Paper 3290-2015

SAS® Analytics on IBM® FlashSystem™ storage: Deployment scenarios and best practices

Harry Seifert, IBM Corporation; Matt Key, IBM Corporation; Narayana Pattipati, IBM Corporation; David Gimpl, IBM Corporation

ABSTRACT

SAS® Analytics enables organizations to tackle complex business problems using Big Data and provide insights needed to make critical business decisions. With actionable and timely insights, organizations can serve customers better, spot market trends sooner and support the launch of new products and offerings. A well-architected enterprise storage infrastructure is needed to realize full potential of SAS analytics. However, as the need for big data analytics and rapid response times increases, the performance gap between server speeds and traditional hard disk drive (HDD) based storage systems can be a significant concern. The growing performance gap can have detrimental effects, particularly when it comes to critical business applications. As a result, organizations are looking for a newer, smarter, faster kind of storage systems to accelerate business insights.

IBM FlashSystem™ storage systems store the data in flash memory modules and they are designed for faster access times and support input/output (I/O) transactions with significantly lower latency per second than HDD-based solutions. Due to their macro-efficiency technology, FlashSystem storage systems also consume lower power and have significantly lower cooling and space requirements.

The paper introduces the benefits of FlashSystem storage for deploying SAS Analytics and highlights some of the deployment scenarios and architectural considerations. The paper also describes best practices and tuning guidelines for deploying SAS Analytics on FlashSystem storage. The deployment scenarios and best practices would help SAS Analytics customers in architecting solutions with IBM FlashSystem storage.

INTRODUCTION

In the IBM Storage portfolio, IBM has two FlashSystem products that emphasize on application performance– the FlashSystem 900 model designed as the premier high performance storage, and our FlashSystem V9000 model which is a full-featured all-flash enterprise software-defined storage system.

FlashSystem V9000 is an all-flash storage solution with a comprehensive feature list for accommodating enterprise storage demands. These functions and system design deliver values that can be classed into three dimensions: scalable performance, enduring economics, and agile integration.

IBM FlashSystem V9000, at its core, delivers highly available storage with latency in the measurement of microseconds. This low latency delivers time once spent on I/O back to the application for increased throughput, higher efficiency, and better end user experience. As workloads increase in size and throughput requirements, IBM FlashSystem V9000 has the capability to scale up with additional flash storage shelves within the same, intuitive management window. As performance needs grow, IBM FlashSystem V9000 can also scale out to add additional controller throughput to the storage namespace.

To soften the cost of high performance storage, IBM FlashSystem V9000 also offers data reduction technologies that apply well to SAS. Thin provisioning is available to prevent overbuying of storage capacity and also includes IBM Real-Time Compression to reduce the footprint of the data. These technologies drive the native price of flash to a price parity of tier one disks and are adjustable on a per-volume granularity.

As not all applications require high performing storage, or not all data within an application for that matter, IBM FlashSystem V9000 includes storage virtualization functionality to couple flash performance with lower tier capacity-oriented storage technology. This is made possible by storage virtualization and

creates a consistency group for applications across flash and disk for copy services and disaster recovery.

For the pinnacle of tier-0 performance there is also IBM FlashSystem 900. This shelf of flash storage maintains the high availability and serviceability of enterprise storage, but foregoes the data management functions offered in IBM FlashSystem V9000 in favor of lowest available latency. This solution is most adopted for critical point solutions.

SAS ANALYTICS WORKLOAD CHARACTERISTICS

The SAS Analytics workloads are compute, IO and memory intensive and majority of the SAS Analytics jobs are either compute intensive or IO intensive or both. The jobs predominantly perform large-block sequential IO operations with a mix of read and write operations; read and write ratio is customer workload specific and it varies from 55:45 to 80:20. In order to meet the IO needs of the SAS Analytics jobs, the underlying infrastructure needs to support a minimum of 100 MBps per core IO throughput.

SAS® Scalable Performance Data Server® (SPDS) IO characteristics are different from general SAS Analytics workloads. SAS SPDS IO is random with dynamic block-sizes, where block size can vary from 8KB to 1MB. And SAS SPDS can run in parallel with traditional SAS Analytics jobs.

SAS workload consists of two different set of file systems – SASDATA and SASWORK. While SASDATA stores persistent data which includes input, output files and SAS code, the SAS work area (SASWORK and SASUTIL file systems) stores temporary data that gets deleted at the end of each SAS job. In typical customer deployments, SASDATA requires much more space for storing persistent data. It is very important to layout and configure these file systems for optimal IO performance.

SAS ANALYTICS DEPLOYMENT ON IBM FLASHSYSTEM STORAGE

The section describes some of the deployment scenarios for SAS Analytics on IBM FlashSystem storage. SAS Analytics customers can deploy their workloads in different ways to realize the benefits of FlashSystems. Some of the deployment scenarios are:

- FlashSystem as a storage array for SAS workloads
- FlashSystem as part of hybrid-storage architecture for SAS workloads
- FlashSystem as a storage array for deploying shared file systems' metadata

FLASHSYSTEM AS A STORAGE ARRAY FOR ENTIRE SAS WORKLOADS

IBM FlashSystem is a matured all-flash storage array that can be used to deploy entire SAS workload – SASDATA, SASWORK and SASUTIL. The FlashSystem with its macro-efficiency design, also consume lower power and have significantly lower cooling and space requirements, all while allowing server processors to run SAS Analytics more efficiently.

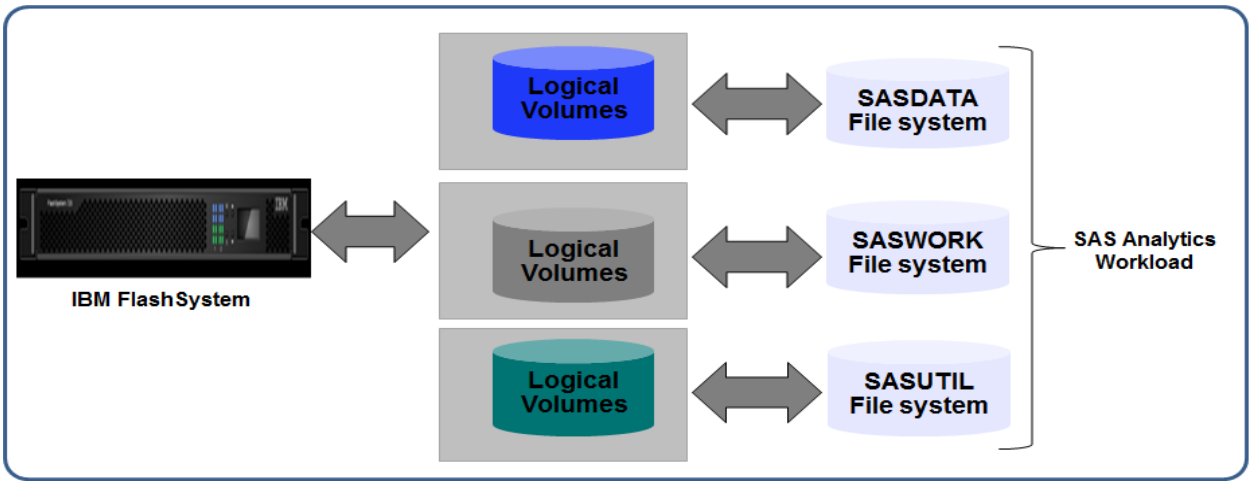

Figure 1. SAS Analytics deployment on FlashSystem array

Latest offering from IBM FlashSystem family FlashSystem 900 supports up to 57 TB of usable capacity in RAID5. SAS analytics customers may have space and/or cooling constraints in their existing data center. IBM FlashSystem is a perfect choice for such environments which provides superior performance in a small form factor (4U). Some of the FlashSystem offerings also support Real-Time Compression (RTC), which offers up to 5:1 data reduction and can deliver flash for less than the cost of a disk array at higher price-performance than a disk-array.

FLASHSYSTEM AS PART OF HYBRID-STORAGE ARCHITECTURE FOR SAS WORKLOADS

When SAS Analytics customers have lot of data running into hundreds of TB or PB, it is not cost effective to deploy entire SAS workload on a FlashSystem. Deploying SAS Analytics on a hybrid storage model with a disk-based storage array (e.g. IBM XIV Storage System and IBM DS 8870) and FlashSystem (all-flash storage) helps customers to get best of both worlds. This solution provides higher storage capacity while accelerating SAS Analytics at the same time. The persistent data of SAS workloads (SASDATA), which has larger physical space requirements can be deployed on disk-based storage system and temporary SAS work area (SASWORK and SASUTIL) can be deployed on FlashSystem.

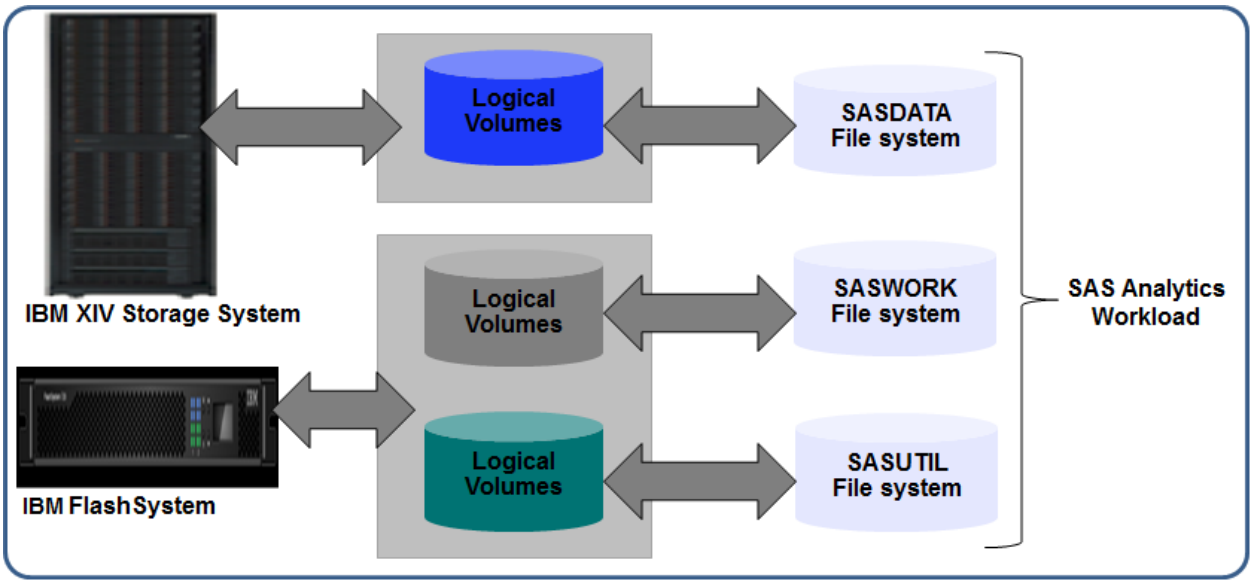

Figure 2. SAS Analytics deployment on hybrid-storage involving FlashSystem and XIV Storage System

FLASHSYSTEM AS A STORAGE ARRAY FOR DEPLOYING SHARED FILESYSTEM METADATA

A shared file system, like IBM Spectrum Scale™ (formerly IBM GPFS™), is a required and integral component of all SAS® Grid Manager deployments, SAS® Enterprise Business Intelligence deployments with load balanced servers on multiple systems, and other types of distributed SAS applications. The speed of the metadata access in a shared file system determines the overall IO performance of the SAS Analytics solutions. For example, IBM Spectrum Scale distributes metadata across the cluster nodes and it is very important to ensure metadata deployed on a faster storage sub system. IBM FlashSystem can be used to deploy shared file system metadata for accelerating file access. When metadata is deployed on FlashSystem, the small block I/O operations of metadata no longer interfere with the large streaming accesses for the data.

Spectrum Scale supports multiple storage pools for a file system. A storage volume can contain data or metadata or both. While creating network shared disk (NSD), designate the FlashSystem volumes to contain metadata-only and XIV volumes to contain data-only. And designate FlashSystem volumes to *system* storage pool and XIV volumes to data storage pool, say *appdata.* Storage pool *system* is used for metadata by default. Spectrum Scale policies can be used to move data to *appdata* storage pool. And while creating the file systems, use different file system block sizes for data and metadata, for optimal IO performance.

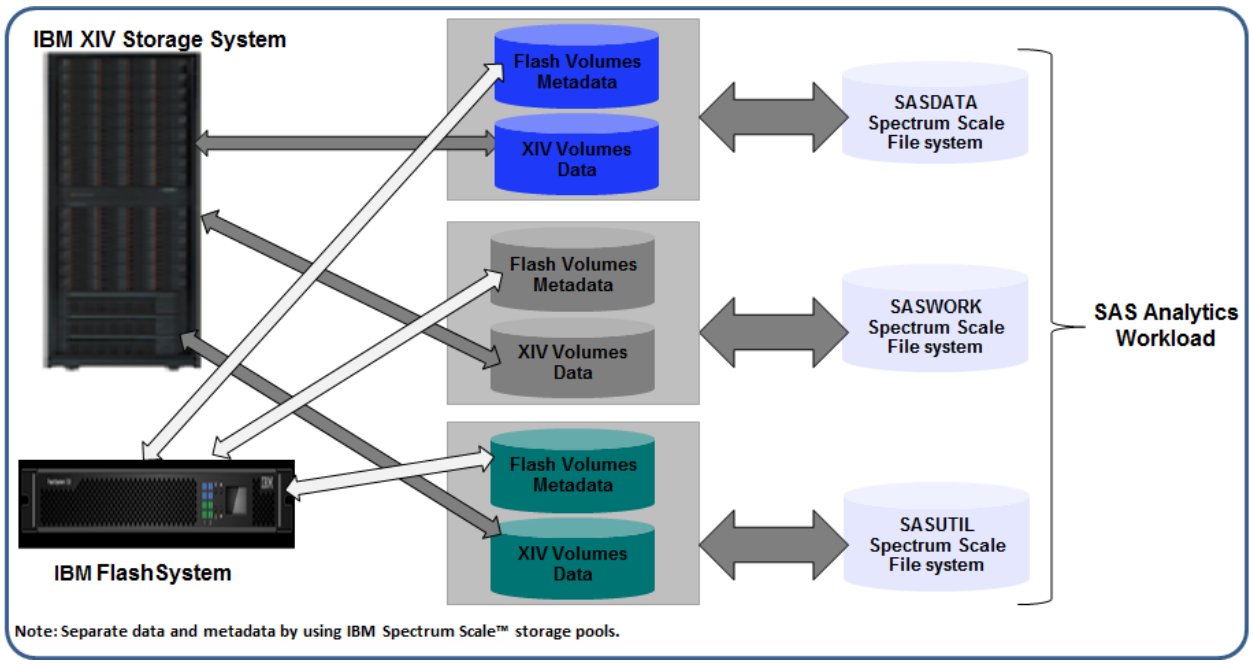

Figure 3. SAS Analytics shared file systems' metadata deployment on IBM FlashSystem

The SAS analytics deployment on hybrid storage environment is described in detail in the following section.

SAS ANALYTICS DEPLOYMENT ON HYBRID-STORAGE ENVIRONMENT WITH FLASHSYSTEM

This section describes deployment of SAS Analytics on IBM POWER8 processor-based servers and hybrid-storage environment involving IBM XIV® Gen3 storage system and IBM FlashSystem 840.

As described in the above sections, SAS workloads typically have two different set of file systems. While SASDATA stores persistent data which includes input and output files, the SAS work area (SASWORK and SASUTIL file systems) stores temporary data that gets deleted at the end of each SAS job. In typical customer deployments, SASDATA requires much more space for storing input files and output files. Hence, it is beneficial to deploy SASDATA file system on XIV Storage System (which is disk based storage) and SAS work area on FlashSystem. The figure below describes the deployment in detail.

IBM Power S822 (8284-22A) is a scale-out server with 2 sockets, 20 cores, and 256 GB memory. The server is configured with a Virtual I/O Server (VIOS) and five client logical partitions (LPARs). The VIOS helps in sharing the Fibre Channel (FC) adapters among the client LPARs by virtualizing the physical FC adapters using N_Port ID Virtualization (NPIV). The test configuration uses a single VIOS, however, it is recommended to use dual-VIOS in production deployments for high availability. IO traffic from XIV and FlashSystem is segregated using separate FC ports (physical as well as virtual) as shown in the above image.

One of the LPARs is used for testing SAS analytics. Logical unit numbers (LUNs) are created on the XIV and FlashSystem and are mapped to the LPARs. The SAS workload Spectrum Scale file systems are created on the mapped LUNs for running SAS analytics jobs. SASDATA file system is created with 16 LUNs mapped from XIV Gen3 storage. 16 LUNs proved to be optimal for the workload used during the

testing. 32 LUNs mapped for SASWORK and SASUTIL each from FlashSystem 840. For AIX clients, 32 LUNs are recommended in the FlashSystem 840 implementation redbook.

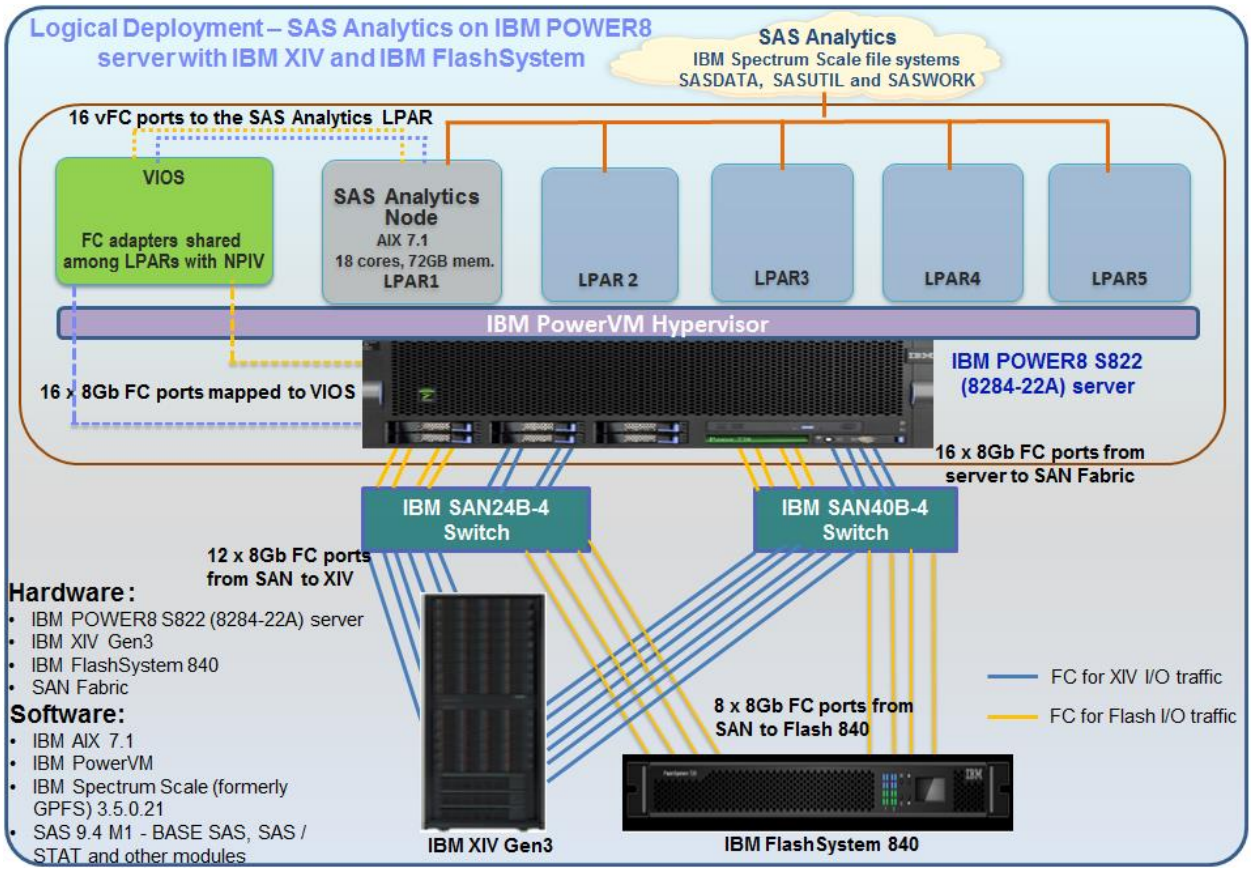

Figure 4. SAS Analytics deployment on hybrid-storage environment with FlashSystem and XIV

CONFIGURATION AND TUNING

This section describes the test environment, configuration and tuning recommendations for deploying SAS Analytics on IBM FlashSystem in a hybrid storage environment.

HARDWARE

Power S822 server configuration

- Model: 8284-22A
- Firmware version: FW810.02 (061)
- Processor architecture: POWER8
- Clock speed: 4116 MHz
- SMT: OFF, 2, 4, 8 (SMT4 is default)
- Cores: 20 (18 cores for the SAS Analytics LPAR and 2 cores for VIOS)
- Memory used: 256 GB (72 GB for the SAS Analytics LPAR and 8 GB for VIOS)
- Internal drives: Four 600 GB (used for booting VIOS and LPARs)
- FC connectivity: Four quad-port 8Gb FC ports (16 ports) attached to the server; used 8 ports during the testing

XIV configuration

- XIV machine type: 2810
- Machine model: 114
- System version: 11.5.0.x
- Drives: 180 SAS drives each with 2 TB capacity and 7200 rpm speed
- Usable capacity : 161 TB
- Modules: 15
- Cache: DDR3 360 GB
- SSD cache: 6 TB
- Connectivity: Six 8Gb dual-port Fibre Channel (FC) adapters (12 ports) connected to storage area network (SAN)
- Stripe size: 1 MB (default)
- SSD cache: Enabled (by default) for all volumes used in workload

FlashSystem 840 configuration

- Canisters : 2
- Flash Modules : 12 x 3.7 TB
- Flash Type : Toshiba 24nm eMLC 512 Gb
- Total Capacity : 37.5 TB (RAID5) (11 member drives and 1 spare)
- Code Level (FW) : 1.1.3.2
- FC Ports connected : 8 x 8Gb FC ports (4 ports attached to each canister) are connected from SAN to FlashSystem 840
- A single RAID5 array is created out of the 12 flash modules with one of the modules as a spare and stripe size used is 4KB.

Storage area network (SAN) configuration

- Two FC switches - IBM System Storage SAN24B-4 Express (24 ports) and IBM System Storage SAN40B-4 (40 ports); both support NPIV
- Sixteen 8Gb dual-port FC ports connected from Power S222 server to the SAN fabric; Eight ports connected to the first switch and eight more ports connected to the second switch
- Twelve 8Gb FC ports connected from SAN Switches to XIV Gen3
- Switch zoning is performed on the SAN switches such that each logical disk assigned from XIV to the LPAR has 12 or 24 paths.

SOFTWARE

- SAS 9.4 M1 64-bit software
- IBM AIX® 7100-03-03-1415
- VIOS 2.2.3.3
- IBM PowerVM Enterprise Edition
- IBM Spectrum Scale (formerly GPFS) 3.5.0.21

TUNING RECOMMENDATIONS

The following list describes the tuning guidelines for SAS workloads to perform optimally on POWER8 processor-based servers with the AIX 7.1 operating system on hybrid-storage environment with XIV Gen3 and FlashSystem 840.

- AIX tuning
  Follow the *SAS on Power / AIX Tuning Guides at* *www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529*.
- XIV Storage System
  - No specific tuning required on XIV Storage Systems. It works using the default environment.
  - XIV by default uses wide stripe and stripes data across all the available disks. The XIV Gen3 system used in testing has 180 disks.
  - XIV uses 1 MB as the stripe size, by default.

- Spectrum Scale tuning
  pagepool 8G (default 1G)
  seqDiscardThreshhold 1G (default 1MB)
  maxMBpS 12000 (default 2048)
  prefetchPct 40 (default 20)
  maxFilesToCache 20000 (default 4000)
  scatterBuffers no (default yes)
  stealFromDoneList no (default yes)
- FC Adapter and logical disk tuning
  Since the IO paths and FC ports are different for XIV and IO traffic, it is recommended to tune them differently to optimize the overall IO performance. The queue depth, max_transfer_size needs to be tuned for XIV and FlashSystem adapters at both the client LPAR and VIOS. Refer to *SAS business analytics deployment on IBM POWER8 processor-based systems* white paper at www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102515.

- FlashSystem specific tuning for LUNs.
  - 32 LUNs per file system
  - Sector / block size while creating LUN: 512 (default)
  - Auto Contingent Allegiance (ACA) support: yes (default)

DEPLOYMENT BEST PRACTICES

The best practices deployed in this section are for a deployment with hybrid-storage involving IBM FlashSystem 840 and IBM XIV Gen3. However, the recommendations can be applied to a scenario where entire SAS workload is deployed on an IBM FlashSystem.

Spectrum Scale (formerly GPFS) file system layout and configuration

SASDATA file system is created out of LUNs mapped from XIV storage and SASWORK and SASUTIL file systems are created out of the LUNs mapped from FlashSystem 840. The sizes of the volumes can vary depending on the total size of the file systems.

- SASDATA (on XIV Gen3):
  - 4 TB total size with 16 LUNs of 256 GB each
  - file system block size: 1 MB
  - file system block allocation type: Cluster (default is scatter)
- SASWORK (on FlashSystem 840):

  - 4 TB total size with 32 LUNs of 172 GB each
  - file system block size: 256 KB / 512 KB
  - file system block allocation type: Scatter (default)
- SASUTIL (on FlashSystem 840):
  - 1.6 TB total size with 24 LUNs of 50 GB each
  - file system block size: 256 KB / 512 KB
  - file system block allocation type: Scatter (default)

Please note that during the testing in lab environment, *scatter* file allocation type proved to be optimal when using FlashSystem. Whereas *cluster* block allocation type proved to be optimal while using XIV Gen3 storage. The allocation type need to be chosen based on the number of LUNs and cluster size in production environments.

Segregate IO traffic

The LPAR used for testing has 16 virtual client FC ports that are mapped to 16 virtual server FC ports at VIOS. The virtual server FC ports in turn are mapped to 16 physical FC ports that are assigned to the VIOS from HMC. IO traffic is segregated at switches by optimal zoning. The zoning details are depicted in the below figures. The IO traffic segregation helps in tuning FC adapters differently to suite the nature and size of the IO.

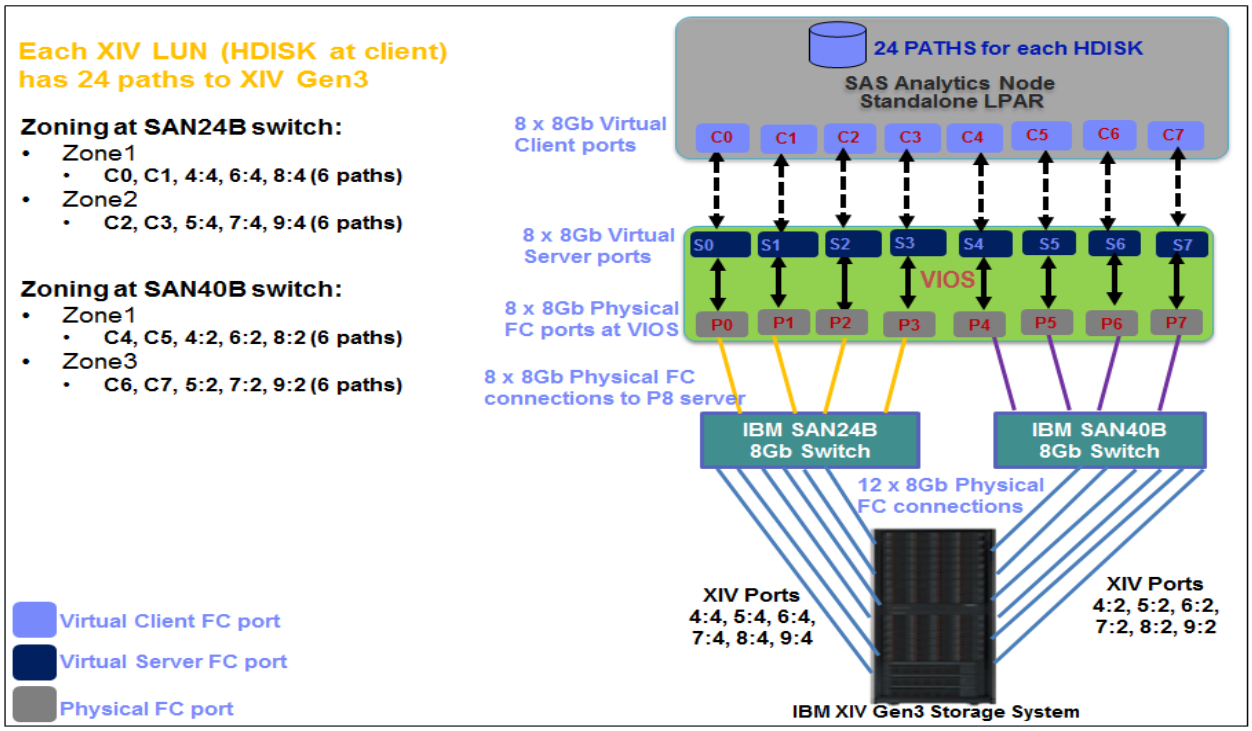

Figure 5. Recommended zoning details for XIV IO traffic in hybrid-storage environment

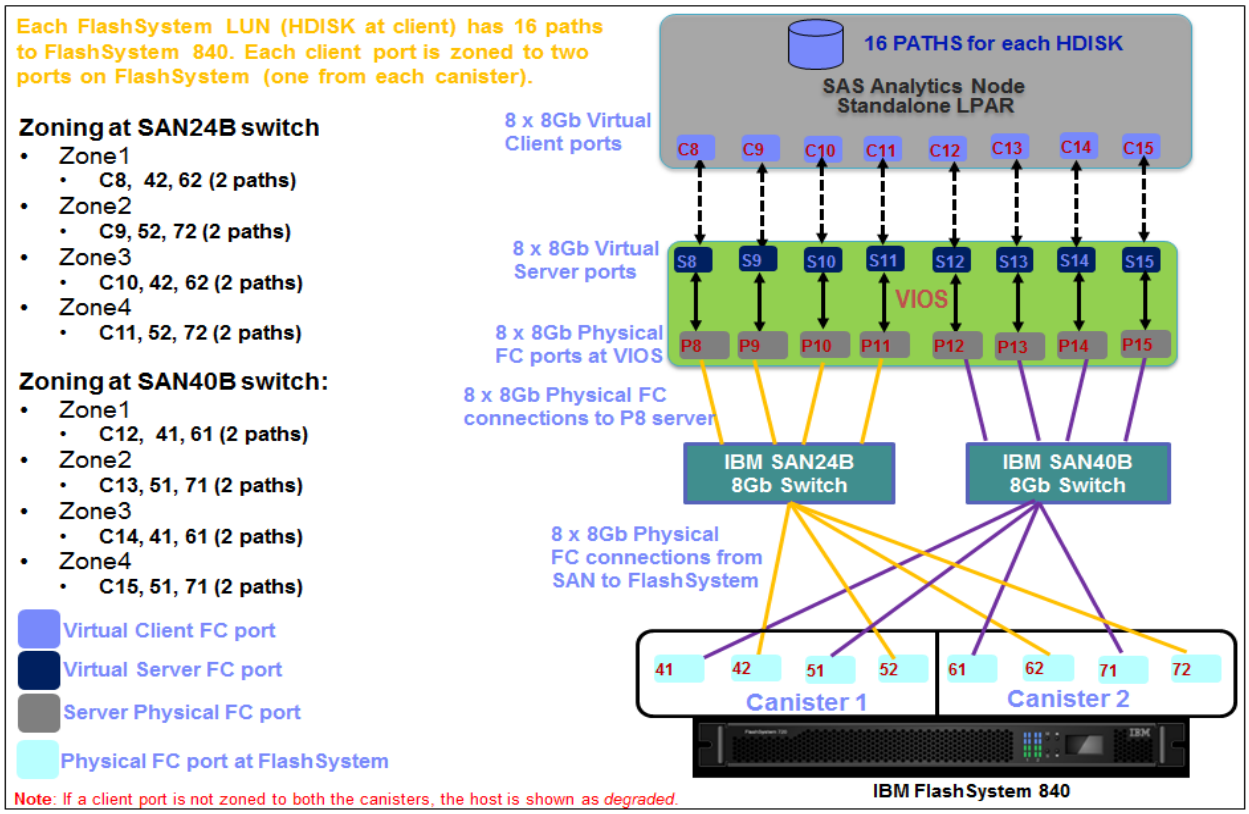

Figure 6. Recommended zoning details for FlashSystem IO traffic in hybrid-storage environment

SAS ANALYTICS PERFORMANCE ON FLASHSYSTEM

SAS WORKLOAD PERFORMANCE IN HYBRID-STORAGE ENVIRONMENT

To understand SAS business analytics performance on IBM Power S822 server and hybrid-storage environment with XIV Gen3 and FlashSystem 840, SAS mixed analytics 20-session workload was run on a single LPAR. The 20-session workload was the appropriate sized workload given the compute and I/O demands of the workload on the Power S822 server with two sockets and 20 cores. The Power S822 server with 20 cores and 256 GB memory can also support a more intensive 30-session workload; however, lower response times are expected compared to a 20-session workload.

20-SESSION WORKLOAD PERFORMANCE

Here is the configuration used for the workload:

- LPAR is configured with 16 cores in dedicated mode with SMT4; 64 GB memory
- VIOS is configured with 2 cores in dedicated mode and SMT4; 8 GB memory
- Used 512KB block size for SASWORK and SASUTIL GPFS file systems that are deployed on FlashSystem (512KB block size proved optimal for GPFS file systems on FlashSystem 840).
- The workload was run with no other competing activity on the server or storage systems.

Performance summary of the workload:

- Workload response time is 1134 minutes, user time is 781 minutes, and system time is 41minutes.

- Combined (at FlashSystem and XIV) peak I/O throughput is 5.0 GBps and sustained I/O throughput is 3.25 GBps, which translates to 200 MBps per core sustained I/O throughput.
- At XIV, peak latency is 3 ms and average latency is 1 ms (for SASDATA file system); At FlashSystem 840, peak latency is 4 ms and average latency is 2 ms (for SASWORK and SASUTIL file systems).
- Average processor usage (user + sys) is 56% and wait is 6%
- At host, average disk service time is 3.5 ms and the peak disk service time is 7 ms
- Total data transferred during the workload is 11 TB (9 TB read and 2 TB write)

Note: POWER8 Processor Utilization Resource Register (PURR) factors were applied on the CPU (user and system) times mentioned above. Refer to *SAS business analytics deployment on IBM POWER8 processor-based systems* paper at www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102515 for more details on PURR and how to apply them

The graphs below depict the IO performance at XIV and Flash for the 20-session workload in hybrid-storage environment.

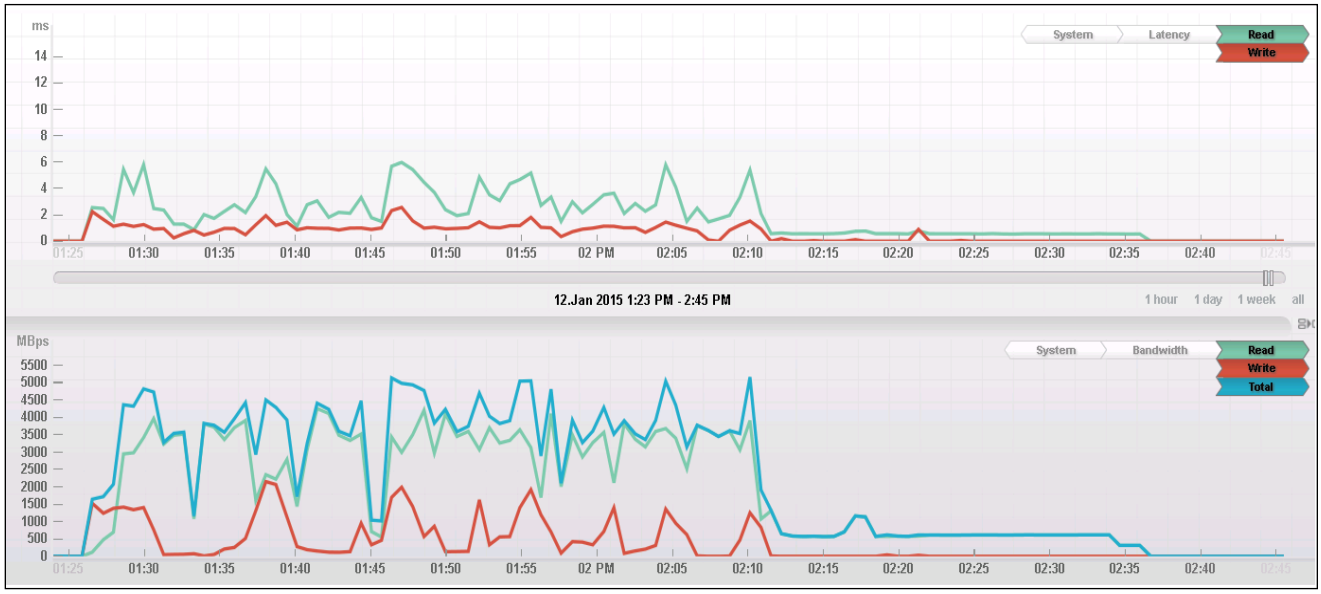

Figure 7: Mixed Analytics 20-session IO performance at FlashSystem for SASWORK & SASUTIL

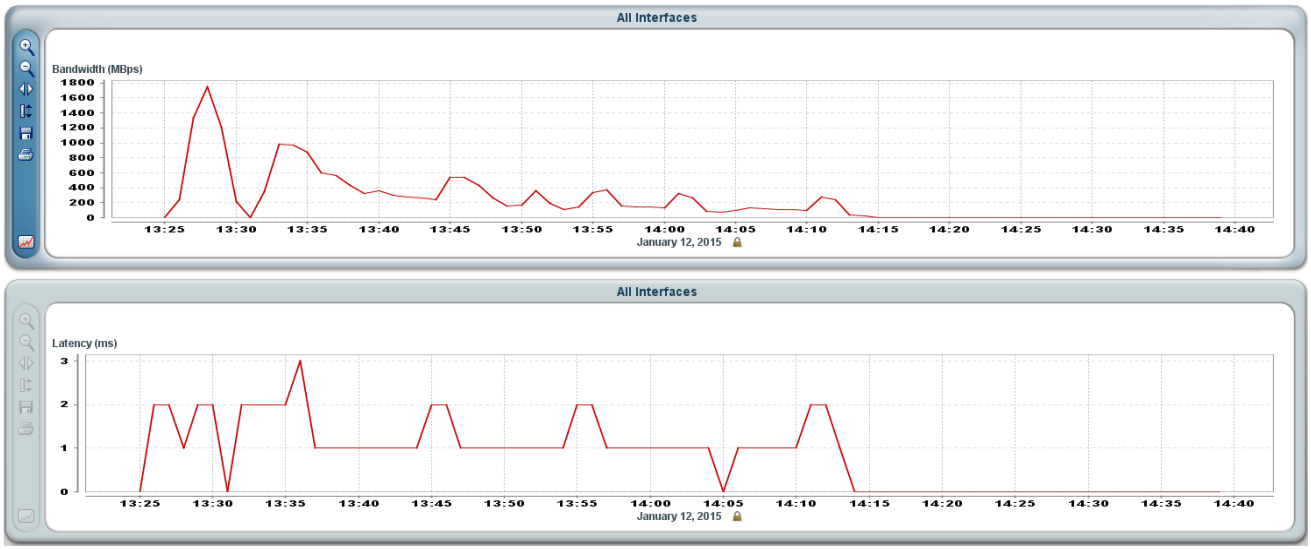

Figure 8: Mixed Analytics 20-session IO performance at XIV for SASDATA

The workload used during the testing is large-block sequential in nature with 80:20 read-write ratios. This is true for many SAS analytics workloads. If you observe figures 9 and 10 above, the SASWORK and SASUTIL file systems, which are deployed on FlashSystem in hybrid-storage environment, contribute to 85 – 95% of the IO generated by the workload. During the testing, it is proved that SAS Analytics workloads could effectively leverage hybrid-storage architecture involving XIV and IBM Flashsystem 840. In the hybrid-storage environment, the XIV storage system is used for deploying persistent data for optimal storage space utilization, the FlashSystem is used for deploying the SAS work area for optimal IO throughput performance.

CONCLUSION

Flash storage technology offers a ray of hope for SAS® Analytics customers who face challenges of IO performance, scalability and longer running jobs. IBM FlashSystem is an all-flash storage system that is proven to work very well with SAS Analytics solutions.

While IBM XIV is a proven disk-based storage system for SAS® workloads, customers can consider hybrid storage architecture to accelerate analytics jobs, keeping the overall total cost ownership (TCO) low. In some cases, SAS analytics customers may have space and/or cooling constraints in their existing data center. To augment storage capacity, buying a full-rack sized disk-based storage subsystem may not be an option for such customers. IBM FlashSystem is a perfect choice for such environments which provides superior performance in a small form factor.

The paper described some of the deployment scenarios for SAS Analytics on Flashsystem along with deployment best practices and performance tuning guidelines.

REFERENCES

SAS Published white paper Oct 2014 “A Survey of Shared File Systems: Determining the Best Choice for your Distributed Applications” at http://support.sas.com/rnd/scalability/papers/SurveyofSharedFilepaper_20131010.pdf.

Beth Hoffman, Narayana Pattipati, 2015 “SAS business analytics deployment on IBM POWER8 processor-based systems” published at www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102515.

Narayana Pattipati September 2012 “SAS 9.3 grid deployment on IBM Power servers with IBM XIV Storage System and IBM GPFS” published at www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102192.

Narayana Pattipati May 2014 “Accelerate insights with SAS Business Analytics and IBM FlashSystem” published at http://public.dhe.ibm.com/common/ssi/ecm/en/tsw03263usen/TSW03263USEN.PDF

ACKNOWLEDGMENTS

Authors would like to thank Beth Hoffman, Executive IT Specialist, Big Data Analytics ISV Solutions, IBM Corporation, for reviewing the paper.

RECOMMENDED READING

How to Maintain Happy SAS Users (presented at NESUG 2007) paper
http://www.nesug.org/proceedings/nesug07/as/as04.pdf

IBM Storage Systems - www.ibm.com/systems/storage

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Harry Seifert
ISV Solutions, Sales and Distribution, IBM Corporation
Phone (W): 1-720-396-7015
seifert@us.ibm.com

Narayana Pattipati
ISV Technical Enablement, IBM Corporation
Phone: + 91-80-41774245
npattipa@in.ibm.com

Matt Key
FlashSystem Technical Sales, IBM Corporation
Phone (W): 1-713-278-6272
mkey@us.ibm.com

Dave Gimpl
FlashSystem Optimized Solutions, IBM Corporation
Phone (W): 1-919-543-0526
gimpl@us.ibm.com