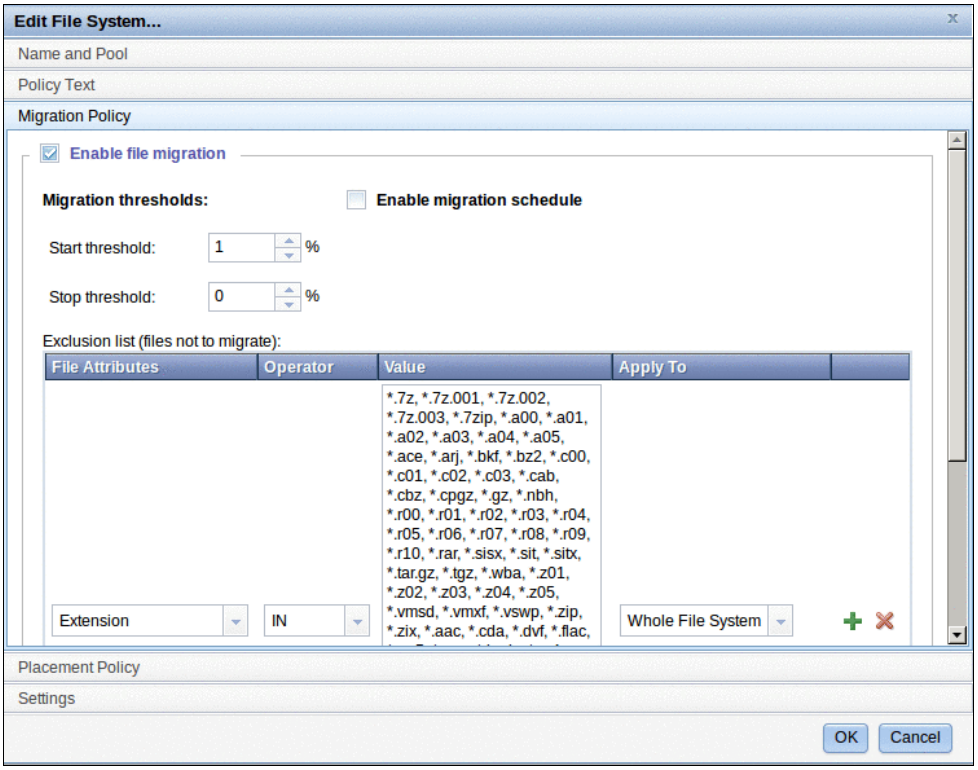
Create and run a migration policy

We suggest that you the files in the system pool that can be compressed. To do this, you need to define and then execute a migration policy that moves the compressible files to the compressed pool.

The screen capture in Figure 16-29 on page 324 shows how to define a migration policy. The default list of non-=compressible file types or an adapted version that matches the system environment (see Example 16-1 on page 296) can be entered in the exclusion rule for this migration policy. In this example, the policy is executed automatically because of the low thresholds, with an upper threshold of 1% and a lower threshold of 0.

The GUI policy editor was used to create the migration policy because it nicely formats the long list of extensions to exclude from migration into the correct policy language statements. Alternatively, the CLI can be used to create such a policy (that method is more cumbersome) and to run the policy by using the **`runpolicy`** command, which executes the migration policy only once and at the time that the command starts to run. The creation and CLI-based execution of a migration run is described in 16.6.2, "Making a selectively compressed file system uncompressed" on page 326.

Figure 16-29 on page 324 shows the GUI migration policy editor, which can be used to define which files should not be migrated to the compressed pool.



☑ Enable file migration

Migration thresholds: ☐ Enable migration schedule

Start threshold: 1 %

Stop threshold: 0 %

Exclusion list (files not to migrate):

File Attributes	Operator	Value	Apply To
Extension	IN	*.7z, *.7z.001, *.7z.002, *.7z.003, *.7zip, *.a00, *.a01, *.a02, *.a03, *.a04, *.a05, *.ace, *.arj, *.bkf, *.bz2, *.c00, *.c01, *.c02, *.c03, *.cab, *.cbz, *.cpgz, *.gz, *.nbh, *.r00, *.r01, *.r02, *.r03, *.r04, *.r05, *.r06, *.r07, *.r08, *.r09, *.r10, *.rar, *.sisx, *.sit, *.sitx, *.tar.gz, *.tgz, *.wba, *.z01, *.z02, *.z03, *.z04, *.z05, *.vmsd, *.vmxf, *.vswp, *.zip, *.zix, *.aac, *.cda, *.dvf, *.flac,	Whole File System

Figure 16-29 GUI migration policy editor

Because the threshold for the migration was set low, the system starts the migration job automatically. You can view the active jobs by using the **`lsjobstatus`** CLI command and follow migration job results by using the **`showlog`** command. See Example 16-12.

Example 16-12 View the migration policy and monitor the migration job by using CLI commands

```
# ******* View the GUI generated policy rule for migration
# ******* The exclusion list is truncated in the example
[7802378.ibm]$ lspolicy -D converttocompressed
RULE 'generatedMigrationRule0'
  MIGRATE
    FROM POOL 'system'
    THRESHOLD(1,0)
  TO POOL 'compressed'
    WHERE NOT (
      (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001'
... OR LOWER(NAME) LIKE '%.zip' OR LOWER(NAME) OR LOWER(NAME) LIKE '%.docx'))
RULE 'default' SET POOL 'system'
EFSSG1000I The command completed successfully.

# ******* List current active jobs, for all jobs used the --all keyword
[7802378.ibm]$ lsjobstatus
File system         Job            Job id Status  Start time              End
time/Progress RC Message
converttocompressed auto_migration 42    running 11/13/12 3:19:45 PM IST
EFSSG1000I The command completed successfully.

# ******* Show the detailed log of the job, the log can already be viewed
# ******* while it is still running. Log truncated in this example
```

```
[7802378.ibm]$ showlog 42
Primary node: mgmt002st001
Job ID : 42
[I] GPFS Current Data Pool Utilization in KB and %
compressed69121048576000.006592%
system79142401048576007.547607%
[I] 238008 of 12336640 inodes used: 1.929277%.
[I] Loaded policy rules from /var/mmfs/tmp/tspolicyFile.mmapplypolicy.943899.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP =
2012-11-13@13:19:47 UTC
parsed 1 Placement Rules, 0 Restore Rules, 1 Migrate/Delete/Exclude Rules,
   0 List Rules, 0 External Pool/List Rules
RULE 'generatedMigrationRule0'
  MIGRATE
    FROM POOL 'system'
    THRESHOLD(1,0)
  TO POOL 'compressed'
    WHERE NOT (
      (LOWER(NAME) LIKE '%.7z' OR LOWER(NAME) LIKE '%.7z.001' OR LOWER(NAME) LIKE
'%.hqx' OR LOWER(NAME) LIKE '%.docx'))
RULE 'default' SET POOL 'system'
[I] Directories scan: 223268 files, 10751 directories, 0 other objects, 0
'skipped' files and/or errors.
[I] Summary of Rule Applicability and File Choices:
 Rule#Hit_CntKB_HitChosenKB_ChosenKB_IllRule
  013464649087601346464908760ORULE 'generatedMigrationRule0' MIGRATE FROM POOL
'system' THRESHOLD(1,0) TO POOL 'compressed' WHERE(.)
[I] Filesystem objects with no applicable rules: 99362.
[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 4908760KB: 134646 of 134646 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 0KB: 0 of 0 candidates;
 Chose to list 0KB: 0 of 0 candidates;
 0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
compressed49156721048576004.687950%
system30234001048576002.883339%
[I] A total of 134646 files have been migrated, deleted or processed by an
EXTERNAL EXEC/script;
   11 'skipped' files and/or errors.
-----------------------------------------------------------
End of log - auto_migration job ended
-----------------------------------------------------------

EFSSG1000I The command completed successfully.
```

Remove the migration policy and enable a placement policy

The migration policy can now be disabled by clearing the placement, and a placement policy can be defined. For a compressed file system pool, a placement policy can be created in the same way that the file system was created with a compressed pool. The placement policy ensures that only compressible files are added to the new pool in the future.

16.6.2 Making a selectively compressed file system uncompressed

It is possible to revert a selective compressed file system to uncompressed while there is full data access by following these steps:

1. Analyze how much physical space is needed to move all data to the system pool.

   It is easiest to do the conversion when the uncompressed space needed is available as free disk space, in addition to the compressed pool. If not, you must take a step-by-step approach and remove volumes from the compressed pool that is emptied.
2. Create a default placement rule for the file system. See Example 16-13.

   Example 16-13 Default policy text

   ```
   RULE 'default' SET POOL 'system';
   ```
3. Create and run a migration rule that moves data to the system pool.

   To migrate data back to the system pool, create a simple policy rule that moves all data to the system pool from the compressed pool. If the policy is created in the GUI policy text editor, the migration policy can be applied to a file system to ensure that the migration starts automatically and is executed regularly. Because the migration rule must run only once, another approach is chosen. The CLI must be used to create a policy, start a policy run, and monitor the migration of data.

   Example 16-14 shows the simple policy rule.

   Example 16-14 Simple policy rule

   ```
   # Create a policy rule - this policy is not applied to any file system
   [7802378.ibm]# mkpolicy migratetosystem -R "RULE 'migratetosystem' migrate from
   pool 'compressed' to pool 'system';"
   EFSSG1000I The command completed successfully.

   # The runpolicy command will only run the policy once on our file system
   [7802378.ibm]$  runpolicy markus_test -P migratetosystem
   EFSSA0184I The policy is started on markus_test with JobID 806.
   # The showlog command provides the output of policy runs
   [7802378.ibm]$ showlog 806
   Primary node: mgmt001st001
   Job ID : 806
   [I] GPFS Current Data Pool Utilization in KB and %
   compressed 6531584 39845888 16.392116%
   system 20494848 49283072 41.585979%
   [I] 1000507 of 1200640 inodes used: 83.331140%.
   [I] Loaded policy rules from
   /var/opt/IBM/sofs/PolicyFiles/policy4252244384126468096.
   Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP =
   2012-11-15@07:39:05 UTC
   parsed 0 Placement Rules, 0 Restore Rules, 1 Migrate/Delete/Exclude Rules,
      0 List Rules, 0 External Pool/List Rules
   RULE 'migratetosystem' migrate from pool 'compressed' to pool 'system'
   [I] Directories scan: 923603 files, 72924 directories, 0 other objects, 0
   'skipped' files and/or errors.
   [I] Summary of Rule Applicability and File Choices:
    Rule# Hit_Cnt KB_Hit ChosenKB_Chosen nkB_Ill Rule
     0574246 49452407 424649452400RULE 'migratetosystem' MIGRATE FROM POOL
   'compressed' TO POOL 'system'
   [I] Filesystem objects with no applicable rules: 422280.
   ```

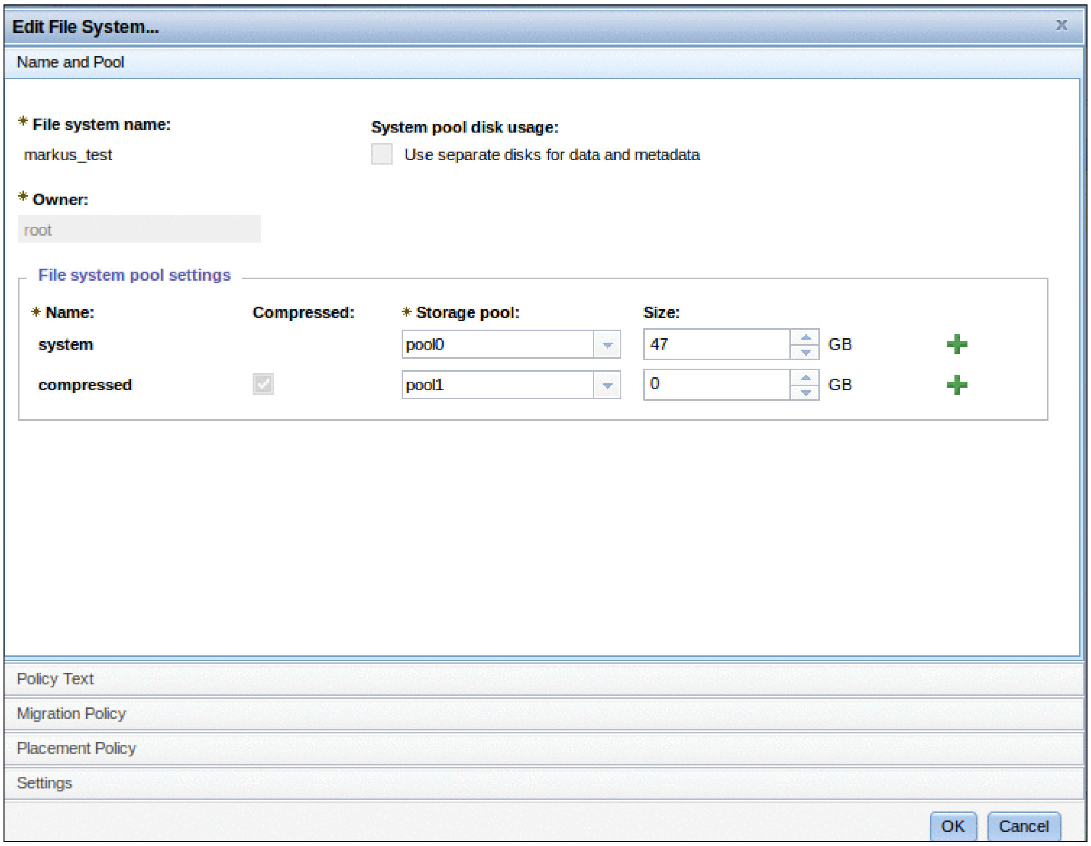
```
[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 4945240KB: 574246 of 574246 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 0KB: 0 of 0 candidates;
 Chose to list 0KB: 0 of 0 candidates;
 0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
compressed1586344398458883.981199%
system254728564928307251.686827%
[I] A total of 574246 files have been migrated, deleted or processed by an
EXTERNAL EXEC/script;
   0 'skipped' files and/or errors.
-----------------------------------------------------------
End of log - runpolicy job ended
-----------------------------------------------------------

EFSSG1000I The command completed successfully.
```

4. Remove all volumes of the compressed pool

   After the policy run is completed, all data is moved to the system pool and the volumes of the compressed pool can be removed. You can use the Edit File System dialog window in the GUI to perform this task. Setting the capacity of the compressed pool to zero removes all compressed volumes and removes the file system pool.

   Figure 16-30 shows how to remove the compressed volumes.



Edit File System...

Name and Pool

* File system name: markus_test

System pool disk usage:
□ Use separate disks for data and metadata

* Owner: root

File system pool settings

* Name:	Compressed:	* Storage pool:	Size:	
system		pool0	47 GB	
compressed	☑	pool1	0 GB	

Figure 16-30 The compressed file system pool size is set to 0, which triggers the removal of all related compressed volumes

16.7 Compression storage space savings reporting

Understanding capacity in a virtualized storage system can be a challenge. Probably the best approach is to review the various layers for storing data, from different points of view:

Data: The data stored by the host operating system onto the volume.

Shares: A logical unit that the user writes the data into

File system: Logical unit in the Storwize V7000 Unified system that is created from file system pools

Pools: Composed from MDisks

File system pool: The storage container of compressed volumes

By design, the compression technology that is implemented in the Storwize V7000 Unified system is visible to the host. It compresses the client data before writing it to the disk and extracts the data as it is read by the host. The data size looks different from a different point of view. For example, the compressed size is not reflected to the user and can be seen only from the Storwize V7000 Unified system point of view.

16.7.1 Reporting basic overview

The following are basic concepts of the Storwize V7000 Unified system's reporting of compressed data:

Real capacity Real capacity is the storage capacity that is allocated to a volume copy from a storage pool. A compressed volume is, by default, a thin-provisioned volume that you can use to allocate space on demand. When a new volume is created, there is an option to define the size that creates as a percentage of the original volume size. By default, it is 2%. The volume expands automatically according to use.

Used capacity The amount of real size that is used to store data for the volume, which is sometimes called *compressed size*.

Before compression size The size of all of the data that was written to the volume, calculated as though it was written without compression. This size is reflected on the host operating system.

Virtual capacity Virtual capacity is the volume storage capacity that is available to a host. In a fully allocated volume, the virtual capacity and real capacity are the same. In a thin-provisioned or compressed volume, however, the virtual capacity can be much larger than the real capacity.

Compression Savings The ratio between the compressed and uncompressed size.

16.7.2 Reporting of compression in the GUI

Reporting of compressed data is available in the GUI in the following windows:

- **System level:** Monitoring → System
- **File system level:** Monitoring → Capacity and in files and file systems
- **Storage pool level:** Pools → MDisks by pools
- **Volume level:** Volumes by pool or file system page

System level

To view the file system level, select **Monitoring** → **System**.

Figure 16-31 shows three areas of the system disks:

- Compressed size
- Before compression size
- Virtual capacity

As you can see in Figure 16-31, by looking at the tube to the left of the Storwize V7000 Unified system, the compression ratio is approximately 50%. The second area of the *capacity before compression* is greater than the lower section of the *compressed capacity*.

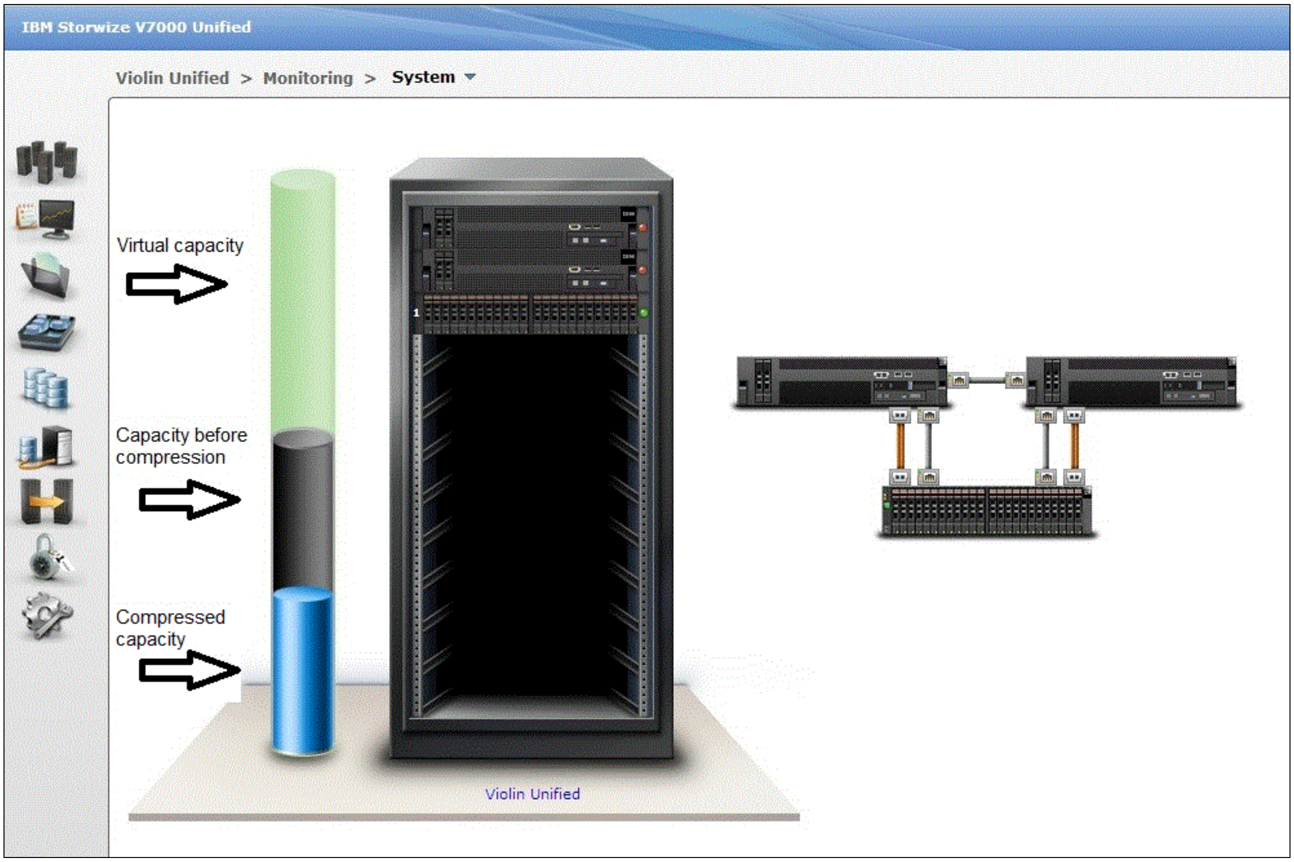

IBM Storwize V7000 Unified
Violin Unified > Monitoring > System
Virtual capacity
Capacity before compression
Compressed capacity
Violin Unified

Figure 16-31 Areas of the system disks

Note: The reporting view from the IBM Storwize V7000 Unified system is different from the reporting from the Storwize V7000. For more information about Storwize V7000 reporting, see *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859:

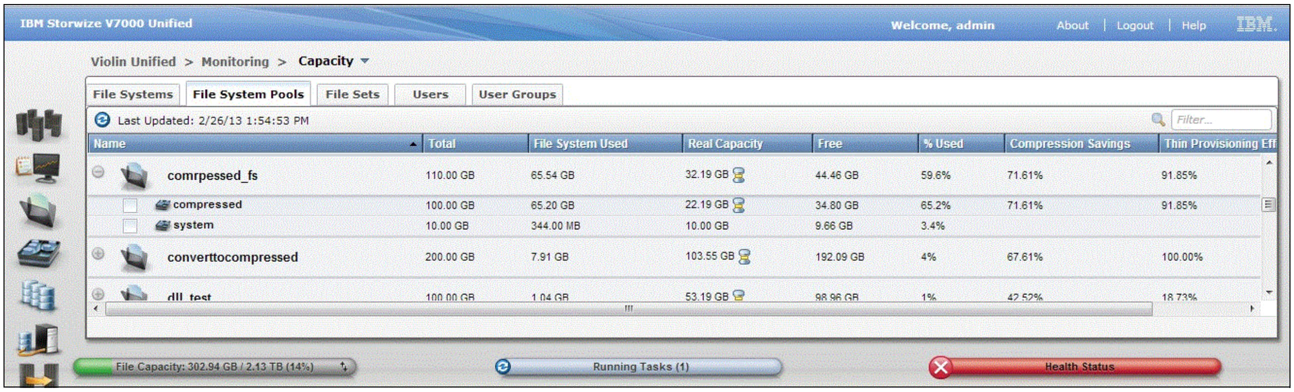
`http://www.redbooks.ibm.com/Redbooks.nsf/RedpieceAbstracts/redp4859.html?Open`

File system level

To get information at the file system level, click the **Monitoring** icon and choose the **Capacity** option. Choose the **File System Pools** tab, and you can see the compression savings of each file system and the amount of used space from each file system pool.

In Figure 16-32, you can see the real capacity, file system used capacity, and the compression savings capacity of each file system. By clicking the plus sign to expand the view, you can see it by file system pool.

In this example, in the `compressed_fs` file system, you can see that the *real capacity* of the system and the *total capacity* are equal. This is because an uncompressed file system is fully allocated. But these values of the compressed pool are different. The real capacity is the allocated size from the pool, which is 22.19 GB in this case.



Last Updated: 2/26/13 1:54:53 PM

Name	Total	File System Used	Real Capacity	Free	% Used	Compression Savings	Thin Provisioning Eff
comrpessed_fs	110.00 GB	65.54 GB	32.19 GB	44.46 GB	59.6%	71.61%	91.85%
compressed	100.00 GB	65.20 GB	22.19 GB	34.80 GB	65.2%	71.61%	91.85%
system	10.00 GB	344.00 MB	10.00 GB	9.66 GB	3.4%		
converttocompressed	200.00 GB	7.91 GB	103.55 GB	192.09 GB	4%	67.61%	100.00%
dll_test	100.00 GB	1.04 GB	53.19 GB	98.96 GB	1%	42.52%	18.73%

Figure 16-32 Real capacity, file system used capacity, and the compression savings capacity of each file system

Storage pool level

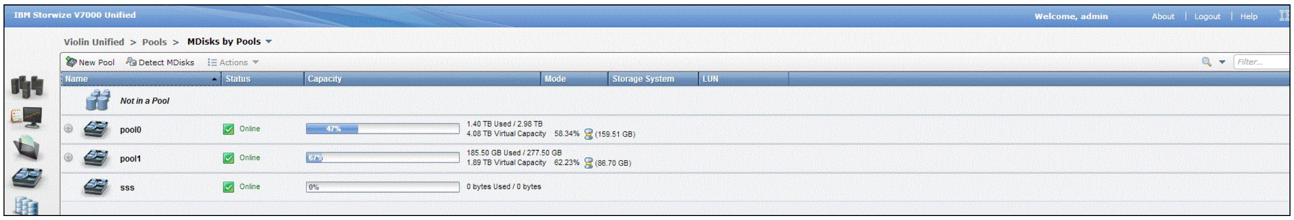
To get information at the storage pool level, click the Pools icon and select **MDisks by pool** from the drop-down menu. This provides the compression ratio benefits of the entire pool.

> **Note:** The pool contains volumes that are used by the file system of the IBM Storwize V7000 Unified system and VDisks of the Storwize V7000 block devices.

Figure 16-33 shows the amount of used capacity and the virtual capacity.

Used size: The percentages in the shaded blue part are the amount of used space from the allocated capacity. In this example, 1 TB and 2.98 TB are used, which means that 47% of `pool0` is used.

Virtual capacity: The numbers also present the virtual capacity. This number is the sum of all volume sizes in the pool. In this example, `pool0` has 4.08 TB of virtual capacity.



Name	Status	Capacity	Mode	Storage System	LUN
Not in a Pool					
pool0	☑ Online	47% 1.40 TB Used / 2.98 TB 4.08 TB Virtual Capacity 58.34% (159.51 GB)			
pool1	☑ Online	67% 185.50 GB Used / 277.50 GB 1.89 TB Virtual Capacity 62.23% (86.70 GB)			
sss	☑ Online	0% 0 bytes Used / 0 bytes			

Figure 16-33 Amount of used capacity and the virtual capacity

> **Note:** The file system consists of volumes. Therefore, this view includes the file system volumes and the block device volumes that are using the same storage pool.

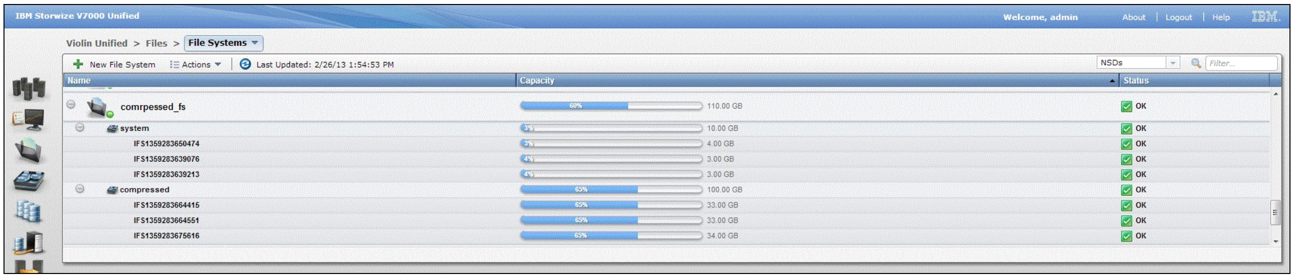
Volume level

There are two methods to report at a volume level.

Method 1: Click **Files** → **File systems**.

Expand the file system and the file system pools to see the volumes that make up each of them. The screen capture in Figure 16-34 shows the used capacity of each volume in the file system per file system pool.

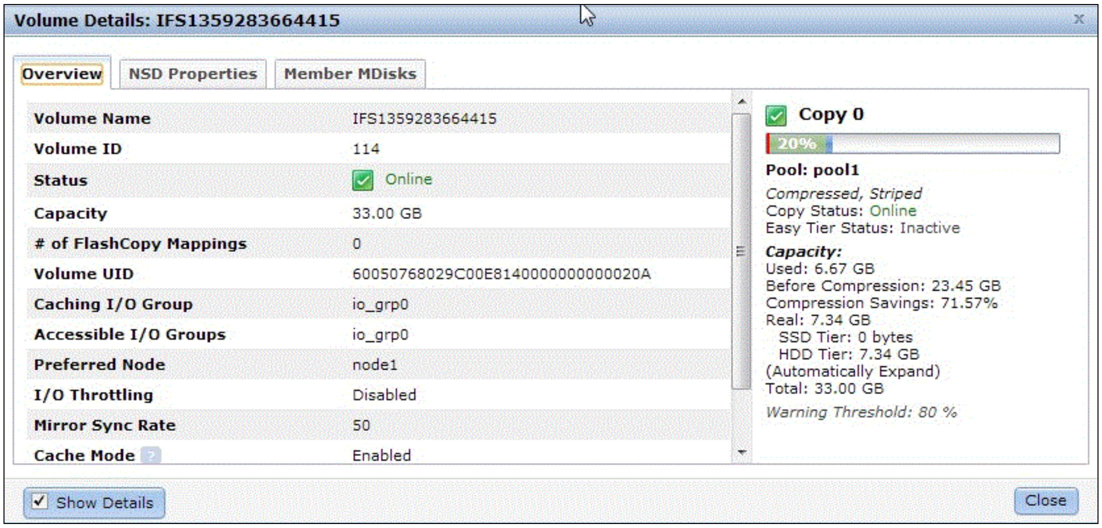
To see more details, right-click a volume and select **Properties**.


IBM Storwize V7000 Unified
Welcome, admin
About | Logout | Help
IBM.
Violin Unified > Files > File Systems
New File System
Actions
Last Updated: 2/26/13 1:54:53 PM
NSDs
Filter...

Name	Capacity		Status
comrpessed_fs	60%	110.00 GB	OK
system	3%	10.00 GB	OK
IFS1359283650474	3%	4.00 GB	OK
IFS1359283639076	4%	3.00 GB	OK
IFS1359283639213	4%	3.00 GB	OK
compressed	65%	100.00 GB	OK
IFS1359283664415	65%	33.00 GB	OK
IFS1359283664551	65%	33.00 GB	OK
IFS1359283675616	65%	34.00 GB	OK

Figure 16-34 Volumes used capacity

In Figure 16-35, you can see the before-compression size, real size, and the compression savings of the volume. In this example, the original data size is 23.45 GB, which was compressed to 6.67 GB. Therefore, the compression savings is 71.57%.



Volume Name	IFS1359283664415
Volume ID	114
Status	✓ Online
Capacity	33.00 GB
# of FlashCopy Mappings	0
Volume UID	60050768029C00E8140000000000020A
Caching I/O Group	io_grp0
Accessible I/O Groups	io_grp0
Preferred Node	node1
I/O Throttling	Disabled
Mirror Sync Rate	50
Cache Mode ?	Enabled

✓ Copy 0
20%
Pool: pool1
Compressed, Striped
Copy Status: Online
Easy Tier Status: Inactive
Capacity:
Used: 6.67 GB
Before Compression: 23.45 GB
Compression Savings: 71.57%
Real: 7.34 GB
SSD Tier: 0 bytes
HDD Tier: 7.34 GB
(Automatically Expand)
Total: 33.00 GB
Warning Threshold: 80 %

☑ Show Details

Figure 16-35 Before compression size, real size, and the space savings of the volume

Ensure that the **Show Details** check box is checked.

> **Note:** The compressed data is divided equally between the volumes in the same file system pool, so the compression savings shown in one volume are reflected in the entire file system pool.

Method 2: Click the **Pools** icon and select **Volumes by pool**. The table provides the compression savings for each volume in a specific pool. This view is more useful for block device volumes than for file systems.

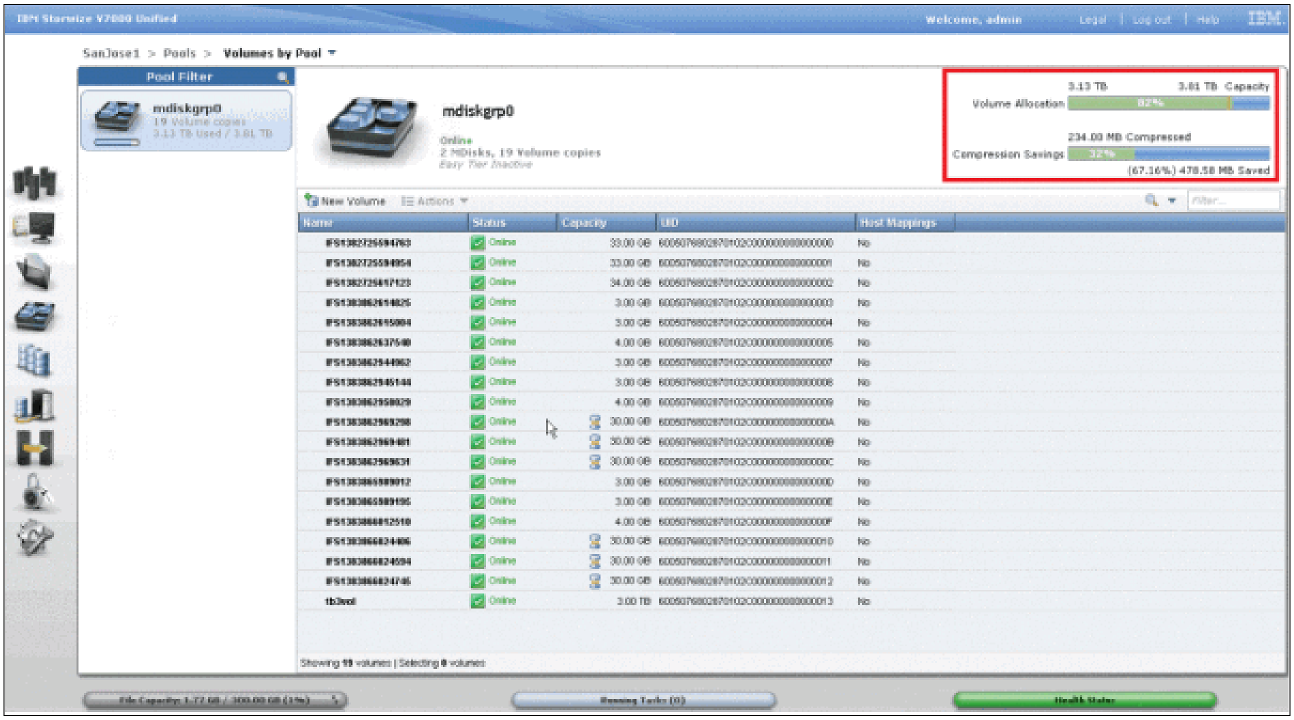
In the graph at the upper-right side of Figure 16-36, you can see the volume capacity, 3.13 TB, and the virtual capacity, 3.81 TB. The volumes that are configured in this pool are over-allocated. Their total size is 3.81 TB. The second bar reports the compression savings. The original size of the data is the compressed size + saved size - 234.00 MB + 478.58 MB = 712.58 MB. The compressed size is 234.00 MB. The compression ratio of the entire storage pool mdiskgrp0 is 32%, which is calculated with by this formula:

Compressed size ÷ total size

For this example:

234.000 ÷ 712.58 = 32%

The savings ratio is ~67%.



Name	Status	Capacity	UID	Host Mappings
IFS1382725594763	Online	33.00 GB	6005076802870102C000000000000000	No
IFS1382725594954	Online	33.00 GB	6005076802870102C000000000000001	No
IFS1382725617123	Online	34.00 GB	6005076802870102C000000000000002	No
IFS1383862614825	Online	3.00 GB	6005076802870102C000000000000003	No
IFS1383862615004	Online	3.00 GB	6005076802870102C000000000000004	No
IFS1383862637540	Online	4.00 GB	6005076802870102C000000000000005	No
IFS1383862944962	Online	3.00 GB	6005076802870102C000000000000007	No
IFS1383862945144	Online	3.00 GB	6005076802870102C000000000000008	No
IFS1383862958029	Online	4.00 GB	6005076802870102C000000000000009	No
IFS1383862969298	Online	30.00 GB	6005076802870102C00000000000000A	No
IFS1383862969481	Online	30.00 GB	6005076802870102C00000000000000B	No
IFS1383862969631	Online	30.00 GB	6005076802870102C00000000000000C	No
IFS1383865989012	Online	3.00 GB	6005076802870102C00000000000000D	No
IFS1383865989195	Online	3.00 GB	6005076802870102C00000000000000E	No
IFS1383866012510	Online	4.00 GB	6005076802870102C00000000000000F	No
IFS1383866824406	Online	30.00 GB	6005076802870102C000000000000010	No
IFS1383866824594	Online	30.00 GB	6005076802870102C000000000000011	No
IFS1383866824746	Online	30.00 GB	6005076802870102C000000000000012	No
tb3vol	Online	3.00 TB	6005076802870102C000000000000013	No

Showing 19 volumes | Selecting 0 volumes

Figure 16-36 Volume capacity of 3.13 TB and virtual capacity of 3.81 TB

Consideration: The reported numbers are dynamically updated as data is sent from the Storwize V7000 cache to the receive-any control element (RACE) component. This updating causes a slight delay between the host writes and the updated reporting (typically only a few seconds).

16.7.3 Compression reporting by using the CLI

The CLI commands provide information in a volume and storage pool level. Therefore, to see the capacity reporting of a file system through the CLI, you need to know the VDisks name that makes up the file system.

You can generate reports on compressed data by using the following CLI commands:

System level: `lssystem`

Specific storage pool:	**`lsmdiskgrp [<mdisk_grp_name>]`**
All storage pools:	**`lsmdiskgrp`**
Specific volume:	**`lssevdiskcopy [<vdisk_name>]`**
All volumes in a pool:	**`lssevdiskcopy -filtervalue mdisk_grp_name=[<mdisk_grp_name>]`**
All volumes in system:	**`lssevdiskcopy`**

Example 16-15 shows the output of the **`lssystem`** command.

Example 16-15 lssystem output

```
[7802378.ibm]$ lssystem
id 00000200A7003A05
name Violin Unified
location local
partnership
bandwidth
total_mdisk_capacity 3.3TB
space_in_mdisk_grps 3.3TB
space_allocated_to_vdisks 1.57TB
total_free_space 1.7TB
total_vdiskcopy_capacity 5.97TB
total_used_capacity 1.48TB
total_overallocation 183
total_vdisk_capacity 5.95TB
total_allocated_extent_capacity 1.58TB
[...]
compression_active yes
compression_virtual_capacity 4.58TB
compression_compressed_capacity 166.52GB
compression_uncompressed_capacity 412.73GB
[...]
```

Example 16-16 shows the output of the **`lsmdiskgrp`** command for a specific pool.

Example 16-16 lsmdiskgrp output

```
[7802378.ibm]$ lsmdiskgrp pool0
id 0
name pool0
status online
mdisk_count 2
vdisk_count 129
capacity 2.98TB
extent_size 256
free_capacity 1.58TB
virtual_capacity 4.08TB
used_capacity 1.34TB
real_capacity 1.40TB
overallocation 136
warning 80
easy_tier auto
easy_tier_status inactive
tier generic_ssd
tier_mdisk_count 0
```

```
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier generic_hdd
tier_mdisk_count 2
tier_capacity 2.98TB
tier_free_capacity 1.58TB
compression_active yes
compression_virtual_capacity 2.78TB
compression_compressed_capacity 113.89GB
compression_uncompressed_capacity 273.40GB
```

Example 16-17 shows the output of the **`lsmdiskgrp`** command for the entire node.

Example 16-17   lsmdiskgrp output for entire node

```
[7802378.ibm]$ lsmdiskgrp
0  pool0 online 2 129 2.98TB  256 1.58TB  4.08TB  1.34TB   1.40TB   136  80  auto  inactive  yes  2.78TB 113.89GB
```

Example 16-18 shows the **`lssevdiskcopy`** output.

Example 16-18   lssevdiskcopy -nohdr output

```
[7802378.ibm]$ lssevdiskcopy -nohdr -filtervalue mdisk_grp_name=pool0
0   Nitzan                    0 0 pool0 400.00MB 4.63MB   24.00MB  19.38MB  1666 on 80     no  yes 4.44MB
1   CSB-SYSTEM_AG_COMPRESSED 0 0 pool0 50.00GB  21.24GB  22.25GB  1.02GB   224  on 80     no  yes 49.96GB
3   bosmatTest               0 0 pool0 100.00GB 15.10GB  17.11GB  2.01GB   584  on 80     no  yes 28.04GB
12  VM_Compressed            0 0 pool0 34.00GB  7.09GB   7.78GB   711.68MB 436  on 80     no  yes 15.56GB
18  export for esx 200       0 0 pool0 100.00GB 3.25MB   2.02GB   2.01GB   4961 on 80     no  yes 0.00MB
19  New_VM_Comp              0 0 pool0 300.00GB 22.12GB  28.13GB  6.01GB   1066 on 80     no  yes 48.63GB
20  New_VM_Clear             0 0 pool0 100.00GB 33.96GB  35.96GB  2.00GB   278  on 80 256 yes no  33.96GB
25  Nitzan_01                0 0 pool0 400.00MB 0.75MB   17.25MB  16.50MB  2318 on 80 256 yes no  0.75MB
26  VM_Comp_P0_0             0 0 pool0 150.00GB 5.19GB   8.20GB   3.01GB   1828 on 80     no  yes 11.21GB
27  VM_Comp_P0_1             0 0 pool0 150.00GB 5.19GB   8.20GB   3.01GB   1829 on 80     no  yes 11.21GB
28  VM_Comp_P0_2             0 0 pool0 150.00GB 5.20GB   8.20GB   3.01GB   1829 on 80     no  yes 11.21GB
29  VM_Comp_P0_3             0 0 pool0 150.00GB 5.19GB   8.20GB   3.01GB   1829 on 80     no  yes 11.21GB
30  VM_Comp_P0_4             0 0 pool0 150.00GB 5.19GB   8.20GB   3.01GB   1828 on 80     no  yes 11.21GB
36  Copy Mirror for Pendulum 0 0 pool0 500.00GB 494.69MB 10.48GB  10.00GB  4768 on 80     no  yes 1.31GB
50  IFS1352811948934         0 0 pool0 33.00GB  521.38MB 1.18GB   683.38MB 2804 on 80     no  yes 1.57GB
59  IFS1352811949068         0 0 pool0 33.00GB  513.53MB 1.18GB   691.21MB 2804 on 80     no  yes 1.56GB
60  IFS1352812042137         0 0 pool0 34.00GB  516.28MB 1.20GB   708.93MB 2841 on 80     no  yes 1.55GB
85  lev_production_backup    0 0 pool0 500.00GB 17.49GB  27.50GB  10.02GB  1817 on 80     no  yes 26.83GB
91  IFS1352369292204         0 0 pool0 1.00GB   3.31MB   36.48MB  33.17MB  2806 on 80     no  yes 0.00MB
92  IFS1352369292340         0 0 pool0 1.00GB   3.31MB   36.48MB  33.17MB  2806 on 80     no  yes 0.00MB
93  IFS1352369304265         0 0 pool0 2.00GB   3.31MB   56.96MB  53.65MB  3595 on 80     no  yes 0.00MB
94  jq_vol1                  0 0 pool0 5.00GB   3.25MB   118.40MB 115.15MB 4324 on 80     no  yes 0.00MB
95  jq_vol2                  0 0 pool0 5.00GB   8.19MB   118.40MB 110.21MB 4324 on 80     no  yes 4.94MB
100 IFS1352638421928         0 0 pool0 33.00GB  972.16MB 1.61GB   681.28MB 2043 on 80     no  yes 14.03GB
101 IFS1352638422076         0 0 pool0 33.00GB  972.19MB 1.61GB   681.09MB 2043 on 80     no  yes 14.04GB
102 IFS1352638501197         0 0 pool0 34.00GB  988.22MB 1.65GB   701.51MB 2060 on 80     no  yes 14.27GB
```

For more information about capacity reporting of block device disks, see *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859:

http://www.redbooks.ibm.com/redpapers/pdfs/redp4859.pdf

Related publications

The publications that are listed in this section are considered particularly suitable for more detailed information on the topics that are covered in this book.

IBM Redbooks publications

The following IBM Redbooks publications provide additional information about the topic in this document. Some publications that are cited in this list might be available in softcopy only:

- *Implementing the IBM System Storage SAN Volume Controller V6.3*, SG24-7933
- *Implementing the IBM Storwize V7000 V6.3*, SG24-7938
- *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859
- *Introduction to Storage Area Networks and System Networking*, SG24-5470
- *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116
- *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363
- *IBM System Storage Business Continuity: Part 1 Planning Guide*, SG24-6547
- *IBM System Storage Business Continuity: Part 2 Solutions Guide*, SG24-6548
- *Get More Out of Your SAN with IBM Tivoli Storage Manager*, SG24-6687
- *IBM Tivoli Storage Area Network Manager: A Practical Introduction*, SG24-6848
- *DS8000 Performance Monitoring and Tuning*, SG24-7146
- *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364
- *Using the SVC for Business Continuity*, SG24-7371
- *IBM System Storage SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521
- *IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services*, SG24-7574
- *IBM XIV Storage System Architecture and Implementation*, SG24-7659
- *IBM Tivoli Storage Productivity Center V4.1 Release Guide*, SG24-7725
- *IBM SAN Volume Controller 4.2.1 Cache Partitioning*, REDP-4426

Other publications

These publications are also relevant as further information sources:

- *IBM System Storage SAN Volume Controller: Planning Guide*, GA32-0551
- *IBM System Storage Open Software Family SAN Volume Controller: Planning Guide*, GA22-1052
- *IBM System Storage SAN Volume Controller: Service Guide*, GC26-7901
- *IBM System Storage SAN Volume Controller Model 2145-8A4 Hardware Installation Guide*, GC27-2219
- *IBM System Storage SAN Volume Controller Model 2145-8G4 Hardware Installation Guide*, GC27-2220
- *IBM System Storage SAN Volume Controller Models 2145-8F2 and 2145-8F4 Hardware Installation Guide*, GC27-2221
- *IBM SAN Volume Controller Software Installation and Configuration Guide,* GC27-2286
- *IBM System Storage SAN Volume Controller Command-Line Interface User's Guide*, GC27-2287
- *IBM System Storage Master Console: Installation and User's Guide*, GC30-4090
- *Multipath Subsystem Device Driver User's Guide*, GC52-1309
- *IBM System Storage SAN Volume Controller Model 2145-CF8 Hardware Installation Guide*, GC52-1356
- *IBM System Storage Productivity Center Software Installation and User's Guide*, SC23-8823
- *IBM System Storage Productivity Center Introduction and Planning Guide,* SC23-8824
- *Subsystem Device Driver User's Guide for the IBM TotalStorage Enterprise Storage Server and the IBM System Storage SAN Volume Controller*, SC26-7540
- *IBM System Storage Open Software Family SAN Volume Controller: Installation Guide*, SC26-7541
- *IBM System Storage Open Software Family SAN Volume Controller: Service Guide*, SC26-7542
- *IBM System Storage Open Software Family SAN Volume Controller: Configuration Guide*, SC26-7543
- *IBM System Storage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, SC26-7544
- *IBM System Storage Open Software Family SAN Volume Controller: CIM Agent Developers Reference*, SC26-7545
- *IBM System Storage Open Software Family SAN Volume Controller: Host Attachment Guide*, SC26-7563
- *Command-Line Interface User's Guide*, SC27-2287
- *IBM System Storage Productivity Center User's Guide Version 1 Release 4,* SC27-2336
- *IBM TotalStorage Multipath Subsystem Device Driver User's Guide*, SC30-4096
- *IBM System Storage SAN Volume Controller V5.1.0 - Host Attachment Guide*, SG26-7905
- *IBM Tivoli Storage Productivity Center IBM Tivoli Storage Productivity Center for Replication Installation and Configuration Guide*, SC27-2337
- *IBM TotalStorage Multipath Subsystem Device Driver User's Guide*, SC30-4096

Online resources

These websites are also relevant as further information sources:

- IBM TotalStorage home page:

  http://www.storage.ibm.com

- IBM System Storage Interoperation Center (SSIC)

  http://www.ibm.com/systems/support/storage/ssic/interoperability.wss

- IBM Tivoli Storage Manager for Storage Area Networks:

  http://www.ibm.com/software/products/us/en/tivostormanaforstorareanetw/

- IBM SAN Volume Controller supported platforms:

  http://www-1.ibm.com/servers/storage/support/software/sanvc/index.html

- IBM TotalStorage Virtualization home page:

  http://www-1.ibm.com/servers/storage/software/virtualization/index.html

- IBM SAN Volume Controller support page:

  http://www-947.ibm.com/systems/support/supportsite.wss/selectproduct?taskind=4&brandind=5000033&familyind=5329743&typeind=0&modelind=0&osind=0&psid=sr&continue.x=1

- IBM SAN Volume Controller online documentation:

  http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

- IBM Redbooks publications about the SAN Volume Controller:

  http://www.redbooks.ibm.com/cgi-bin/searchsite.cgi?query=SVC

- Download site for Windows Secure Shell (SSH) freeware:

  http://www.chiark.greenend.org.uk/~sgtatham/putty

- IBM site to download SSH for IBM AIX:

  http://oss.software.ibm.com/developerworks/projects/openssh

- Open source site for SSH for Microsoft Windows and Apple Mac platforms:

  http://www.openssh.com/windows.html

- Cygwin Linux like environment for Windows:

  http://www.cygwin.com

- Microsoft Knowledge Base Article 131658:

  http://support.microsoft.com/kb/131658

- Microsoft Knowledge Base Article 149927:

  http://support.microsoft.com/kb/149927

- Windows Sysinternals home page:

  http://www.sysinternals.com

- Subsystem device driver download site:

  http://www-1.ibm.com/servers/storage/support/software/sdd/index.html

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM

Redbooks

Implementing the IBM Storwize V7000 Unified Disk System

(0.5” spine)
0.475”<->0.873”
250 <-> 459 pages

Implementing the IBM Storwize V7000 Unified Disk System

Consolidates storage and file serving workloads into an integrated system

Simplifies management and reduces cost

Integrates support for IBM Real-time Compression

This IBM Redbooks publication introduces the IBM Storwize V7000 Unified Disk System, a virtualized storage system that consolidates block and file workloads into a single storage system. Advantages include simplicity of management, reduced cost, highly scalable capacity, performance, and high availability. It also offers improved efficiency and flexibility through built-in solid-state drive optimization, thin provisioning, IBM Real-time Compression, and nondisruptive migration of data from existing storage. The system can virtualize and reuse existing disk systems, which offers a greater potential return on investment.

We suggest that you familiarize yourself with the following Redbooks publications to get the most from this book:

- *Implementing the IBM Storwize V7000 V6.3*, SG24-7938
- *Implementing the IBM System Storage SAN Volume Controller V6.3*, SG24-7933
- *Real-time Compression in SAN Volume Controller and Storwize V7000*, REDP-4859
- *SONAS Implementation and Best Practices Guide*, SG24-7962
- *SONAS Concepts, Architecture, and Planning Guide*, SG24-7963

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8010-01 ISBN 0738439185