EMC NetWorker Design
Best Practices with Data Domain

Introduction

Written by:
Ken Ciolek and Natalie Meek
GlassHouse Technologies, Inc.

In today's IT industry where information is crucial to company business growth, protecting the data against loss has become very critical. Even with the development of more cost effective technologies, completely and effectively protecting what is inevitably multiple copies of the same information is a major challenge for many businesses.

Standard practices and technologies are being pushed to their limits in many backup environments, most to an unsuccessful or inefficient outcome. This challenge is common for many EMC NetWorker customers, and finding the right new technology to bridge the gap between current state and desired state is often a long, difficult process.

Data Domain provides an alternative near line storage solution for NetWorker customers who are faced with never-ending data growth and unabated storage expansion associated with ballooning backup and archive data. While NetWorker is one of the most scalable data protection solutions available to the market, data growth and data retention requirements drive near-continual expansion to NetWorker pools.

The scope of this whitepaper focuses on how the Data Domain deduplication storage solution integrates with standard NetWorker architectural and operational environments in order to overcome the growing gap between what is actually being accomplished and what needs to be accomplished.

.

The NetWorker Architecture and Terminology

The EMC NetWorker client / server environment provides the ability to protect your enterprise against the loss of valuable data. In a network environment, where the amount of data grows rapidly as servers are added to the network, the need to protect data becomes crucial. The EMC NetWorker product gives you the power and flexibility to meet such a challenge.

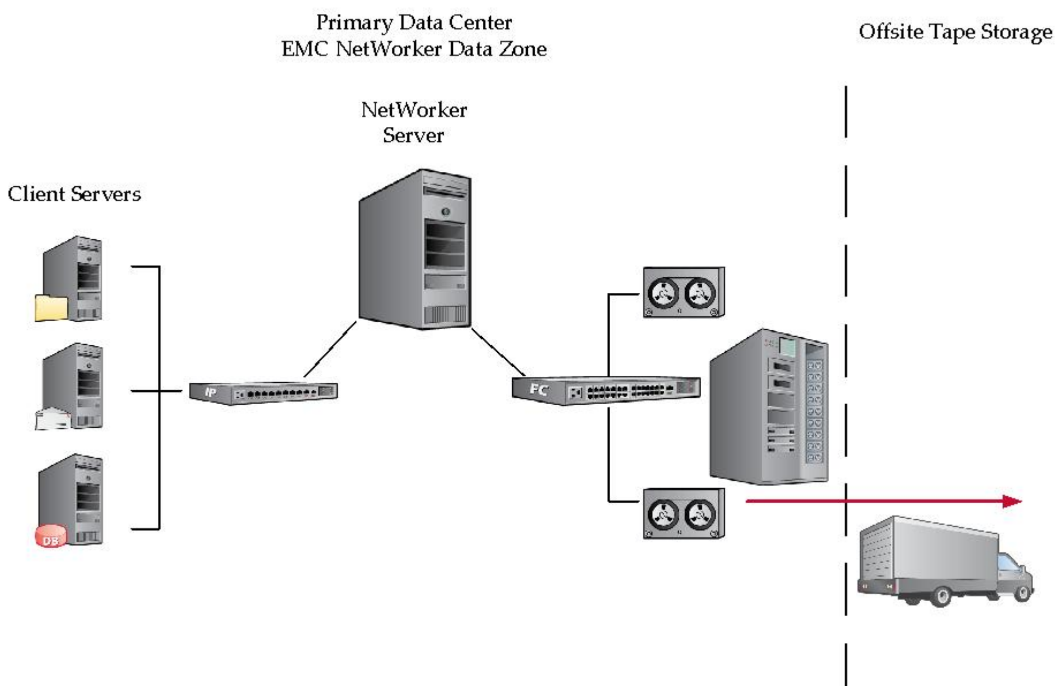

Figure 1: Classic NetWorker Architecture Design

Each NetWorker server instance is supported by multiple self-managed relational databases (resource, client file index and media) and various logs. Historically, customers backed up data directly to tape for nightly backups. Alternatively, more customers have moved to using file and advanced file type devices to backup data to disk for faster write speeds. This data is then migrated daily to physical tape as a replica of the original data (commonly referred to as "clone" data) for purposes of disaster recovery and media recovery. Common terminology used in this whitepaper is provided in the following table.

Term	Definition
Client	A computer whose data can be backed up or recovered by a NetWorker Server.
Storage Node	A Client with the ability (via SAN or SCSI connection) to back up data to a library or another storage location. Storage Nodes can also receive data from other NetWorker clients.
SAN Storage Node	A NetWorker Client with the ability (via SAN or SCSI connection) to back up its own data to a library or another storage location (SAN Storage nodes can not back up other clients).
NetWorker Server	The computer running the NetWorker server software that contains the NetWorker databases (client file index, media and resource) and provides backup and recovery operations to the clients on the same network. Servers can (and usually do) function as storage nodes.
Data Zone	All computers administered for backup and recovery by a NetWorker server. There can only be 1 NetWorker server in a data zone.
Advanced File Type Device	NetWorker term for disk backup locations with the ability to read and write simultaneously. The advanced file type (adv_file) is usually created for large disk locations with the ability to dynamically extended (using a volume manager) if the disk runs out of space while in use as an advanced file type device (usually during backup).
Save Set	A group of files or locations backed up onto a device
Group	One or more clients configured to back up data at a specific time of the day.
Schedule	The level of backup (full, incremental, differential, skip) a client is designated to run during a backup
Browse Policy	The policy to determine how long data will reside in the Client File Index

Term	Definition
Retention Policy	The policy to determine how long data can be recovered.
Client Resource	The client resource lists the save sets to be backed up on a client. It also includes the schedule, directive browse policy, and retention policy for the client. A client can have multiple client resources however; the Savesets can not contain the same information to be located within the same group.
Pools	An attribute in NetWorker that allows you to classify data. Pools are separated by the following information (in hierarchical order): Groups, clients, save sets, level.
Cloning	The act of copying backed up data. Clones function identically as the original backup volume. Cloning can be completed anywhere from one save sets to an entire volume (or multiple volumes).
Staging	Moving data from one storage type to another. Staging also later removes the data from its original location.
Full Backup	A full backup is all files within a save set, regardless of whether they have been modified.
Incremental Backup	An incremental backup only backs up any files that have changed since the previous backup.
Differential Backup	Differential backup is based on a numerical system. NetWorker has 9 levels of backup. Levels 1 through 9 backs up files that have changed since the last lower numbered backup level. Differential backups typically are configured to back up all data that has changed since the last full backup.
Client File Index Database	The database the tracks every file or object that is backed up. There is one Client File Index per physical client.
Media Database	The database that tracks the all save set and volume life cycle information. There is one Media database per data zone
Resource Database	The database that tracks all resource information (such as clients, schedules, groups, etc…). There is one resource data base per data zone

Table 1: NetWorker Terminology

Typical NetWorker Challenges

The typical NetWorker environment supports a handful to thousands of clients. NetWorker scales by adding additional NetWorker storage node/SAN Storage nodes, and associated disk and tape resources. The most common challenges in NetWorker environments include:

- Completing backups, staging and cloning with limited time and physical resources
- Contending with extremely large client backups (several million files per client)
- Scaling NetWorker databases, logs, media management and pools to keep up with sheer demand
- Eliminating redundant data backup locations (multiple full/incremental copies of databases, aggressive backup retention policies, etc.)
- Eliminating performance bottlenecks (NetWorker server type, networking, client issues, etc.)
- Lack of capacity planning and reporting disciplines

Technology Overview

Data Domain reduces unnecessary NetWorker data storage via inline data deduplication and traditional compression. Data deduplication is performed on incoming data streams and allows only the new segments of data to be identified and stored as unique instances within the Data Domain file system. The following table lists key terminology for Data Domain.

Term	Definition
Data Domain System	A standalone Data Domain storage appliance, gateway, or a single controller in a DDX array.
Protected Data Size	The sum total of all file sizes in the set of primary data being backed up.
Logical Storage Size	The total size of all backup images in all pools on a Data Domain system. This total size includes all pools in NetWorker mapped to a Data Domain system instance, which can include primary disk pools, and clone storage pools.
Disk Pool Dump Size	The size of an individual backup image written to a pool (for example, one night's worth of "backup data").
Addressable Capacity	The amount of physical space available on a Data Domain system to store deduplicated and compressed backup images.
Physically Consumed Storage	The amount of addressable capacity on a Data Domain system currently storing backup data and associated metadata.
Cumulative Compression Factor	The ratio of the logical storage size to the physically stored

Term	Definition
	space.
Periodic Compression Factor	The ratio of one or more disk pool dumps to the physically consumed storage for those dumps. Note that the periodic compression factor over any interval beyond the first few days will typically exceed the cumulative compression factor by a large margin because the first version of a file written to a Data Domain system will compress less than subsequent versions. Consider for example two selective backups of 100 GB of protected data over two nights: typical periodic compression factors might be 2:1 the first night and 50:1 the second night, but the cumulative compression factor would only be ~4:1 (200 GB / 50+2 GB) rather than the 25:1 or so one might expect. Note further that while the cumulative compression factor is what determines cost per GB, it is the periodic compression factor that most affects replication bandwidth requirements.
Deduplication	Replacing redundant 4KB to 16 KB segments in incoming data streams with very small references to identical segments already stored on disk. Also known as "global compression".
Local Compression	Standard lossless compression algorithms. The available Local Compression algorithms available on a Data Domain system include LZ (Lempel-Ziv), gz and gzfast.
Cleaning	A periodic process to find any unreferenced segments on a Data Domain system and make that space available for subsequent backups. Because Data Domain systems never overwrite data, file deletes by a NetWorker server do not immediately make space available for subsequent backups — cleaning must run first. This cleaning process is not unique to Data Domain systems. Cleaning may be performed on a Data Domain system at the same time as backup/restore I/O, but because cleaning is a fairly resource intensive process it is best to schedule it for non-peak hours. The default schedule for cleaning is Tuesday morning at 6:00 a.m. but may be rescheduled for any convenient times during the week or manually via script or command line..

Table 2: Data Domain Terminology

Note: Data Domain's patented approach to deduplication is called Global Compression™ in Data Domain product literature, but for purposes of this whitepaper will be referred to as deduplication.

Data Domain data deduplication methods are more granular and variable than fixed segment size data deduplication competitors. Data Domain segment length is variable, ranging from 4-16KB. This is a significant differentiator from competitive products which perform deduplication at the file level or at a block level, resulting in more efficient deduplication capabilities.

Since the rate of primary data change (newly introduced unique 4K to 16K segments) stays about the same from night to night at most sites, the amount of physically consumed storage for a NetWorker backup is roughly the same as the physically consumed storage for an incremental NetWorker backup. The ratio of protected storage size to incrementally consumed physical storage each night stays about the same, but the periodic compression factor of an incremental backup is much lower than the periodic compression factor of a selective backup (because the former is much smaller in size). As a result, it is very inexpensive to include many versions of files in a storage pool on a Data Domain system. The relative size of protected data and incremental backup data, before and after de-duplication and compression is illustrated in the following figure.

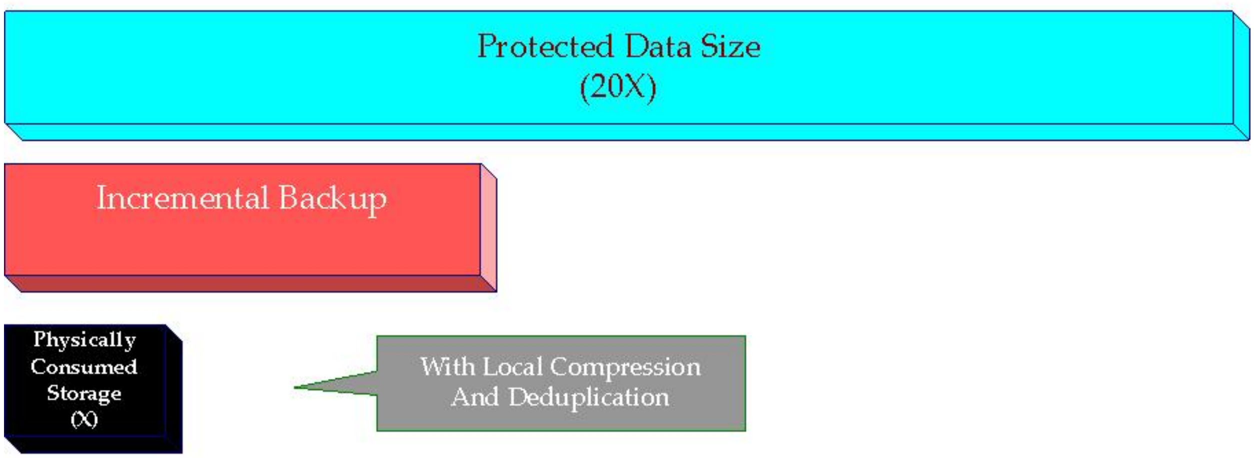

Figure 2: Backup Data Deduplication and Compression

Both deduplication and standard data compression (also referred as 'Local Compression' in product literature) are executed via loss-less compression methods (i.e. no data integrity impact). Lempel-Ziv (LZ) compression is standard, however GZFast or GZ are alternatives available to each Data Domain system instance for standard data compression. As always, backup data should not be compressed prior to attempting additional compression at the device level.

Data Domain Architecture and Models

A base Data Domain system supports a certain capacity of addressable storage (post-RAID, post-spares). Based on backup policy, this will enable 10x-30x more logical capacity. For example, a system that offers 10TB of addressable capacity would offer 100TB to 300TB of logical capacity. .

Each Data Domain system instance supports 200MB/sec average throughput. This base metric applies both to read and write operations, as the architecture is optimized for sequential I/O.

System Name	Physical Capacity (TB)	Logical Data Storage (TB)	Maximum I/O Performance (GB/Hour)
DDX (with 16 arrays)	504	8,800 – 20,000	12,800
DDX (with 8 arrays)	252	4,400 – 10,000	6,400
DD580	31.5	550 – 1,250	800
DD565	23.5	400 – 980	630
DD530	4.5	55 – 140	360
DD510	2.25	25 – 65	290
DD120	0.75	7-18	150

Table 3: 2007 Data Domain Systems, Addressable and Logical Capacity

Note: Logical Data Storage Values above reflect deduplication and compression effects on backup data. The actual values are highly dependent on rate of change and backup policies.

The solution scales modularly by incrementally adding either capacity to an existing Data Domain system instance in the case of the DD580 or the DDX, or adding a new Data Domain system instance to the NetWorker production environment. Multiple Data Domain system instances can be racked and managed through an enterprise console; however logical management of each Data Domain system instance is still required. The following figure illustrates Data Domain system architecture scalability.

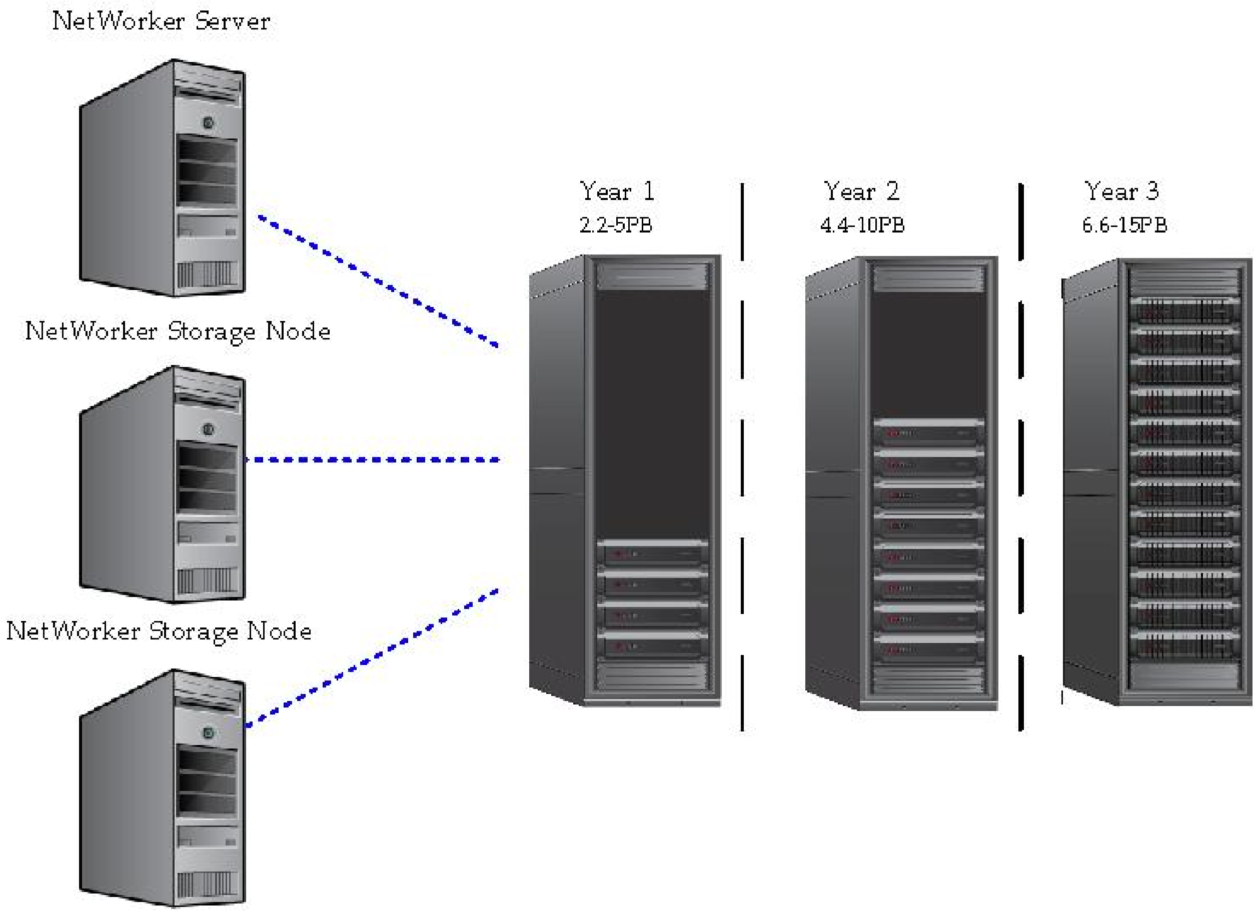

Figure 3: Data Domain System Architecture Scalability

File System and VTL Integration

Data Domain systems support two integration methods with NetWorker, either via network file system mounts or as a standalone Virtual Tape Library (VTL). Data Domain systems can run in a mixed mode capacity, providing both interface methods concurrently to one or many NetWorker server instances. This flexibility affords a great number of integration scenarios for NetWorker. The following figure illustrates both integration scenarios with NetWorker Servers and Storage Nodes.

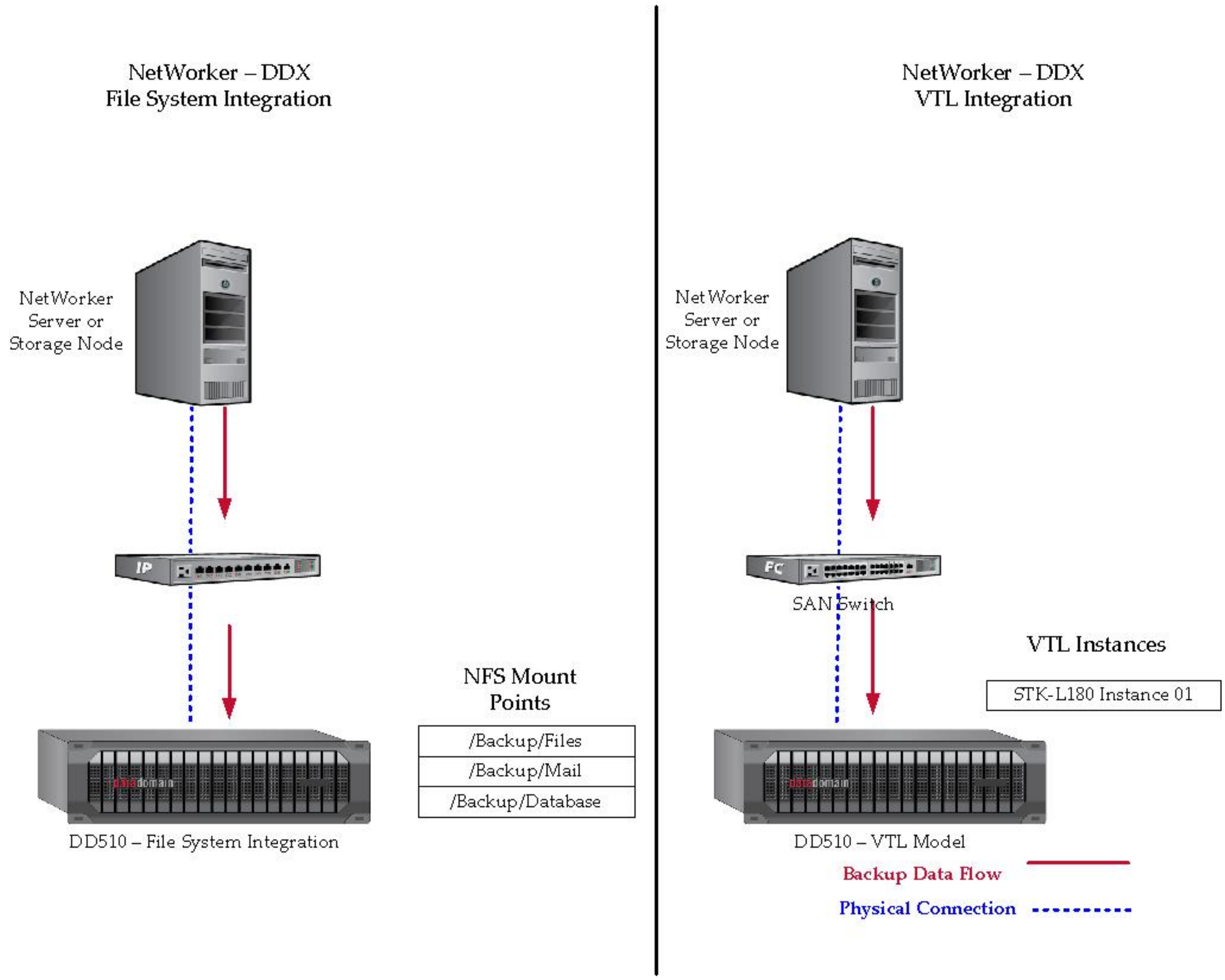

Figure 4: NetWorker - Data Domain System Integration

For network file system access, NetWorker addresses the Data Domain system via a native NFS mount or CIFS. NetWorker addresses the usable space exactly as it would a standard file system mount point (NTFS, JFS, UFS, etc.).

The VTL interface emulates a STK L180 tape library, and requires a fiber-channel connection along with the appropriate NetWorker device driver. NDMP backups are supported with a DDR attached directly to the NAS host via a fibre channel connection. Multiple instances of VTL can be created per Data Domain system instance. Up to 64 LTO tape drives, 10,000 slots, and 100,000 virtual cartridges can be created per Data Domain system instance. As a standalone VTL, existing physical tape resources can be leveraged by native NetWorker capabilities.

Replication

Asynchronous data replication is supported between Data Domain system instances. Once the initial mirror replica is established, only changes to index/metadata and new data segments are replicated to the target site. As a result, WAN bandwidth requirements are reduced by up to 99% and the amount of time to replicate data to an offsite location is reduced significantly

Replication is configured in Collection or Directory mode. Collection mode allows single Data Domain system instances, both NFS and VTL, to be configured in a source-target relationship, with one-way replication only. Directory replication supports many-to-one configurations which are established at the directory/mount level. Directory replication supports bidirectional replication between Data Domain system instances, which is ideal for various DR architectures, including hub-spoke architectures for remote offices. VTL instances emulate NFS Directory replication at the VTL pool level where options are set to indicate the source is a VTL pool.

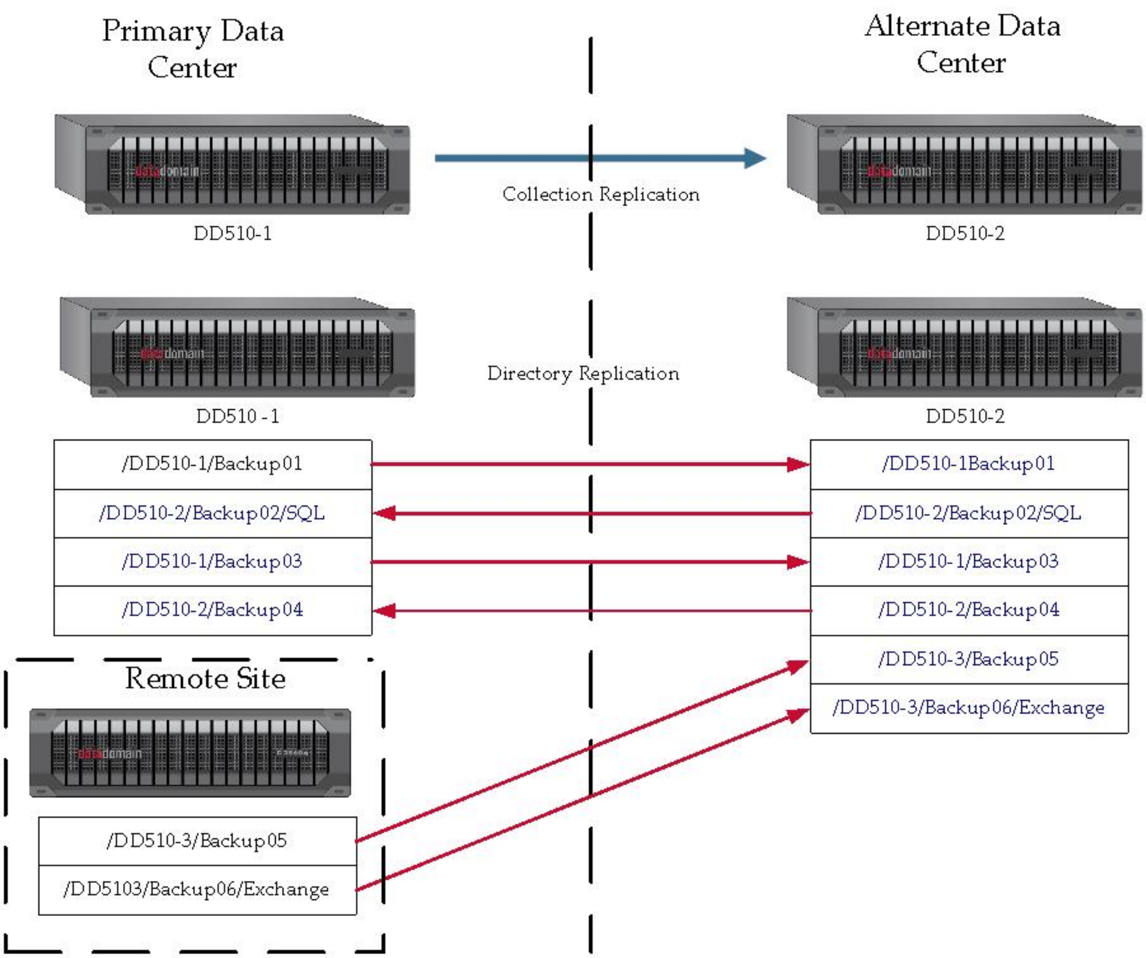

Figure 5: Collection and Directory Replication Modes

How Data Domain Best Fits with NetWorker

A Data Domain system provides an alternative for disk and tape volume pools in NetWorker. The Data Domain file system is optimized for sequential read and write operations. This provides for a great fit with existing NetWorker disk-based or VTL abilities. A Data Domain system is best configured as an advanced file type in NetWorker.

NetWorker databases should continue to be provisioned on traditional disk devices. A Data Domain system should not be used for storing active NetWorker databases, logs, and configuration files. Instead, these NetWorker elements can be backed up to a Data Domain system for operational recovery and replicated to a remote site for disaster recovery.

Note: Some NetWorker environments support extremely high-performance backups for high-volume clients. Typically, specialized designs are implemented to support backups of 1-4 TB/hour. The Data Domain system architecture can be configured to support high-performance workloads (via multiple save streams) with each Data Domain system instance supporting 200 MB/sec aggregate workloads per controller on currently shipping Data Domain systems. Because Data Domain's product architecture is CPU-centric, this number typically changes (upward) with new product releases in a given price band. The top end of the Data Domain controller line, with dual-socket Intel CPU components, has gone from 40 MB/sec. (DD200 in year 2004) to more than 200 MB/sec (DD580/g in year 2007), a factor of five increase over three years. Please check current Data Domain literature for current platform names and throughput.

Planning / Sizing Considerations

Backup Policies and Data Rate of Change

NetWorker policies are unique to each customer environment, but typically follow a common methodology. Most sites use a mixture of incremental backups with full backups run on a regular schedule (weekly or monthly). Incremental backups are more typical because they are faster and take up less space on the backup device. This leaves the full backups taking up more disk and tape space. Every full backup will write redundant data that exists in previous full backups, resulting in a large amount of the budget being lost to consumed disk capacity, more and unnecessary tape storage, offsite charges, and drive resource contention.

The impact of this redundant full backup data becomes much less significant when deduplication is introduced. Data Domain systems facilitate Synthetic Full backups, the goal of which is to create a full backup image from existing backup data. This process allows the NetWorker backup environment to benefit from an 'incremental forever' methodology without officially adopting such a scheme. In the end, though, the change rate of the data is the final arbiter for the amount of backup data stored.

From a NetWorker perspective, a database backup may appear 'net new' each time it is backed up, but from a Data Domain system perspective, the actual data changes may result in minimal new physical storage consumption. Databases, email, and unstructured data (file server data) will benefit the most from data deduplication in most production environments. Data growth issues are compounded by non-working copies of data used for reporting or testing, all of which are typically backed up daily by NetWorker. The net result is a never ending demand for physical storage resources. Data Domain counters the effects of uncontrolled backup data growth.

Deduplication benefits are realized over time and eventually plateau once the backup versioning policy and the incremental backup traffic is fully realized. Since data change rates vary by data type and production environment, a combination of backup policies, data change rate, and data structure impacts Data Domain system sizing estimates.

Sizing

Sizing storage capacities for a data deduplication solution takes into consideration actual data change rates, which are not visible from a NetWorker perspective since NetWorker views data change at the file/object level.

The ratio of current backup data to Data Domain system storage (after data deduplication + compression) varies, but NetWorker customers can expect on average 20x data reduction for typical backup environments. Again, this ratio is wholly dependent on the rate of data change, backup methods being used, and backup policy standards.

Note: For customers using VTL mode, Data Domain systems do not require space pre-allocation for virtual tape volumes. As virtual tape volumes are mounted and filled in the Data Domain system, physical space is not 'hedged' for scratch virtual tape volumes. Other VTL technologies pre-allocate physical tape volumes, regardless of whether or not they are empty, full, or active.

An initial sizing metric for the Data Domain system is to estimate the primary volume of data on all backup clients, or a 1:1 ratio of primary data (not to be confused with backup data) to the Data Domain system addressable capacity. Then run backups against the Data Domain system. Depending on retention policies, compression rates, change rates, etc, three to six months of backups can be retained on disk. As an example, an environment with 5TB of data requiring backups would need a Data Domain system with 5TB of storage. In this case, three to six months of backups could be stored on a single 5TB Data Domain system.

For new customers, Data Domain recommends ongoing capacity planning, through a discipline of sizing, provisioning, utilization, and ongoing measurement. This ideal capacity planning method includes ongoing measurement, and demand forecasting once a subset of production backup data is sent to a NetWorker server using a Data Domain system. Generally, the benefits of data deduplication are realized over time. Only then, do data volumes stabilize within Data Domain system instances.

Note: Although most NetWorker environments limit the effective number of backup save sets due to typical NetWorker setup (for example – 1 year retention policy, weekly full backups which equates to 52 backups per year) , the impact of additional backup save sets to Data Domain systems is often negligible. While this benefit allows more flexibility for aggressive backup policies, NetWorker database size should remain a key consideration in NetWorker policy planning.

Integration planning

A Data Domain system integrates into a NetWorker environment as the primary storage destination for directed backups. In either configuration, NAS or VTL, the Data Domain system can take advantage of several alternate agents and more efficient methodologies. Both instances take advantage of the speed of disk and easily integrate with a previously configured NetWorker environment as a VTL or as disk, an advanced file type device with a NAS configuration.

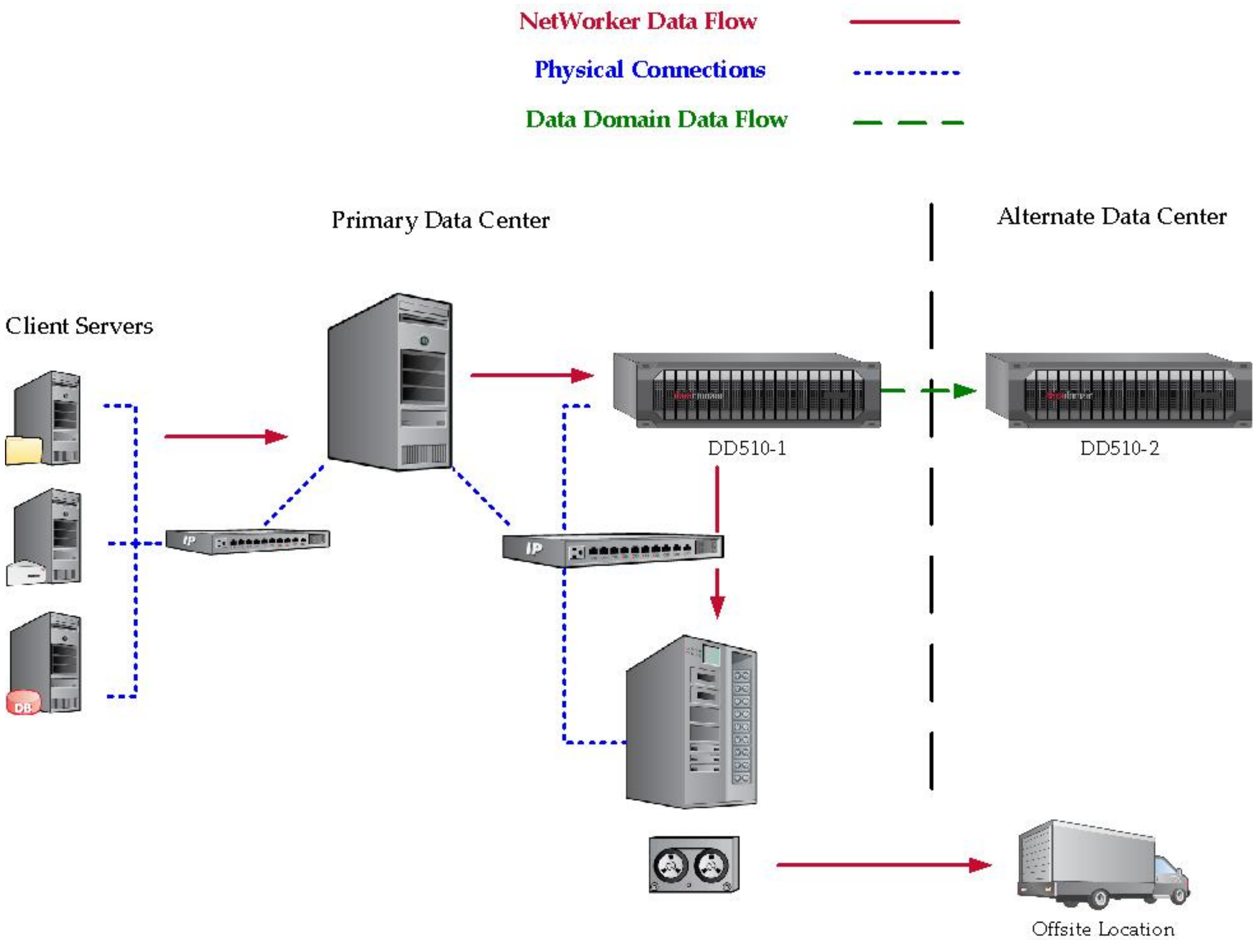

Figure 6: NetWorker Site Integration

One or more directories can be specified as the location of files for an advanced file type device. As a result, multiple Data Domain system instances can be dedicated to a single NetWorker server; however the benefits of data deduplication do not span multiple Data Domain system instances at this time. Pools containing 'like data' can be mapped to individual Data Domain system instances, or to individual mount points within a Data Domain system. For example, some NetWorker administrators create a group associated to advanced file type device (and pool) for all file servers, where a significant volume of primary unstructured data is redundant. Mapping this backup group and associated storage pool(s) to a Data Domain system expands the benefits of data deduplication above and beyond redundant backup copies. The following figure illustrates this concept.

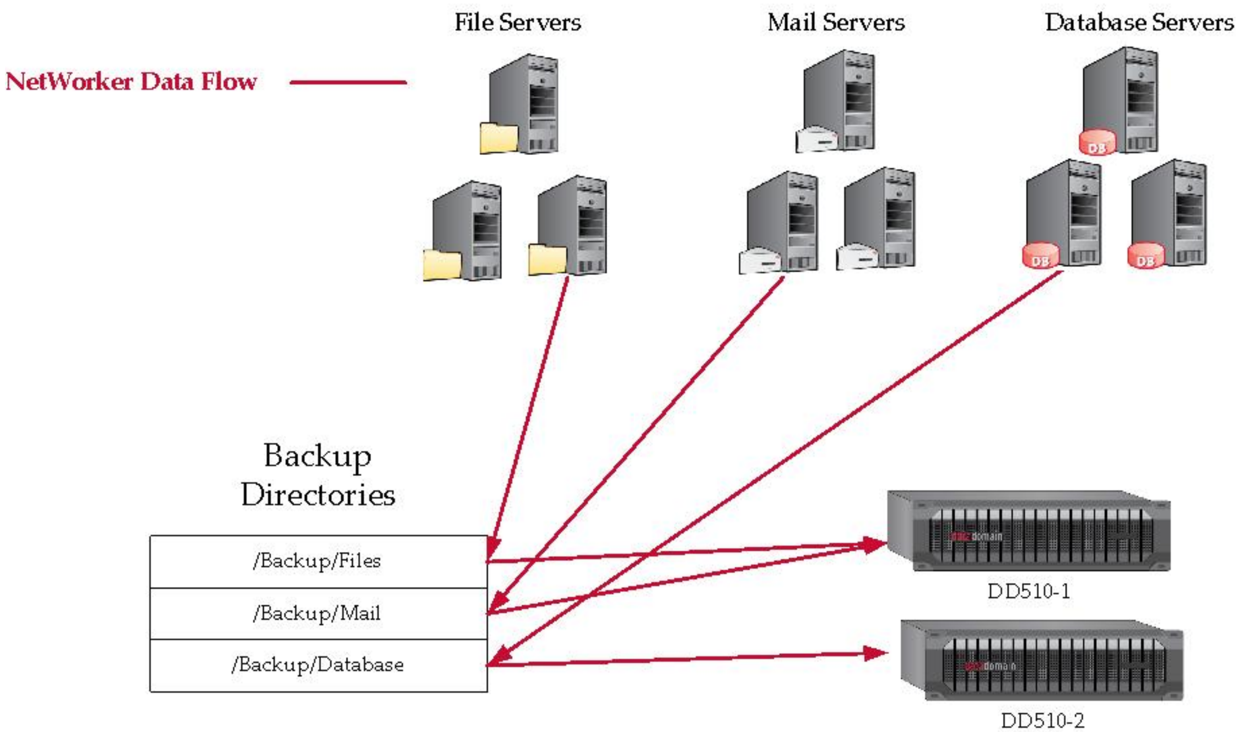

Figure 7: Mapping Data Types to Storage Pools and Data Domain System Instances

There are significant administrative benefits to mapping client data types to pools mounted to specific directories within a Data Domain system instance. These benefits include:

- More granular view of compression statistics, which are available on a directory by directory basis
- Ability to replicate specific directories
- Simplified administration associated with scaling to additional Data Domain system instances

If a Data Domain system instance reaches capacity, individual directories can be migrated to a new Data Domain system instance using replication, with minimal reconfiguration and downtime. For example, if all Oracle, Exchange, and file server backups are mapped to individual Data Domain system directories, a specific backup data type can be migrated to a new Data Domain system to reduce capacity in the original Data Domain system instance, allowing for additional capacity and growth.

Large NetWorker environments contend with a never-ending challenge of performance optimization for system resources. To avoid performance bottlenecking and to obtain optimal Data Domain system performance, multiple Data Domain systems can be deployed to support parallel workload / performance demands. The same architectural concept applies to traditional file system and physical disk layout planning for performance workloads.

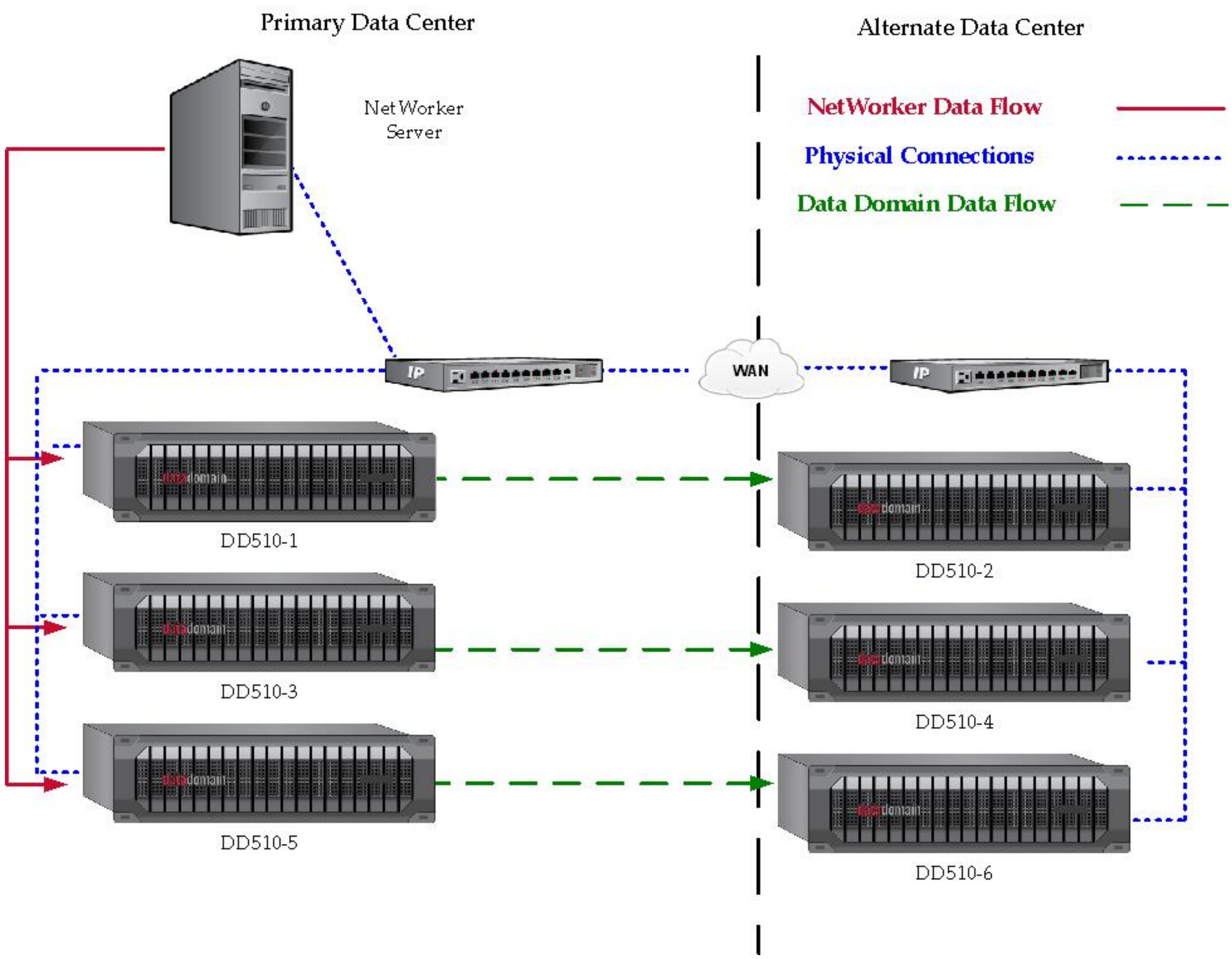

Figure 8: Large Site NetWorker Integration

Operational Considerations

When your infrastructure is configured optimally, administering a NetWorker Data Zone is relatively painless. When any aspect of the architecture, networking, firmware version, server capacity, etc, is misaligned it will show in daily success rates. A Data Domain system seamlessly installs into any architecture and will enable NetWorker administrators to more efficiently take advantage of the resources already in place.

Utilizing the advantages of the advanced file type device that a Data Domain system presents allows a NetWorker administrator to efficiently manage a timeline for daily operations, and successfully meet the timeline. These daily operations consist of the typical routines, outlined by the following diagram.

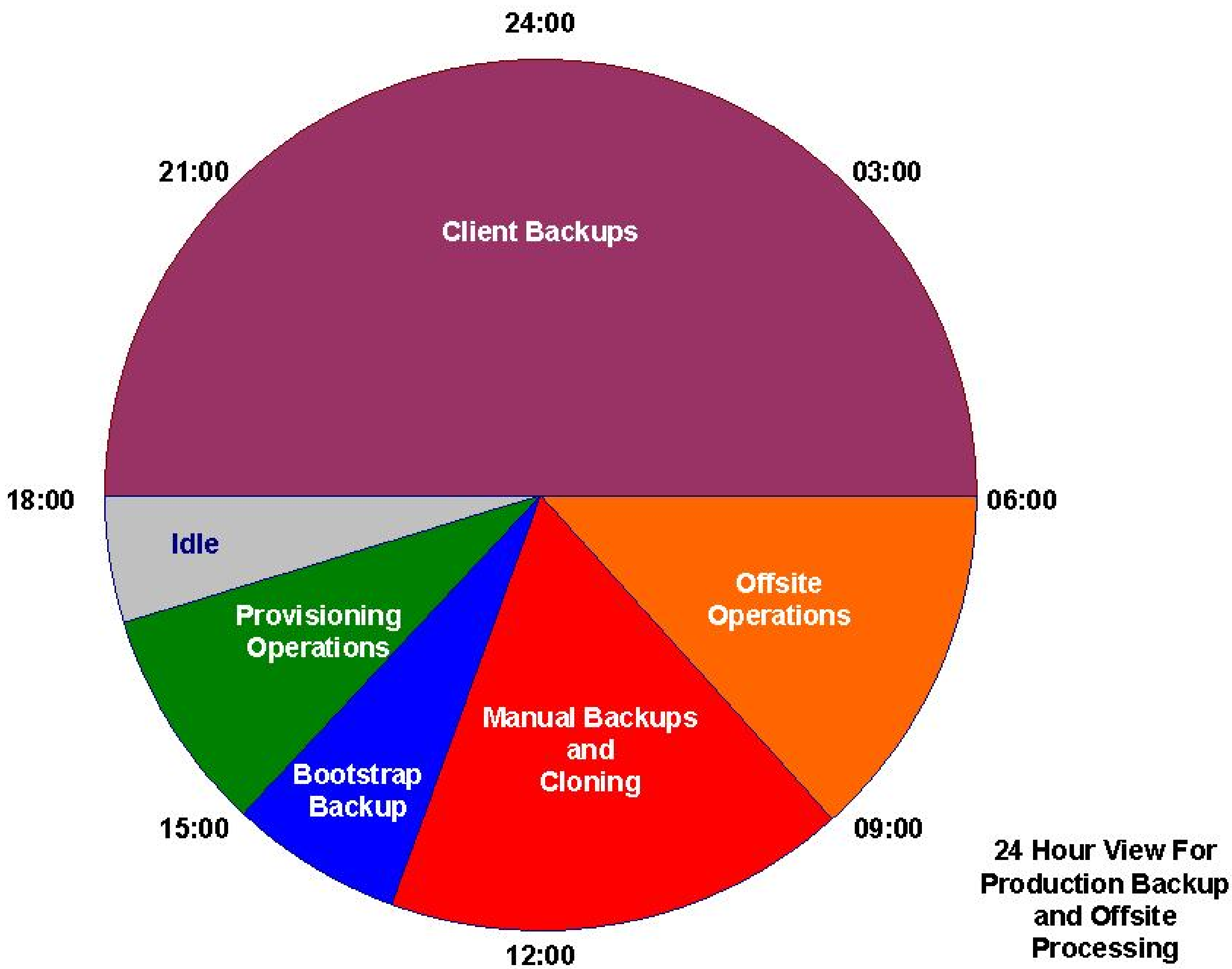

Figure 9: NetWorker Daily Operations

A Data Domain system instance can greatly reduce both the backup window and cloning process. With NetWorker staging or cloning you can easily offload to tape. However, if the Data Domain system architecture is leveraged for replication to offsite storage, any staging or cloning processes can be completely eliminated. Either implementation strategy significantly reduces the amount of time required to complete daily operations as illustrated in the following graphic.

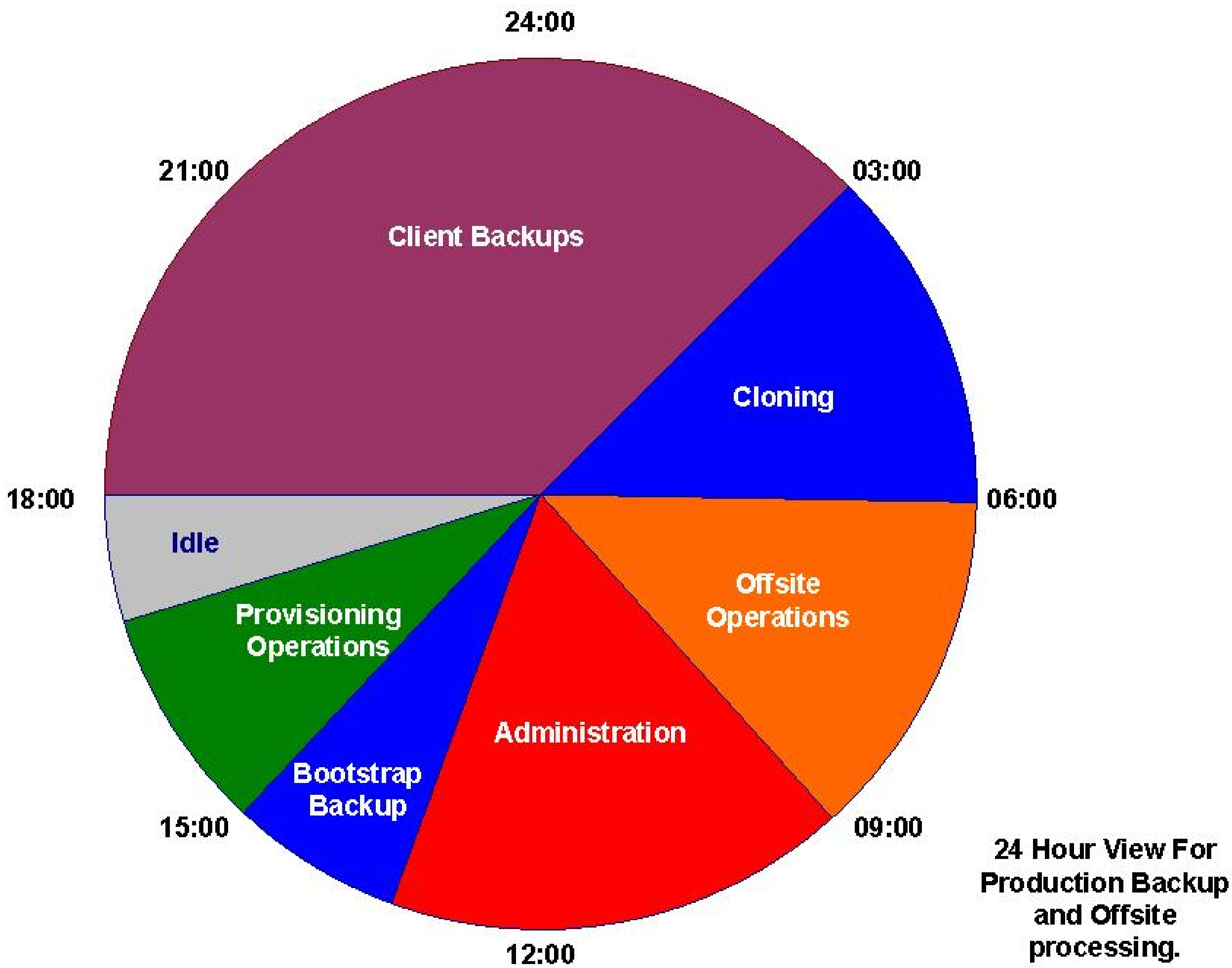

Figure 10: NetWorker Daily Operations with Data Domain

Customers will need to consider capacity planning and space management when implementing Data Domain systems with NetWorker. However, unlike managing disk arrays used for backup and especially tape, management of Data Domain deduplication systems is simple, making it much easier to achieve efficient utilization and ensure sufficient available space.

The Data Domain system requires an operational process to perform 'cleaning' of the Data Domain file system, which effectively reclaims space and optimizes the Data Domain file system for performance. The default schedule for 'cleaning' is to run weekly, however depending on data change rates and available Data Domain system space this process can be scheduled more frequently (2-3 times per week).

Note: The Data Domain system cleaning process can be resource intensive. A 'throttle command' is available to assign the relative priority to cleaning processes to normal backup and restore I/O. We recommend the Data Domain system cleaning processes run periodically as part of standard NetWorker operations, scheduled when backup/restore activity is at minimum levels. See the DD OS Admin Guide for more discussion.

Recovery Considerations

With a tape only backup infrastructure, restore activities are hampered by all of the limitations that come with trying to read data stored on a mechanical device. With the Data Domain system and inline native deduplication, NetWorker sites can leverage an 'incremental forever' approach for production backups without having to change their current methodologies. If offsite cloning is configured to go to tape, an environment will still suffer the delay and charges for tape recalls which put pressure on any application data RTO (Recovery Time Objectives). If the administrator takes advantage of Data Domain's ability to replicate to another Data Domain system at an alternate location, a much shorter RTO is achievable. Data Domain inline depulication also provides better RPO (Recovery Point Objectives) as compared to typical implementations of post-processed deduplication.

The combination of the NetWorker Advanced File Type Devices and the Data Domain systems is the very purpose of backups – to enable faster, more reliable recoveries. By recovering data at the speed of disk, NetWorker Advanced File Type Devices can deliver vast improvements in recovery performance and reduction in downtime costs.

Direct onsite recovery conditions improve first and foremost due to a Data Domain system based on disk hardware. Disk is much more reliable than traditional tape mechanics, and offers much faster search and read times without any delays due to robot mounting /positioning functions. Data Domain takes this one step further with its Data Invulnerability Architecture, a system design that extends the resiliency of the system well beyond that typically found in this class of storage. Backups directly to a Data Domain system also eliminate the complexity of multiplexing and the inconvenience of incremental backups that span across multiple physical tape volumes. Consequently, multiple tape mounts which increase restore operation times in large restore operations will not be required.

NetWorker Server Tuning

NetWorker Server tuning is recommended for new Data Domain system implementations using NFS/CIFS and IP protocol.

Note: Some NetWorker Server configuration changes require a NetWorker server refresh to take effect.

The following table outlines standard server tuning parameters for optimal Data Domain system performance with NetWorker, using NFS/CIFS and IP protocol.

Configuration Type	Comments
AIX – Network Configuration	ifconfig en0 tcp_recvspace 65536 tcp_sendspace 65536 tcp_nodelay 1 tcp_nodelayack sb_max
HPUX – Network Configuration	Enter the following two commands then remount the Data Domain system NFS share to enable the values: ndd -set /dev/tcp tcp_recv_hiwater_def 262144 ndd -set /dev/tcp tcp_xmit_hiwater_def 262144
SOLARIS – System Settings	Create a file /etc/rc3.d/S90ddr. Enter the following two lines in the file: ndd -set /dev/tcp tcp_recv_hiwat 131072 ndd -set /dev/tcp tcp_xmit_hiwat 131072 In the file /etc/system, add the following lines: set nfs:nfs3_max_threads=16 set nfs:nfs3_async_clusters=4 set nfs:nfs3_nra=16 set rpcmod:clnt_max_conns=1 set fastscan=131072 set handspreadpages=131072 set maxpgio=65536
LINUX – Server Settings	echo "4096 262144 1048576" > /proc/sys/net/ipv4/tcp_rmem echo "4096 262144 1048576" > /proc/sys/net/ipv4/tcp_wmem echo 262144 > /proc/sys/net/core/rmem_max echo 262144 > /proc/sys/net/core/wmem_max

Configuration Type	Comments
	echo 262144 > /proc/sys/net/core/rmem_default echo 262144 > /proc/sys/net/core/wmem_default
WINDOWS – Network Configuration	**Note:** Do not modify the Windows registry parameter 'AFD' if the NetWorker server or any associated NetWorker clients are supported by Windows NT 4.0 1. Open REGEDT32 and navigate to: HKEY_LOCAL_MACHINE\SYSTEM\CURRENTCONTROLSET\SERVICES\ AFD\PARAMETERS 2. Add a new DWORD value to the DefaultSendWindow key and set the value to 65536 (decimal). 3. Add a new DWORD value to the DefaultReceiveWindow key and set the value to 65536 (decimal). 4. Within REGEDT32, navigate to the following location: HKEY_LOCAL_MACHINE\SYSTEM\CURRENTCONTROLSET\SERVICES\ TCPIP\PARAMETERS 5. Add a new DWORD value to the GlobalMaxTcpWindowSize key and set the value to 65536 (decimal). 6. Add a new DWORD value to the TcpWindowSize key and set the value to 65536 (decimal). 7. Add a new DWORD value to the Tcp1323Opts key and set the value to 3. 8. Restart the Windows server.

Table 14: NetWorker Server Configuration Guidelines for NFS/CIFS and IP Protocol

Network File System Integration

NFS mounts require IP protocol and a dedicated Gigabit VLAN or direct Gigabit, connection for NetWorker server integration.

Data Domain recommends the following general NFS configuration settings for mounting a Data Domain system to a NetWorker server.

NetWorker Server Platform	NFS Configuration
AIX	mount –v nfs –o proto=tcp,vers=3,intr,hard,combehind,rsize=32768,wsize=32768,llock –n dd200 /backup */mount-point*
Solaris	mount -F nfs -o hard,intr,vers=3,proto=tcp,rsize=32768, wsize=32768 *restorer-name*:/backup */mount-point*
HP-UX	mount –F nfs –o rsize=32768,wsize=32768,hard *restorer-name*:/backup */mount-point*
LINUX	mount -t nfs -o intr,hard,rsize=32768,wsize=32768,proto=tcp,vers=3 ddr:/backup /dd/<mount point>
Windows	CIFS access to Data Domain system is recommended for Windows NetWorker Servers

Table 6: NFS Mount Configuration Guidelines

Note: Data Domain only supports TCP protocol with NFS, and recommends hard-mounts to ensure availability after NetWorker server outages.

Advanced File Type Design Considerations

NetWorker has supported backups to disk device from as early as version NetWorker 5.x with the file-type device option. NetWorker 7.0 included the advanced file type device (which allows simultaneous reads/writes to the device). Both are relatively simple to configure and can be used with staging and/or cloning.

The advanced file type device operations include reclaiming disk space from expired and aborted savesets to optimize the use of disk storage. The advanced file type device is optimized for very large disk devices and is never marked full as would be encountered in tape or file type devices when the volume reaches a full capacity. When the advanced file type device runs out of space, the current backup is put on hold and a notification is generated (this can be configured via email, console or log notifications), indicating that it is waiting for more space. NetWorker will then begin to delete expired save sets and reclaim space until enough data is available for backup to continue. When enough space is available, the backup will continue.

Note: Freeing up space within the Advanced File Type devices does increase the backup window therefore NetWorker Administrators should always monitor the size of the device if it is close to full and meeting the backup window is a large concern.

Within NetWorker, the creation of an advanced file type device appears as two seperate devices. When the advanced file type device is labeled, the NetWorker Server software creates a secondary device with read-only accessibility. The NetWorker software then creates and mounts a volume with a .RO suffix in this device.

Note: The secondary device is read-only. This enables concurrent operations (reading) from the secondary device. NetWorker Administrators should not make any changes to the *_AF_readonly* device resource. Changes made to the primary volume or save set, are automatically propagated to the secondary volume or save set.

NDMP Integration

NDMP protocol is standard with Network Attached Storage (NAS) devices industry wide. Following standard integration techniques used with NetWorker and physical tape libraries, the Data Domain system provides the same functionality for NDMP integration.

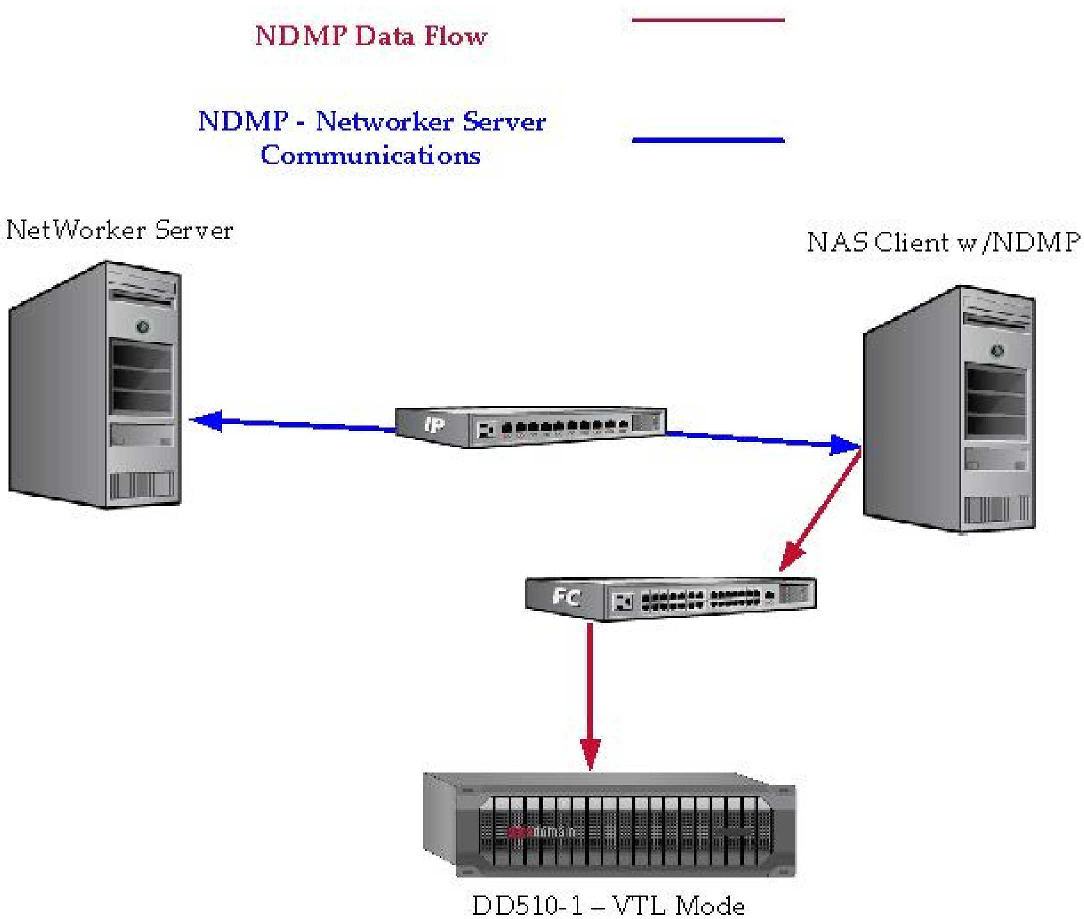

Figure 11: NDMP Integration 1 with NetWorker and Data Domain System

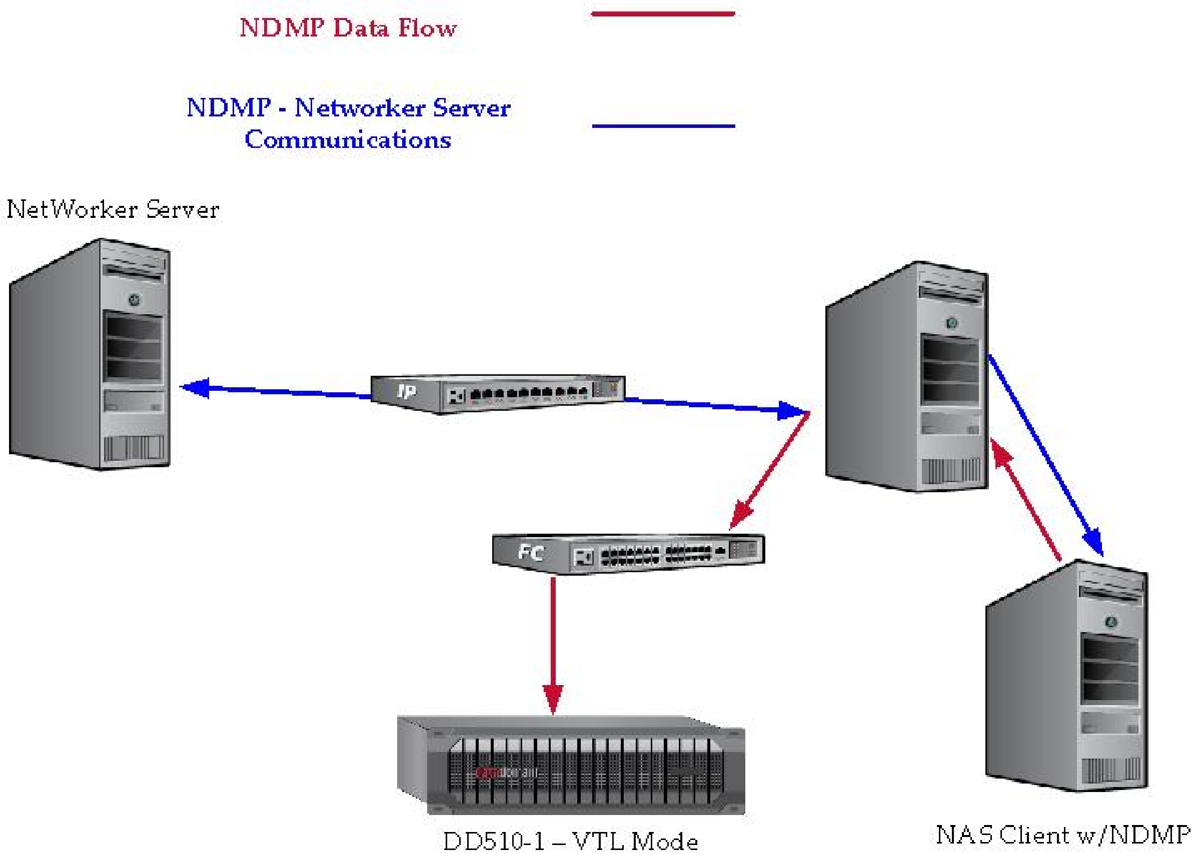

Figure 12: NDMP Integration 2 with NetWorker and Data Domain System

Archiving Integration

Long term data retention in NetWorker is accomplished either via traditional backup or archive functionality with Disk Extender or Email Extender. An Extender based long term backup solution automatically manages the duplication of data across tiers of archive storage, each with a set retention. The Disk and Email Extender solutions provide for a fully functional archiving solution.

For NetWorker sites desiring long term retention a Data Domain system can be leveraged as an upper tier for archive data. The following diagram illustrates NetWorker data movement to a Data Domain system long term tier via Disk Extender, which is configured to periodically migrate data to an archive tape pool for extended retention.

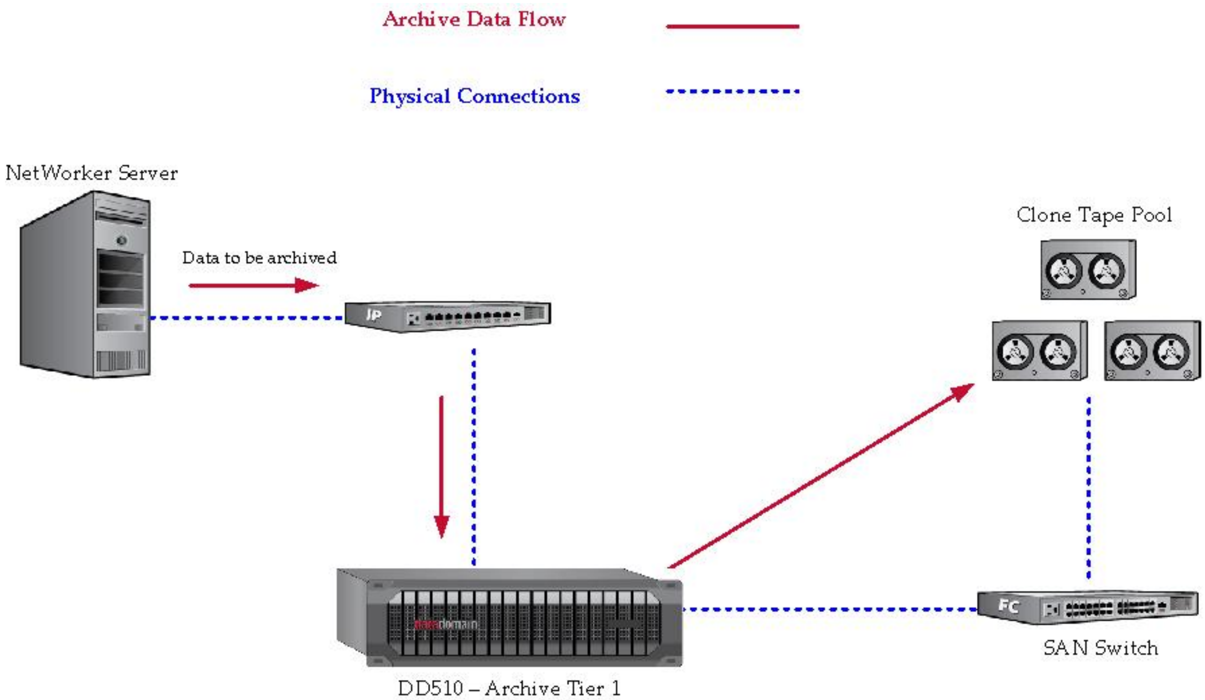

Figure 13: Archive Integration with NetWorker and Data Domain System

NetWorker Database (Bootstrap) Backups

NetWorker bootstrap backups are critical to the recovery of a NetWorker server. A NetWorker bootstrap backup can be done after regular backups by placing the NetWorker server in a production group. The bootstrap is comprised of three components that reside on the NetWorker server: the media database, the resource database, and the NetWorker server's client file index. Bootstrap backups are usually written to a local tape device and it is a good idea to have the NetWorker bootstrap backups complete at least twice a day. One backup should be sent offsite for disaster recovery purposes and the other should be kept onsite for operational recovery. The second bootstrap tape can be created using customized cloning scripts automatically run daily (using a savegroup –O on the group the server is assigned to). In addition, a bootstrap report can be created with vital information to rebuild your NetWorker server.

When a Data Domain system is added to the NetWorker data zone, NetWorker bootstrap backups can be written directly to the Data Domain system. When NetWorker bootstrap backups are sent to the Data Domain system, storage space is not wasted. If a second Data Domain system is introduced at an alternate site, the NetWorker bootstrap can be replicated using Data Domain Replicator Software. This setup will eliminate all of the inefficiently used bootstrap backup tapes. It will also reduce the amount of NetWorker server time required to write two bootstrap backups daily, by leveraging Data Domain replication for the offsite database backup copy.

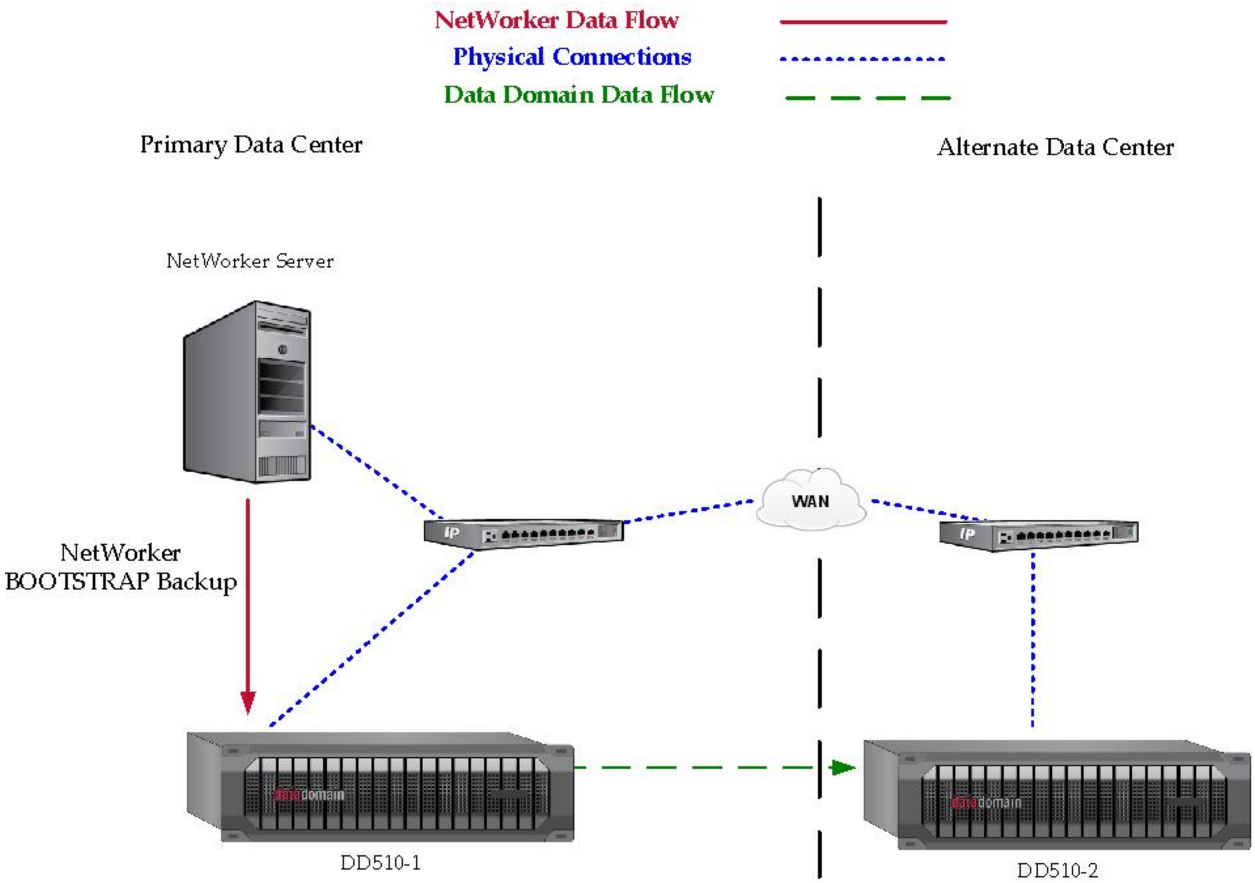

Figure 14: Bootstrap Example

NetWorker Disaster Recovery

Several architectural scenarios employ the use of a Data Domain system as a central mechanism for replicating NetWorker data to an alternate site for purposes of disaster recovery. As with any other method of performing NetWorker disaster recovery operations, the NetWorker database, and configuration files must be available for recovery operations.

The MMRECOV procedure is used to recover a NetWorker server to its original state prior to the disaster. The general sequence of events to recover a NetWorker server includes the following steps.

Note: To start the recover process you will need a new server running the same OS and patch level as the production NetWorker server and it will need to be connected to the Data Domain system at the DR site.

1. Install the NetWorker server software.

2. Configure the NetWorker server to see the Data Domain file systems, they must be configured identical to the way they were in production.

3. Ensure that the disk where you are restoring the bootstrap files contains the directory where the catalog previously resided. Recreate any symbolic links to catalog locations that may have existed

4. Use “MMRECOV” to restore the database

5. Restart the NetWorker services on server.

6. Inventory required media (only if physical tape is managed by NetWorker)

7. Restore the Indexes of the clients (if needed)

8. Resume backup and restore operations in alternate backup environment

Before running any disaster recovery plan it is best to refer to the relevant NetWorker documentation and test before implementation in production environments. A Data Domain system eliminates the need to go to tape at the primary site, and altogether when using an alternate site instance with replication. For larger catalog instances this will decrease the RTO significantly.

Architectural Scenarios

Local Data Domain Primary Storage with Manual Offsite Tape

In the scenario illustrated in the following figure, a Data Domain system instance is used to replace traditional NetWorker DiskBackup option..

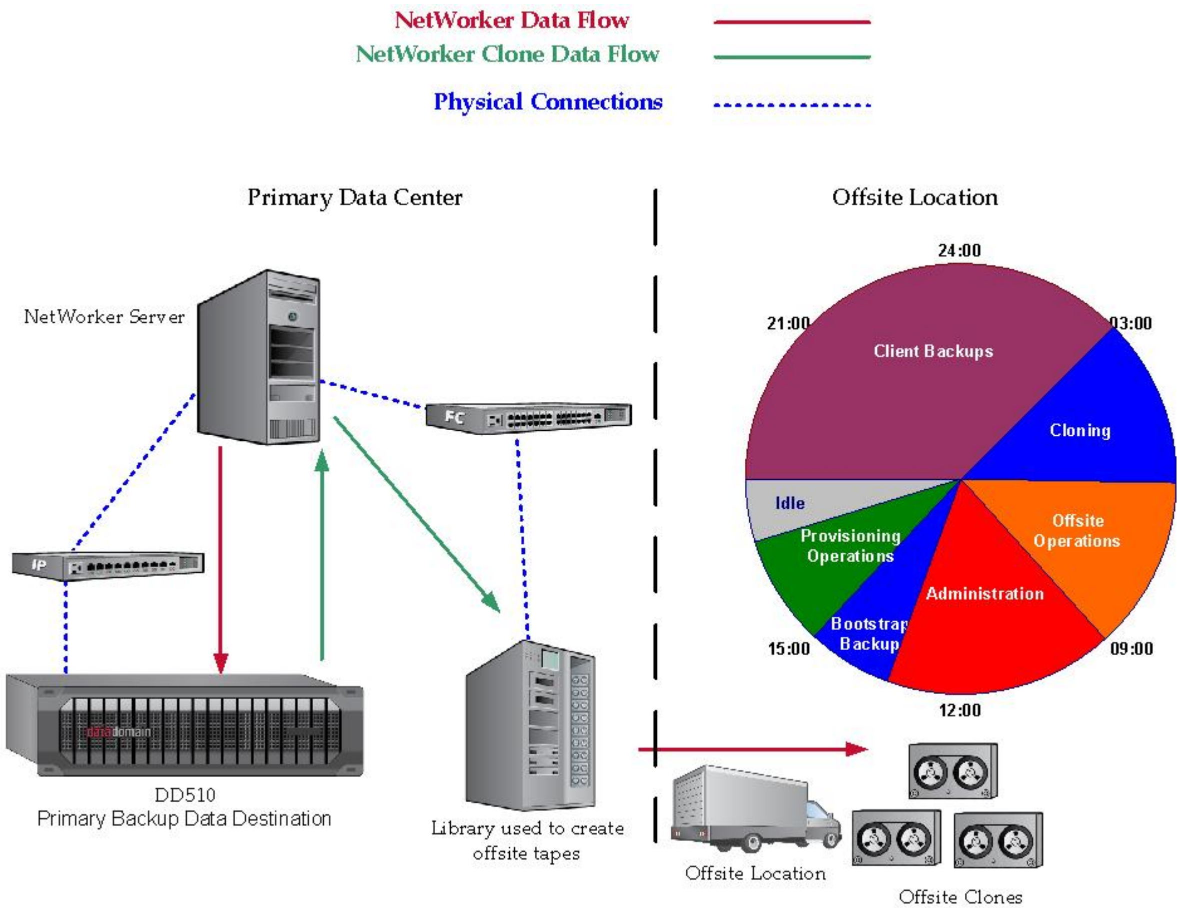

Figure 15: Local Data Domain System Primary Backup Storage with Manual Tape Vaulting

A tape library is used to create physical tapes for bootstrap backups and the tapes are manually transported offsite via courier. Due to the use of a Data Domain system as the primary storage pool for backups, the migration step has been removed from the daily batch processing tasks, shortening the daily operational workload.

Local Data Domain Primary Backup Storage with NetWorker cloning

In the scenario illustrated in the following figure, a Data Domain system instance is used to replace traditional NetWorker disk and tape cloning. A tape library is accessible via extended SAN at the alternate site, and is used for daily tape cloning.

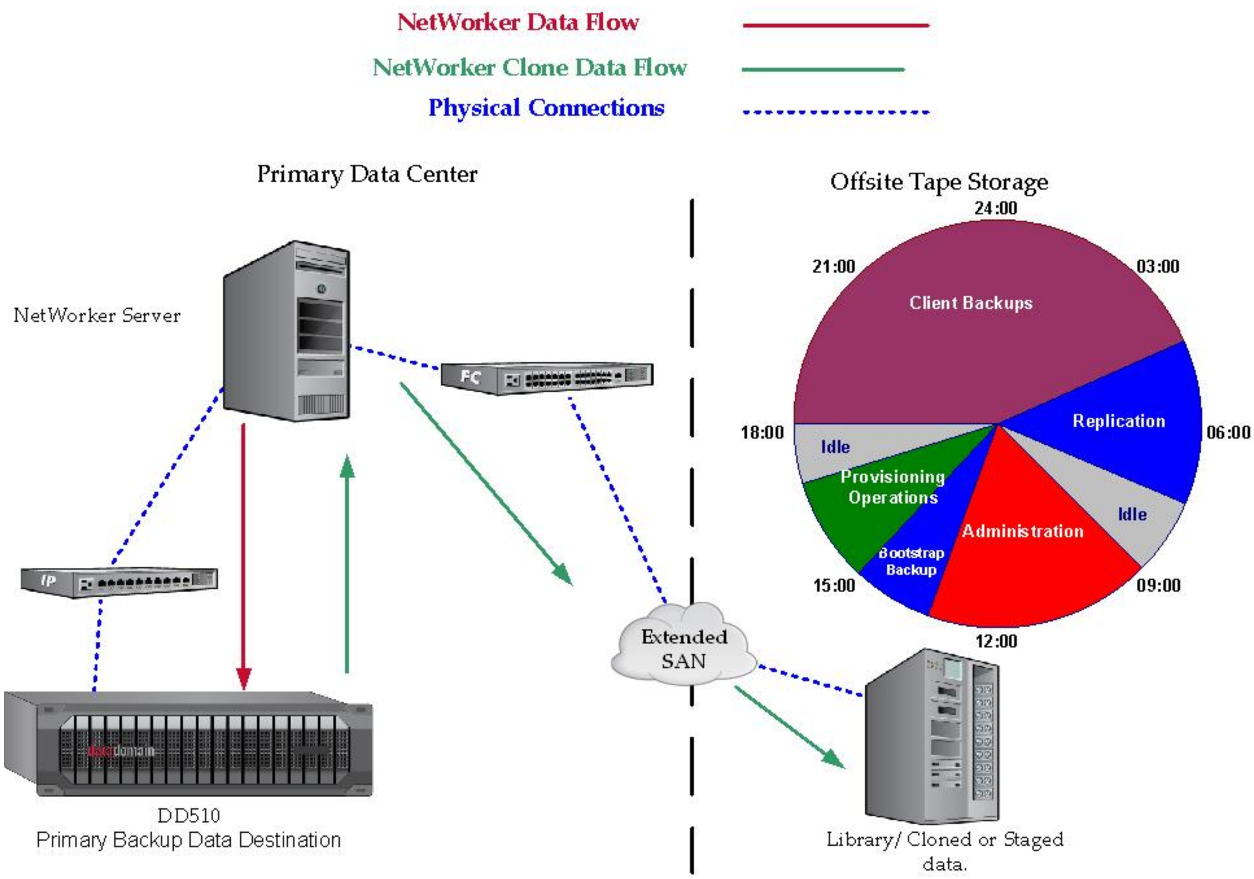

Figure 16: Local Data Domain System Primary Backup Storage with NetWorker Cloning

Again, tape cloning and staging has been removed from the daily tasks, shortening the daily operational workload. This scenario benefits sites with existing tape cloning or staging and offsite tape infrastructure in place.

Local Data Domain Primary Backup Storage with Data Domain System Replication to Alternate Site

In the scenario illustrated in the following figure, physical tape is eliminated by using a Data Domain system at the primary data center to hold the primary storage pools. Another Data Domain system is placed at the alternate datacenter and data is replicated between data centers using Data Domain Replicator Software. The use of the Data Domain replication dramatically reduced the network bandwidth needed for electronic vaulting of data.

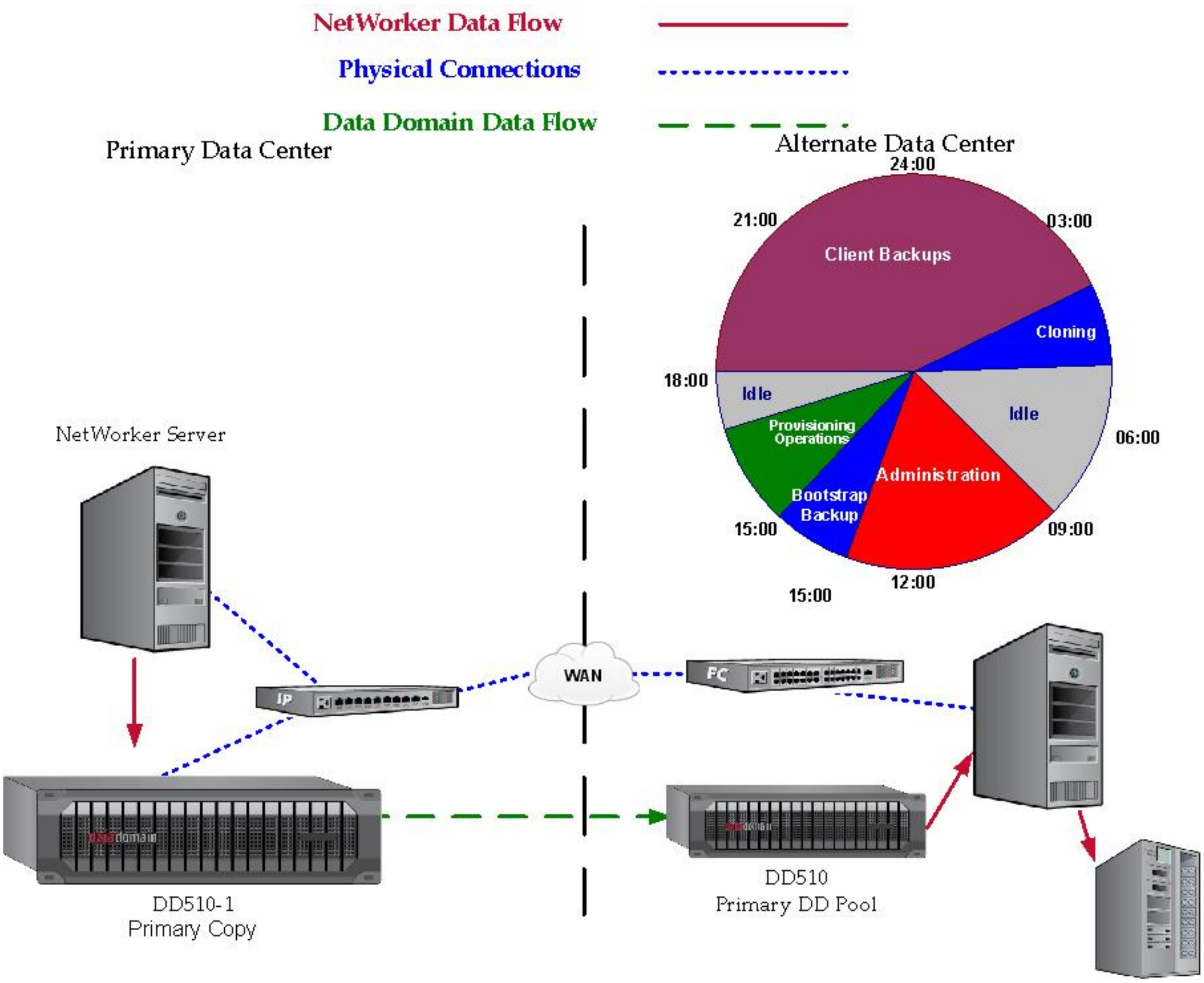

Figure 17: Local Data Domain System Primary Backup Storage with Replication to Offsite Data Domain System

By utilizing the Data Domain systems at both the primary and alternate data center, the daily operational tasks for backup storage pools are reduced.

Mixed Production / DR with Data Domain Replication

In the scenario illustrated in the following figure, multiple NetWorker servers clone data to the Data Domain systems at the opposite site.

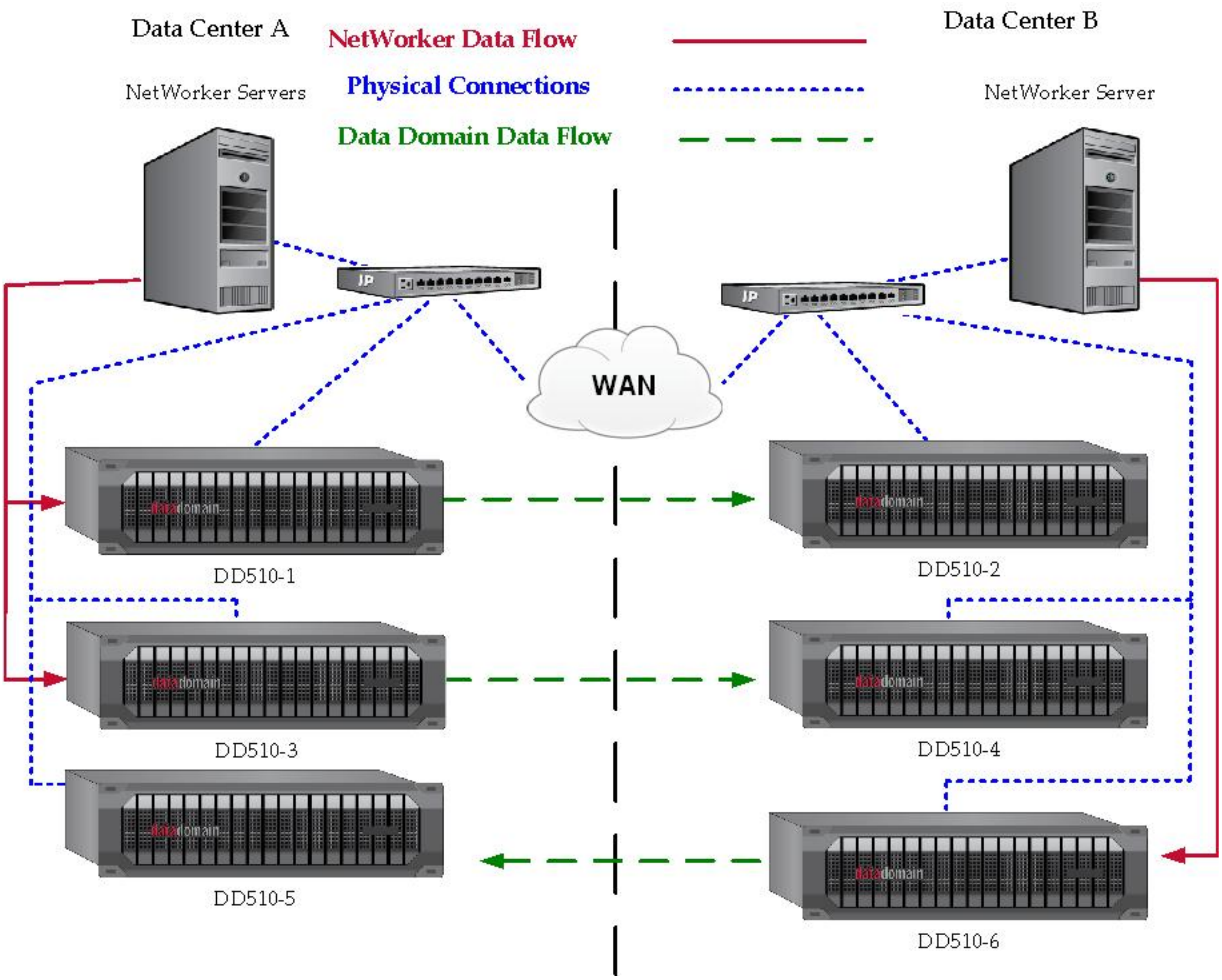

Figure 18: Mixed NetWorker Production / DR with Data Domain Replication

Each data center supports a mixture of production and disaster recovery capacity, and each NetWorker server clones data to the alternate site using Data Domain replication.

Central Disaster Recovery Data Center Configuration

In the scenario illustrated in the following figure, a centralized DR site supports multiple production data centers, by replicating NetWorker backup data to the DR site via Data Domain replication.

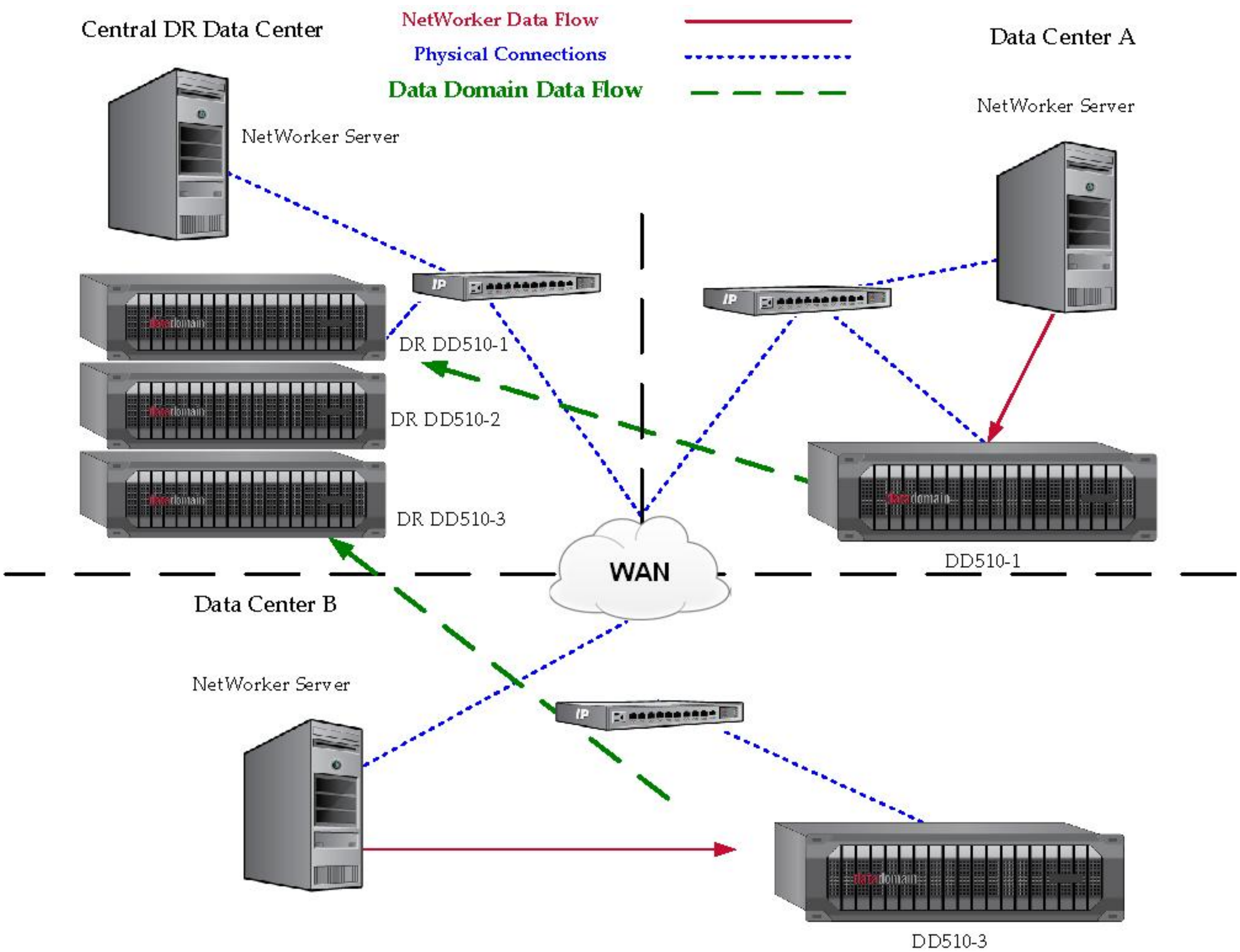

Figure 19: Central Disaster Recovery Data Center Configuration

Remote Office Local and Disaster Recovery Configuration

In the scenario illustrated in the following figure, a centralized DR site supports multiple remote sites, which replicate NetWorker backup data to a single Data Domain system instance the DR site using Data Domain replication.

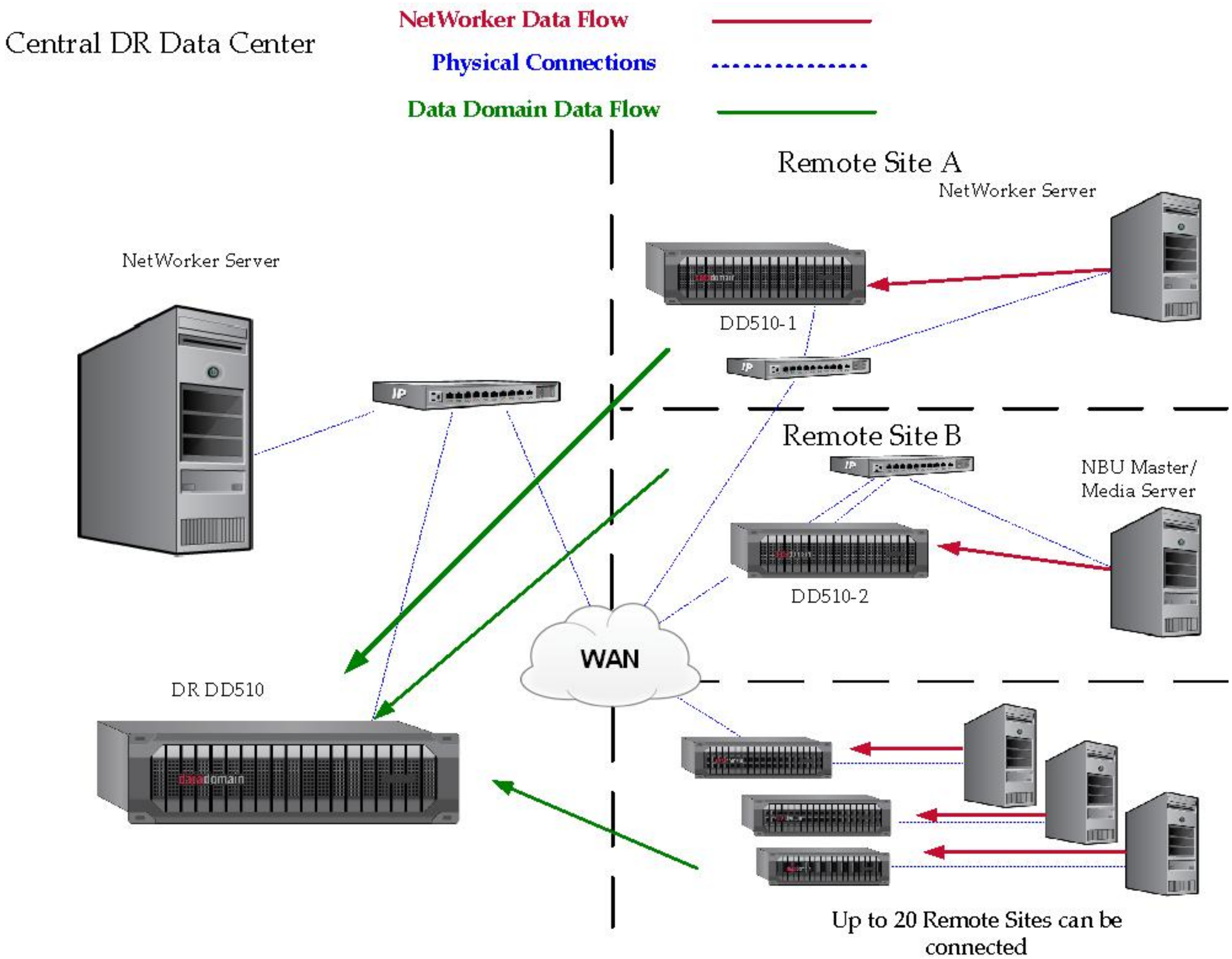

Figure 20: Remote Office Local and Disaster Recovery Configuration

Conclusion

Data Domain inline deduplication storage offers a variety of new storage architecture strategies for NetWorker customers. The NetWorker application functions 'as-is' with the Data Domain system interface methods, minus the complexities and management overhead associated with physical tape media. For the enterprise, disaster recovery replication of NetWorker data has traditionally required significant architectural measures and management overhead. Data Domain replication provides a viable offsite disaster recovery and longer term retention alternative.

We see a particularly compelling case for Data Domain to revolutionize the way small, medium, large and remote NetWorker sites manage physical storage resources and replication for disaster recovery. For extremely large NetWorker instances, appropriate planning, sizing, and integration strategies make the Data Domain system a viable solution for the enterprise.

About GlassHouse Technologies, Inc.

GlassHouse Technologies is the industry's leading independent consulting firm with proven experience transforming IT infrastructure. GlassHouse's proprietary methodology aligns business processes and information technology systems, transforming our client's existing infrastructure into scalable, compliant, cost-efficient and tightly organized environments. GlassHouse consultants architect, implement and operate IT environments to drive high performance and agility. GlassHouse clients include, Allianz, Morgan Stanley, Aetna and Wells Fargo.