Symantec NetBackup (NBU) Design Best Practices with Data Domain

GlassHouse Whitepaper

Written by:
Brian Sakovitch and Kelley Alexander
GlassHouse Technologies, Inc.

Introduction

Protecting the ever expanding data that collectively is the information that businesses rely on for growth has come to the forefront of the IT industry. Rarely does a company reduce their net storage capacity or the amount of data they retain, even in the face of stiff budgetary constraints. And though more cost effective technologies have been developed, completely and effectively protecting what is inevitably many copies of the same data is a major challenge for many businesses.

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

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

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

NBU Architecture and Terminology

NetBackup (NBU) is a client–server software solution designed for enterprise customers. The core product functionality includes backup/recovery, archive/retrieval, and disaster recovery. NBU allows for a flexible solution. The following figure illustrates a typical NBU architecture.

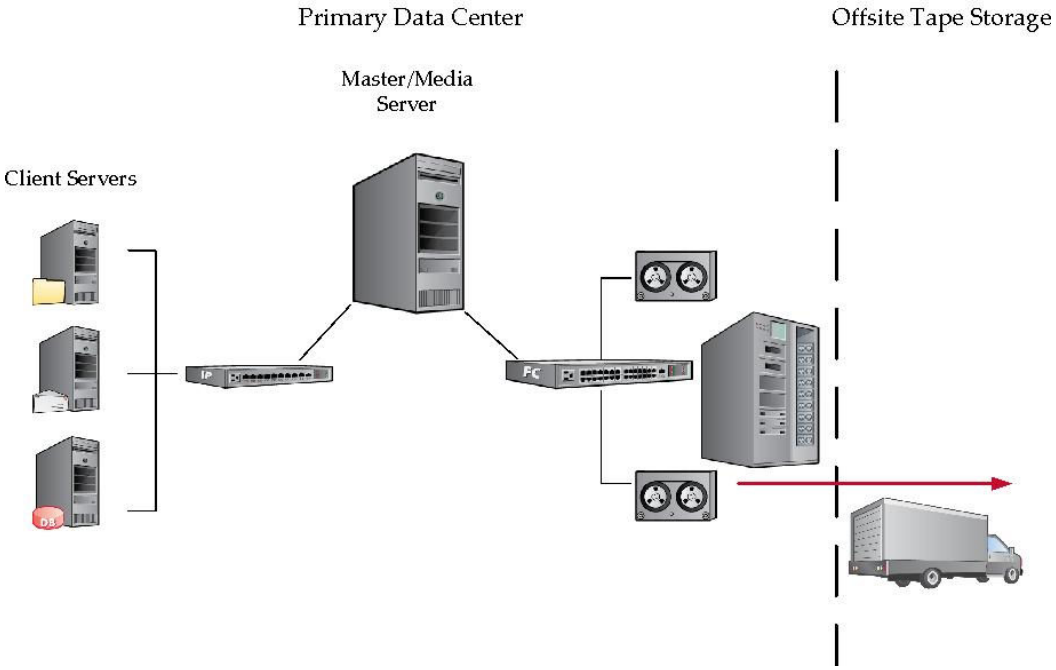

Figure 1: Common NetBackup Architecture

Each NBU master server records and manages a backup image Catalog. An environment can be configured with a number of media servers, or one server acting as both a master and media server, managing shared or individual pools of storage units. These storage units include a variety of platforms that allow for both onsite and offsite data copies to be made. More commonly, these platforms are tape libraries copying from one onsite tape to another sent offsite. Common terminology used in this whitepaper is provided in the following table.

Term	Definition
Master Server	The primary server in an NBU environment controlling scheduling, tracking client backups, managing tape media and the NBU catalog (see below). This server can also be configured as a Media Server.
Media Server	This server manages the data movement between client and storage destination.
Catalog	A group of files containing vital information about performed backups. Included is information such as configuration, status, errors, what files and folders have been backed up, and tracking information on the location of the backed up data.
Full Backup	A backup that attempts to record all data in the set file list.
Differential Incremental Backup	A backup that only records data that has changed since the last backup.
Cumulative Incremental Backup	A backup that records data that has changed since the last Full backup.
Storage Unit	A device where backup data is sent to be stored. This can either be a library with a number of tape drives, direct attached tape drives, a configured disk pool, etc.
Storage Unit Group	A logical entity that pools the resources of a number of storage units for use by NBU policies.
Policy	The information provided to run a backup; client, file list, retention, schedule, etc.
Inline Copy	A backup duplicate made synchronously with the backup. This type of duplication relies on the throughput speed of the slowest Storage Unit resource.
Vault	A NBU agent option that allows the configuration of logical profiles designed to automate the duplication of backups, perform catalog backups, eject media, and generate reports.
Volume Pool	Groupings of available storage that is free for use, or is in use, by NBU backup data. Typically this refers to groups of tape media and configured disk.
High Water Mark	This value is the capacity level for a Disk Storage Unit at

Term	Definition
	which, upon being reached, the NBU administrator is notified or staging to tape resources begins.

Table 1: NBU Terminology

Typical NBU Challenges

The typical NBU environment supports anywhere from a handful to thousands of clients. NBU scales by adding additional storage unit resources, and associated media server resources to manage them. The most common challenges in NBU environments include:

- Leveraging available storage unit resources for maximum utilization and throughput.
- Contending with extremely large client backups
- Eliminating redundant data backups (copies of databases, aggressive backup retention policies, etc.)
- Eliminating performance bottlenecks (NBU servers, networking, client, tape drives, etc.)
- Managing and storing large amounts of tape media
- Lack of reporting disciplines.
- Keeping up with catalog growth.

Technology Overview

Data Domain reduces unnecessary NBU 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 "active" set of primary data being backed up.
Logical Storage Size	The total size of all backup images on a Data Domain system.
Disk Pool Dump Size	The size of an individual backup image set written to a storage unit (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 space.

Term	Definition
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 Restorer 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 NBU 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.

Since the rate of primary data change (newly introduced unique 4K to 16K segments) at most sites stays about the same from night to night at most sites, the amount of physically consumed storage for subsequent full NBU backups is roughly the same as the physically consumed storage for a differential incremental NBU 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 full backup (because the former is much smaller in size). As a result, it is very inexpensive to include many versions of files 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.

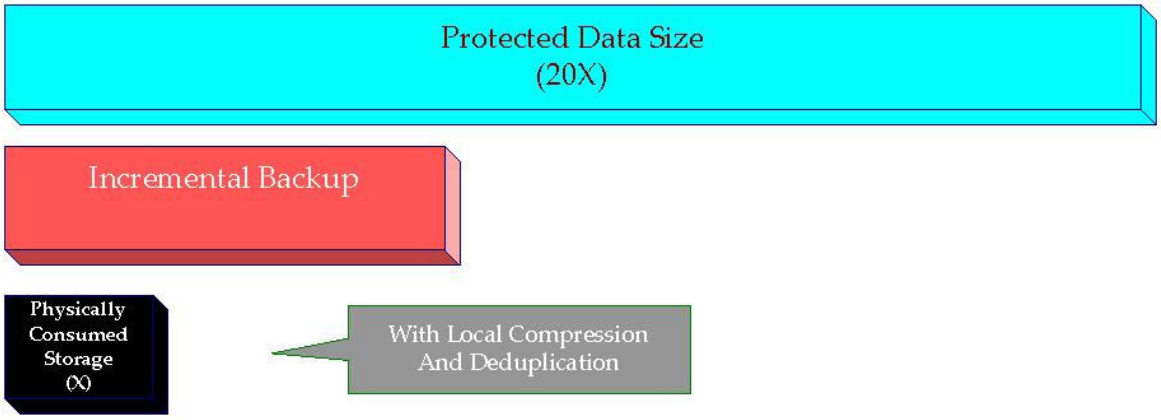

Figure 2: Backup Data Deduplication and Compression

Both deduplication and standard data compression (also referred to as 'Local Compression' in product literature) are executed via lossless 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

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 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 to the NBU 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.

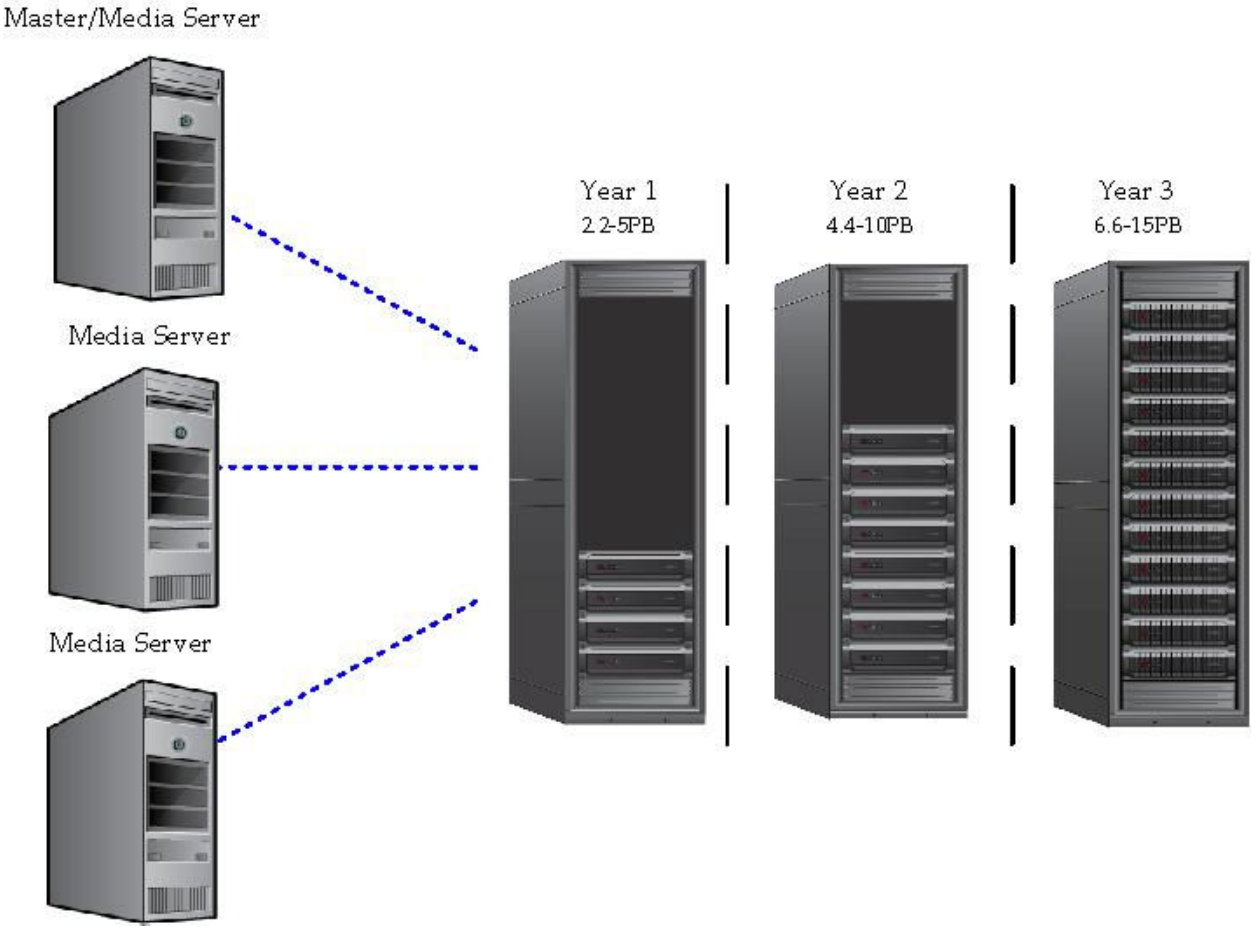

Figure 3: Data Domain System Architecture Scalability

File System and VTL Integration

Data Domain systems support two integration methods with NBU, 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 NBU media servers. This flexibility affords a great number of integration scenarios for NBU. The following figure illustrates both integration scenarios with NBU media servers.

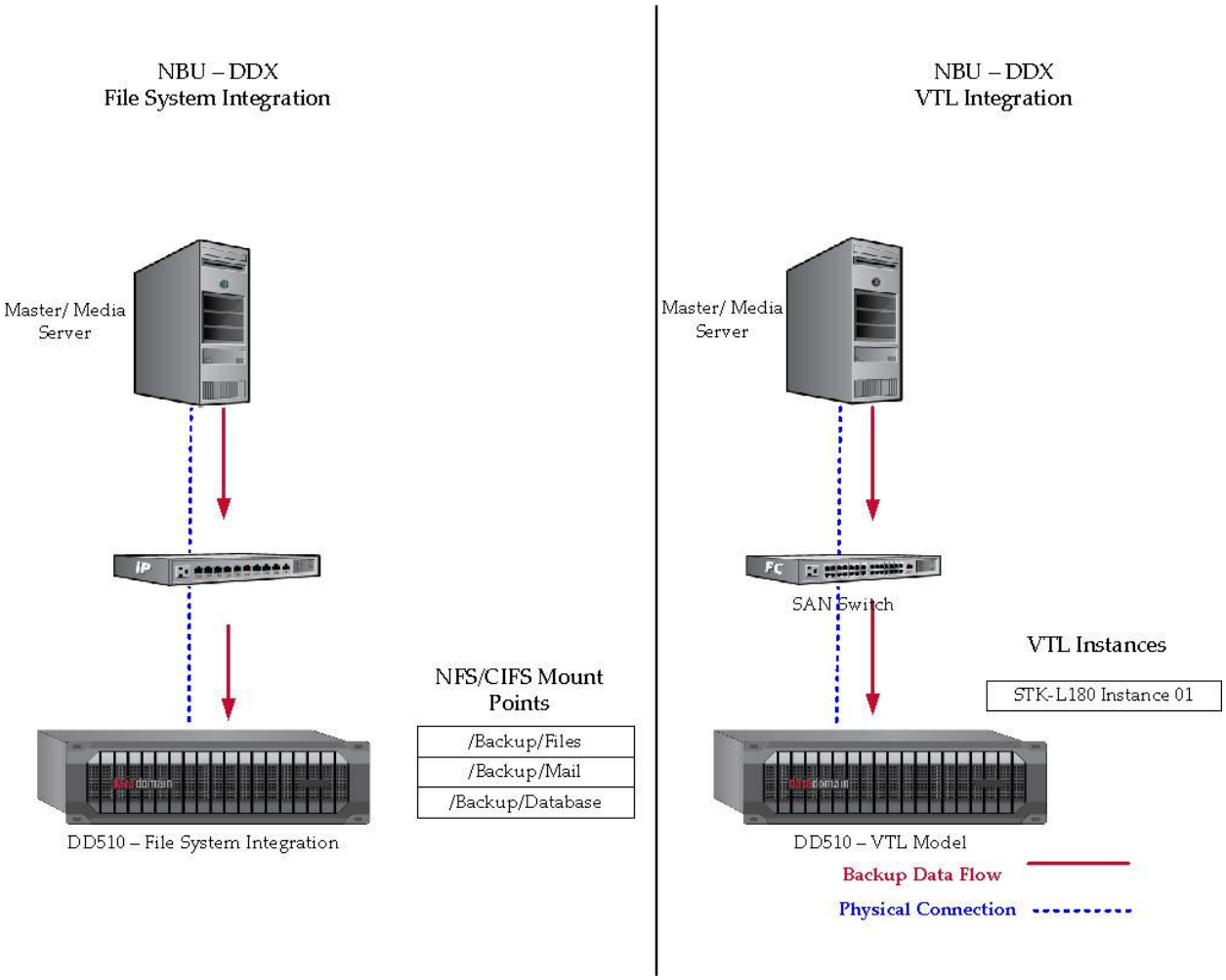

Figure 4: NBU Data Domain System Integration

For network file system access, NBU addresses the Data Domain system via a native NFS/CIFS mount. NBU 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 NBU device driver. NDMP backups are supported only via the Data Domain VTL interface. 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 NBU 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 implementations 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.

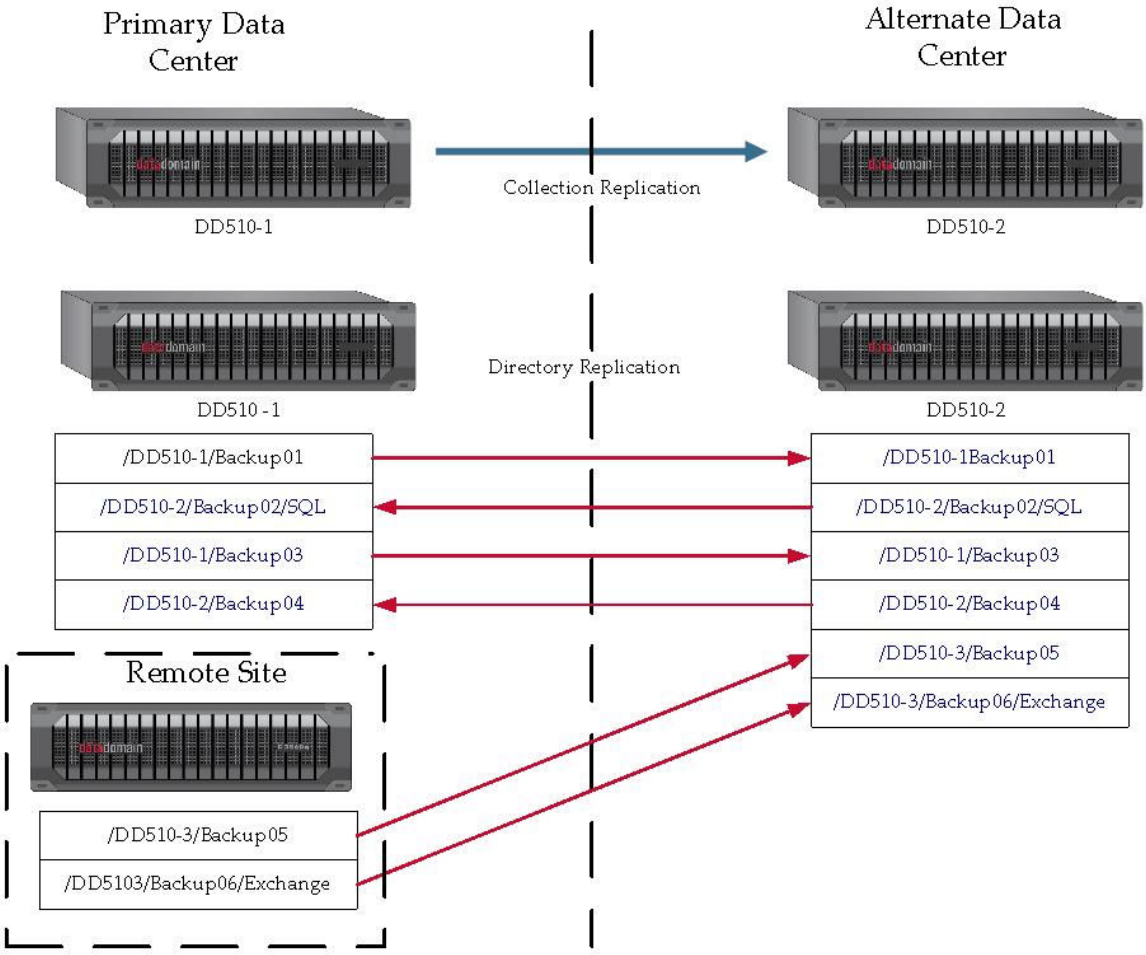

Figure 5: Collection and Directory Replication Modes

How Data Domain Best Fits with NBU

A Data Domain system provides an alternative for disk and tape volume pools in NBU. The Data Domain file system is optimized for sequential read and write operations. This provides for a great fit with existing NBU disk-based or VTL abilities when coupled with the NBU Vault option. With the OpenStorage (OST) plug-in NBU can control the replication performed between Data Domain system instances and can track the data movement in its catalogs. This capability enables a very powerful option for offsite replication, or DR solutions.

The NBU catalog should continue to be provisioned on traditional disk devices where these NBU elements can be backed up to a Data Domain system for operational recovery and replicated to a remote site for disaster recovery.

Note: Some NBU environments are architected to support extremely high-performance backups for high-volume clients. Typically, specialized designs are implemented to support backup speeds of 1-4 TB/hour. The Data Domain system architecture can be configured to support high-performance workloads (via multiple parallel instances) 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

NBU policies are more or less unique to each customer environment, but generally follow a common methodology. Most environments use a mixture of incremental backups with the occasional full backups run on a regular schedule (weekly or monthly). Differential incremental backups are more common because they are faster and have a smaller footprint on available backup storage. This leaves the full backups consuming a larger chunk of disk and tape storage pools. Every full backup will write redundant data already existing 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 NBU Synthetic Full backups, the goal of which is to create a full backup image from existing backup data. This process allows the NBU 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 NBU 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 the majority of 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 NBU. 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 the characteristics of data 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 NBU perspective since NBU views data change at the file/object level.

The ratio of current backup data to Data Domain system data stored (after data deduplication and compression) varies, but on average NBU 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 tally up the primary volume of data on all backup clients, or a 1:1 ratio of primary data to the Data Domain system's 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 NBU media 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.

Integration planning

A Data Domain system integrates into a NBU 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 NBU environment as a VTL or as disk storage unit (DSU) with a NAS configuration.

The following figure illustrates NBU integration with a Data Domain system in a primary and alternate site configuration. In this architecture a single NBU master / media server is shown. In a production environment, a single NBU media server can be successfully mapped to multiple Data Domain system instances, or vice versa with the appropriate NBU licensing. Note that the customer becomes 'DR Ready' much faster with this inline deduplication and replication approach compared to systems that do deduplication as a post-process or where tapes are physically transported off-site..

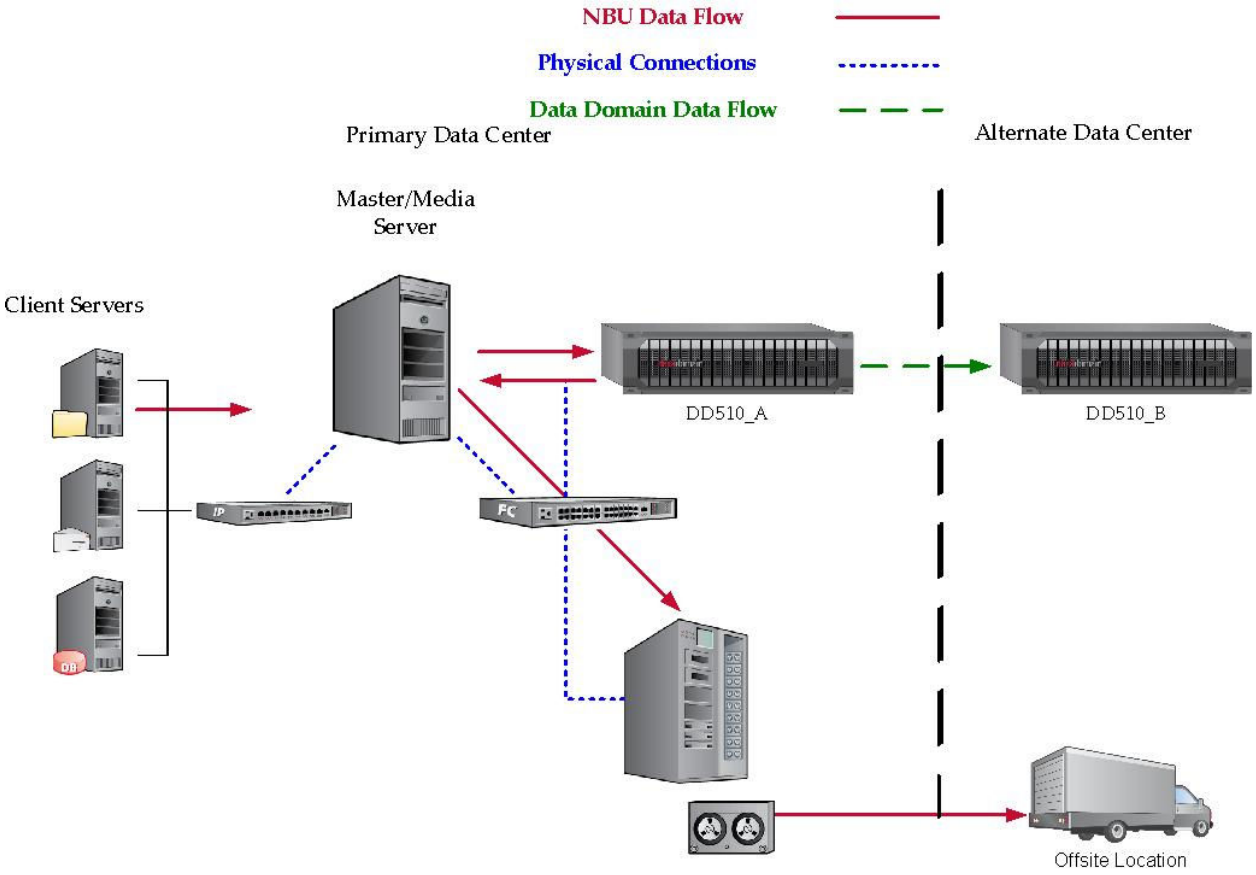

Figure 6: NBU Site Integration

In a disk-based integration scheme, more administrative options are available for Data Domain system integration. One or more directories can be specified as the location of files for a storage unit (disk). As a result, multiple Data Domain system instances can be dedicated to specific application types; however, the benefits of data deduplication do not span multiple Data Domain system instances at this time. 'Like data' can be mapped to individual Data Domain system instances or to individual mount points within a Data Domain system in order to take the greatest advantage of deduplication technology. The following figure illustrates this concept.

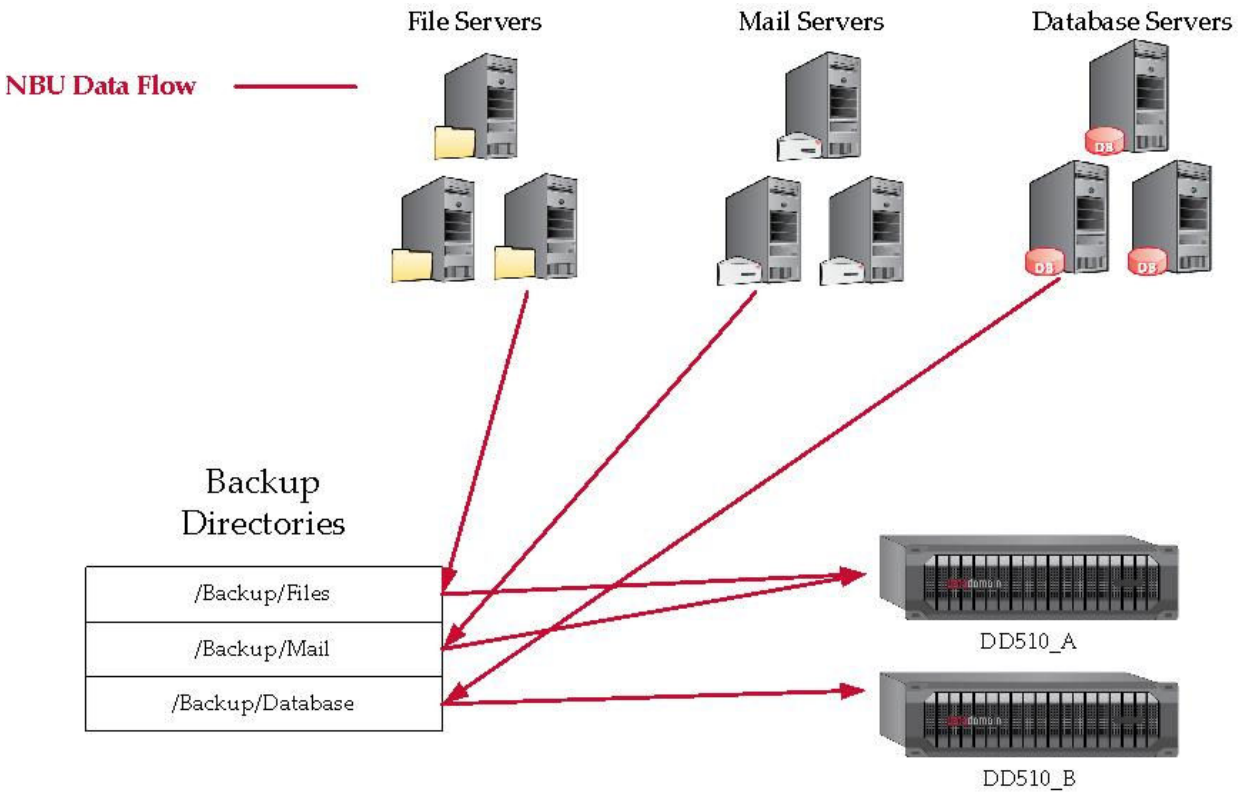

Figure 7: Mapping Data Types to Backup Directories and Data Domain System Instances

There are significant administrative benefits to mapping client data types 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 required. 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 of the remaining data types.

Large NBU 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. At present, each Data Domain system presents a separate Storage Unit for NBU.

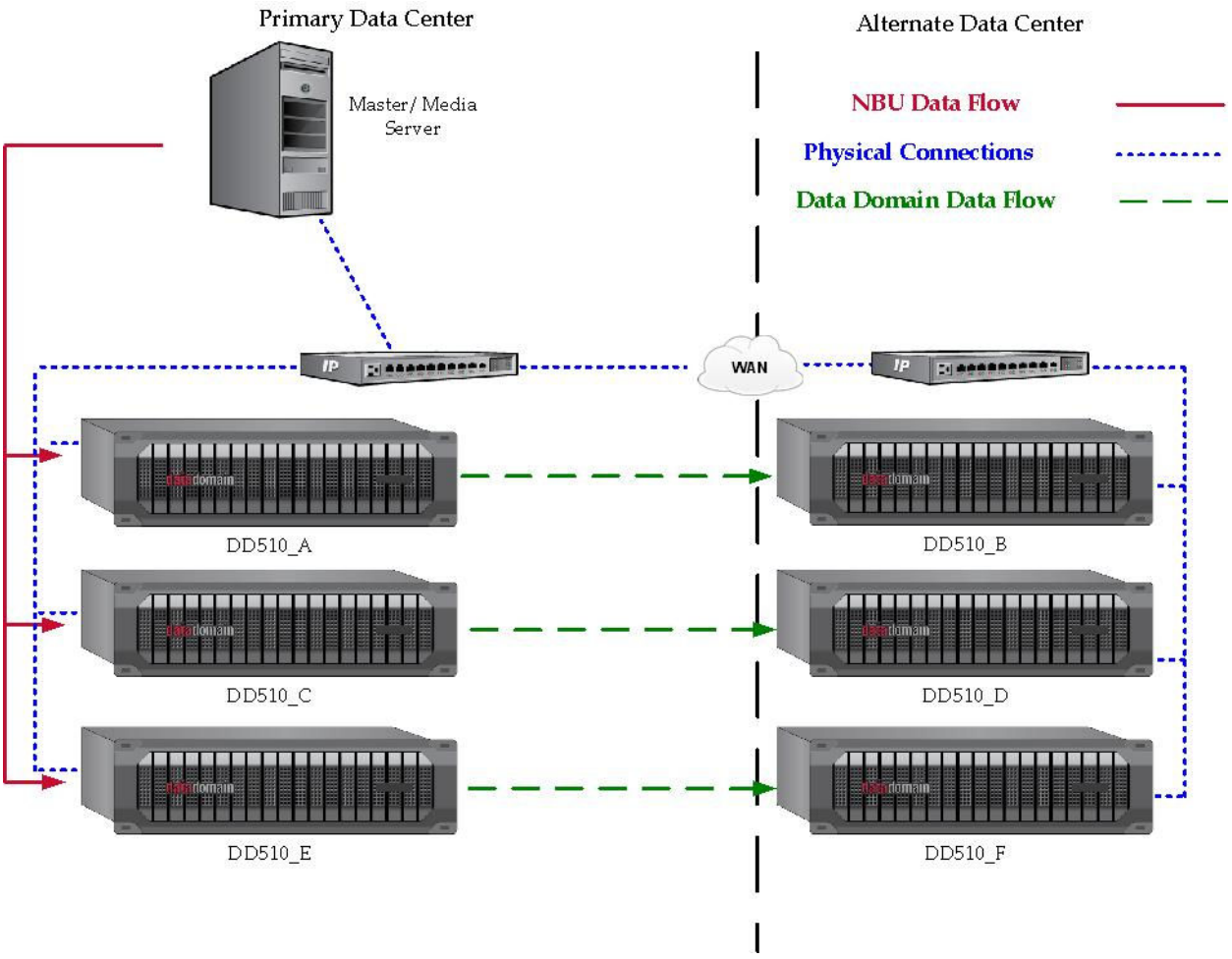

Figure 8: Large Site NBU Integration

Offsite Duplication

A significant aspect of all data protection solutions is the ability to send important data to a distant location where the intent is that it will be kept safe from any disaster that may affect the primary data site. That data can then be used to restore operations within the tolerance of a specified Recovery Point Objective (RPO) and Recovery Time Objective (RTO). Offsite storage solutions can also play a large part in data archival, and NBU allows for several duplication methods for both measures.

If a tape-only environment is configured to push enough data to keep multiple drives spinning, inline copy is an option for duplication. This option can be configured to make up to four (4) online copies of the backup as it is being taken. The speed of these copies, however, is limited by the slowest resource each copy is being directed toward.

A Data Domain system makes offline duplication much faster and easier when used in conjunction with the NBU Vault option. NBU Vault can automate the duplication process for backup images selected by a specific client, a date range, etc. Using an offline duplication method with a Data Domain system has several advantages.

- Frees resources and allows for maximum efficiency of tape backup infrastructure resources during backup hours

- A Data Domain system, being a disk based deduplication appliance, can quickly push data over a relatively unburdened dedicated backup storage network during backup off-hours
- Disk to disk to tape over a dedicated network greatly helps reduce wear and tear on tape drives by eliminating the need to make a tape at the site of the local backup.
- The Vault option allows for a completely automated system and frees administrator time to concentrate toward other tasks

The OST plug-in is also another method of controlling the deletion and replication of backup images from a single point. OST allows NBU, versions 6.5 or higher, to better communicate with disk-based storage devices. This option is available to NBU customers that utilize a Data Domain system, allowing NBU to integrate with and take advantage of the benefits provided by the Data Domain systems.

The configuration for this solution requires OpenStorage (OST) Server running on the Data Domain system, and an OST plug-in installed on the NBU media server. The NBU administrator will have the ability to control the backup, restore, deletion, and duplication of backup images written to this OST disk storage type. The backup images are organized into Logical Storage Units (LSU) on the Data Domain system and multiple backup images can be stored in an LSU. The replication abilities between Data Domain system instances will greatly enhance the efficiency of both offsite, and DR image duplication when used in conjunction with the NBU OST solution.

The main advantage of instituting offline duplications with a Data Domain system, regardless of method, is speed. This in turn pushes a shorter active backup and duplication window, and makes more time available for system administrative tasks.

Operational Considerations

When outside infrastructure is configured optimally, administering NBU environments is relatively painless. When any aspect of the architecture; networking, firmware version, server capacity, etc, is misaligned it will show in reduced daily success rates. A Data Domain system has the potential to seamlessly install into any architecture and will enable NBU administrators to more efficiently take advantage of the resources already in place.

Utilizing the advantages of the disk based backup destination that a Data Domain system presents allows a NBU administrator to efficiently set a timeline for daily operations, and meet that timeline. These daily operations consist of the typical routines, outlined by the following diagram.

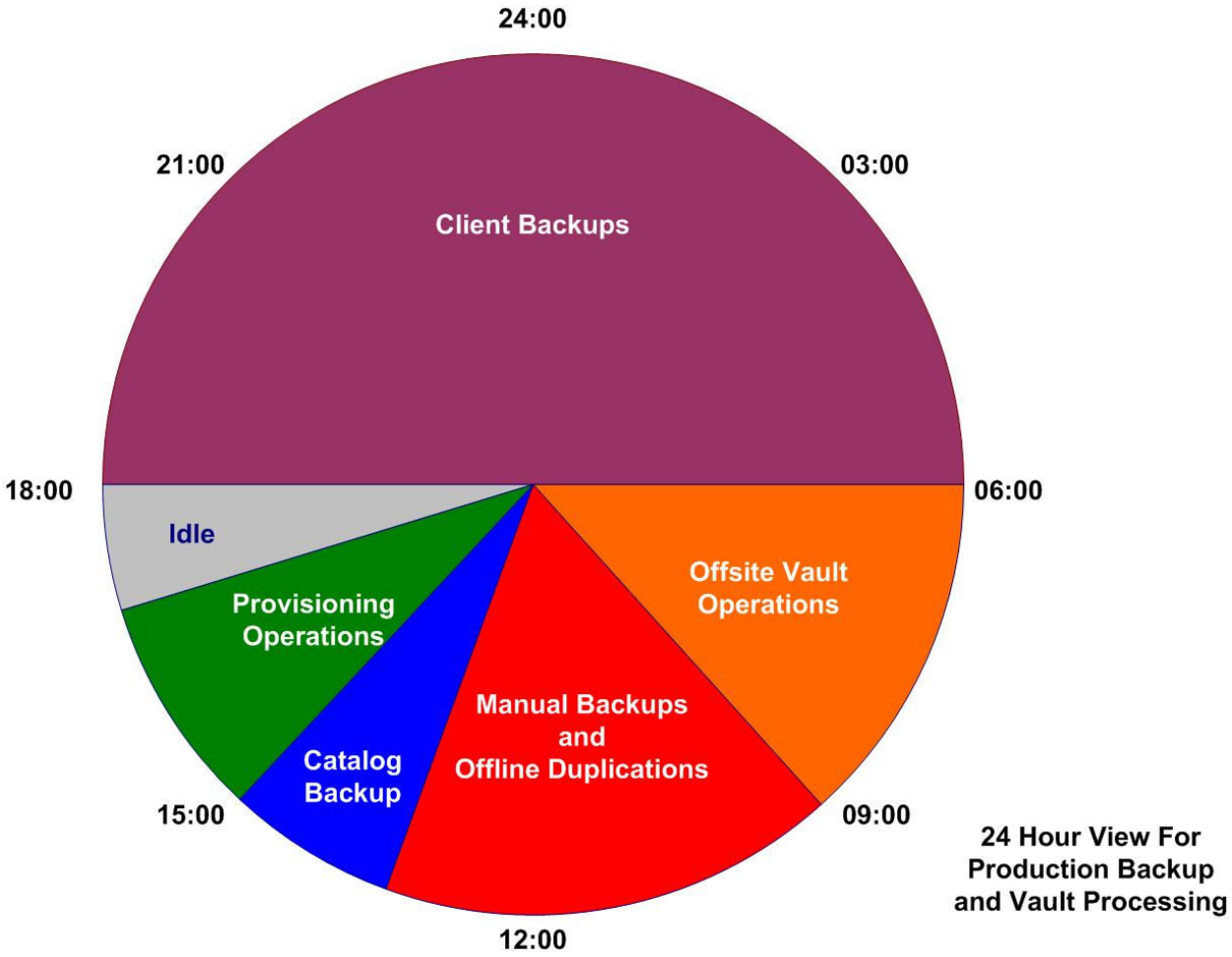

Figure 9: NBU Daily Operations

A Data Domain system instance can greatly reduce both the backup window and duplication process. With the NBU Vault agent, or another duplication manager, offline duplications to tape become a better option than inline copy. However if the Data Domain system architecture is leveraged for replication to offsite storage, any manual tape vaulting processes and the need for agent based offline duplication can be completely eliminated. Either implementation strategy significantly reduces the amount of time required to complete daily operations as illustrated in the following graphics.

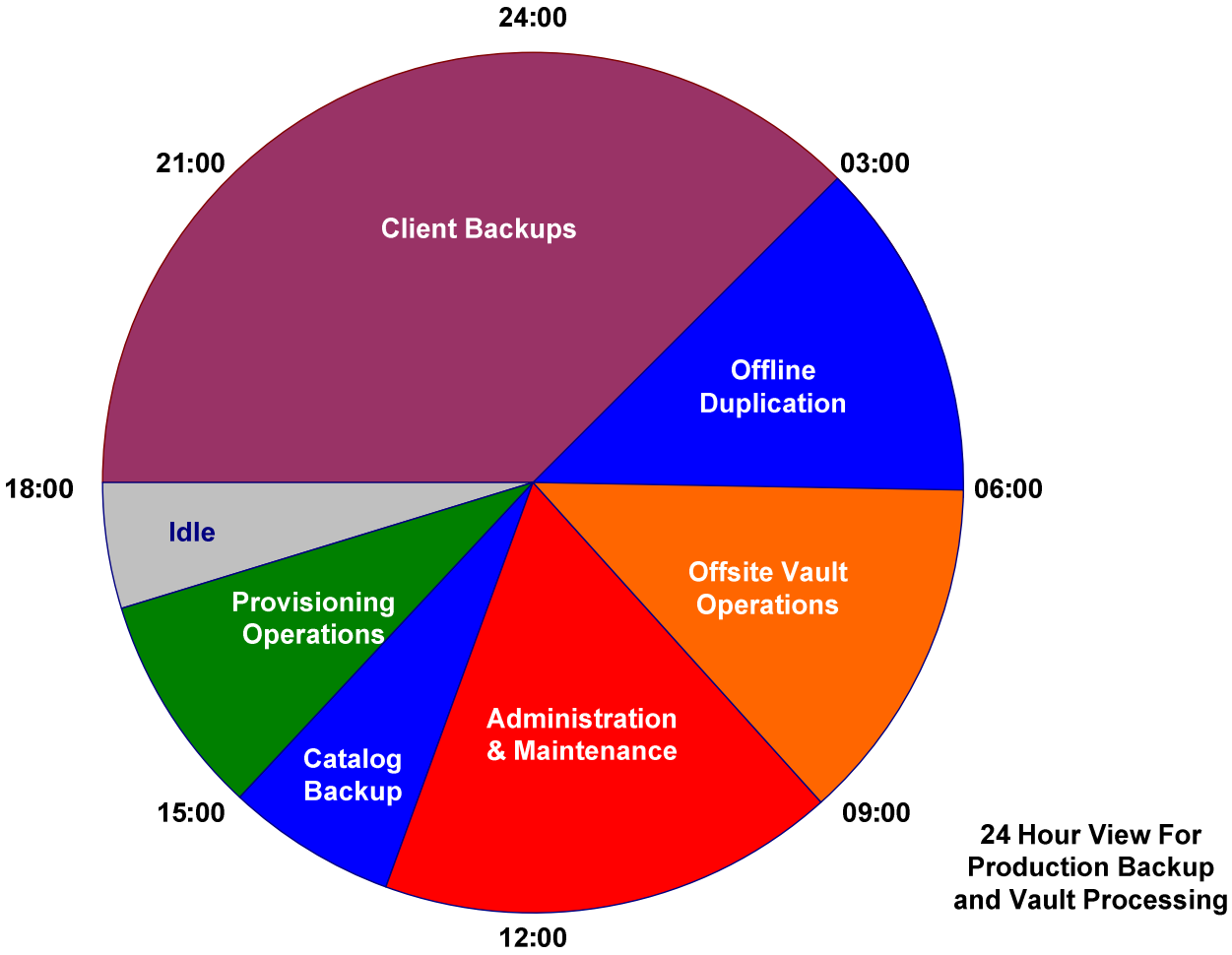

Figure 10: NBU Daily Operations with Data Domain and Tape Duplication

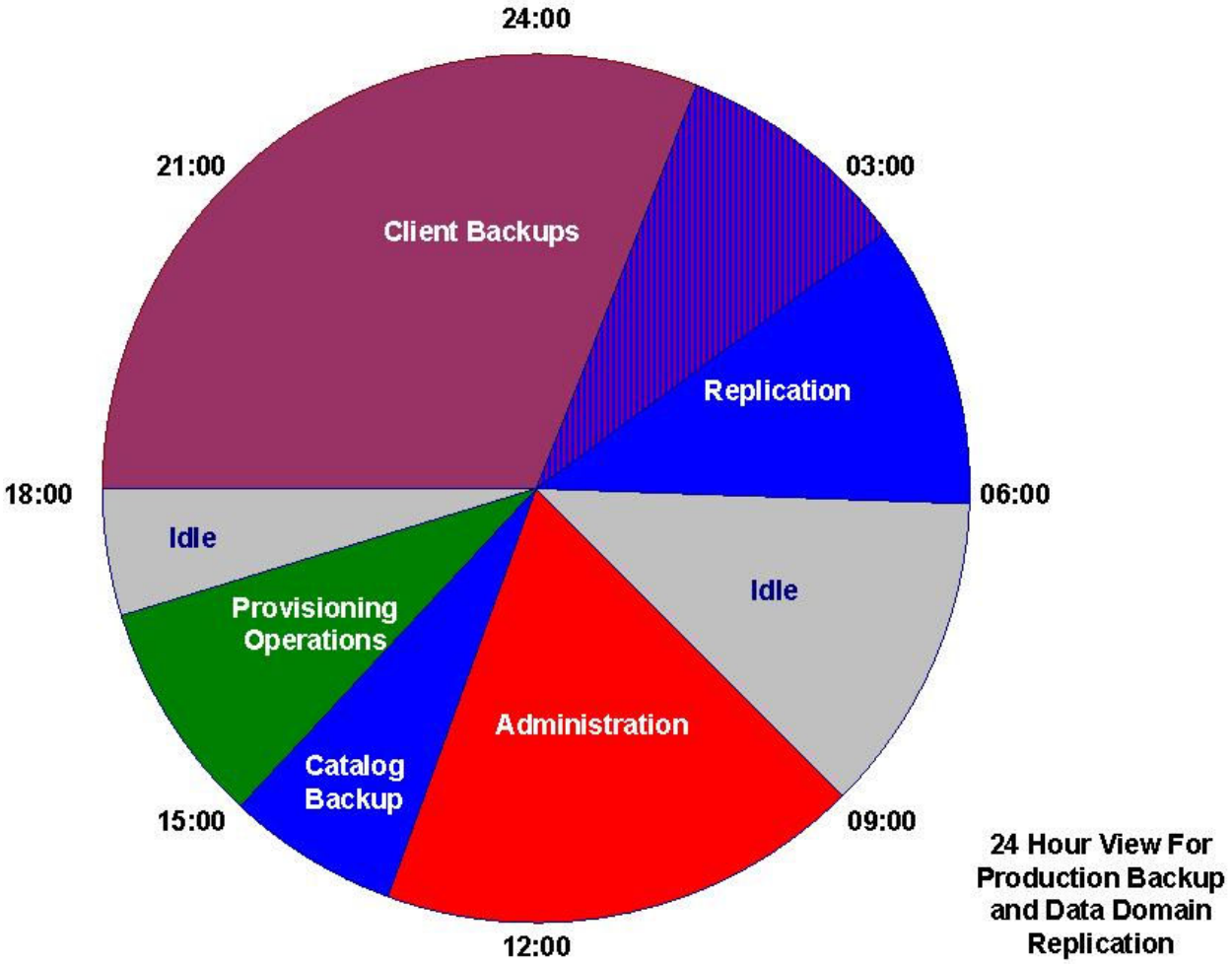

Figure 11: NBU Daily Operations with Data Domain Replication

Just like physical disk and tape resource management for NBU, Data Domain system instances must be managed to ensure efficient utilization and sufficient available space. As illustrated in the following figure, the NBU administrator should monitor available Data Domain system space as a routine operational task.

While 'expiration' processing in NBU immediately deletes data from file type device class storage pools, the data is not immediately deleted from a Data Domain system. The Data Domain system requires the administrator to perform or schedule '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 capacity, 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 compared to normal backup and restore I/O. We recommend the Data Domain system cleaning processes run periodically as part of standard NBU 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 storage infrastructure, restore activities are hampered by all of the limitations that traditionally come with trying to read data stored on a mechanical device. With the Data Domain system and inline native deduplication, NBU sites can leverage Data Domain systems for production backups without having to change their current methodologies. If offsite duplications are 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.

.

NBU 5.X + offers what are called Synthetic backups with disk based storage units. Synthetic Full backups are created from combining a traditional full backup and subsequent incremental backups. NBU is also able to create Synthetic Cumulative backup images from a set of cumulative or differential images. Using these synthetic backups allows an administrator to recover from one or two images, greatly simplifying restore operations.

Direct onsite recovery conditions improve first and foremost due to a Data Domain system based on inline deduplication and resilient disk storage. 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.

Integration Basics

NBU Server Tuning

NBU master server tuning is recommended for new Data Domain system implementations using NFS/CIFS and IP protocol

Note: Some server configuration changes require a server reboot to take effect.

The following table outlines standard NBU server tuning parameters for optimal Data Domain system performance with NBU, 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 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 echo 262144 > /proc/sys/net/core/rmem_default echo 262144 > /proc/sys/net/core/wmem_default
WINDOWS – Network	**Note:** Do not modify the Windows registry parameter 'AFD' if the NBU

Configuration Type	Comments
Configuration	server or any associated NBU 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 4: NBU 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 NBU media server integration. Jumbo frames may be used to maximize TCP/IP protocol efficiency and minimize processor overhead on the NBU media server. Network connection trunking is not currently supported by the Data Domain system.

Data Domain recommends the following general NFS configuration settings for mounting a Data Domain system to a NBU master / media server.

NBU Server Platform	NFS Configuration
AIX	mount –v nfs –o proto=tcp,vers=3,intr,hard,combehind,rsize=32768,wsize=32768,llock –n dd200 *restorer-name*:/backup */mount-point*

NBU Server Platform	NFS Configuration
Solaris	mount -F nfs -o nolock,hard,intr,vers=3,proto=tcp,rsize=32768, wsize=32768 *restorer-name*:/backup /*mount-point*
HP-UX	mount –F nfs –o nolock,rsize=32768,wsize=32768,hard *restorer-name*:/backup /*mount-point*
LINUX	mount -T nfs -o nolock,hard,intr,nfsvers=3,tcp,rsize=32768, wsize=32768,bg *restorer-name*:/backup /dd/<mount point>
Windows	CIFS access to Data Domain system is recommended for Windows NBU Master / Media Servers

Table 5: NFS Mount Configuration Guidelines

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

Disk Storage Unit Design Considerations

NBU allows disk type storage units to be defined simply as a Disk Storage Unit, or as a Disk Staging Storage Unit (DSSU). Disk storage units are just what they advertise and allow for backups to be mapped to NFS mount points for storage, or can be located directly on the SAN. NBU 5.X and above allow for the creation of a DSSU which serves as a temporary warehouse for backup images before being moved off to another backup storage destination.

When configuring a Data Domain system as a disk storage unit, time needs to be spent on developing a simple and obvious file space mapping. Mapping client data types to specific destinations allows for easy administration when replicating backup images of common application types. In NBU 6.0, or higher, the default High Water Mark is set at 98%, but for a longer lead time and to allow for more efficient space planning a lower mark should be set by the NBU administrator. A High Water Mark of 95% is recommended by Symantec and can be changed when creating a disk storage unit or afterwards. The Vault option is recommended for automating duplication operations without inline copy.

The typical disk storage unit can be taken a step further and configured as a DSSU. This method treats disk as cache when taking an initial backup. Using this method does not require the Vault agent to use for duplication to another backup storage source as a disk staging schedule would handle those operations. When this type of storage unit fills up it automatically finds the oldest image that has been copied to another storage destination and deletes it.

NDMP Integration

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

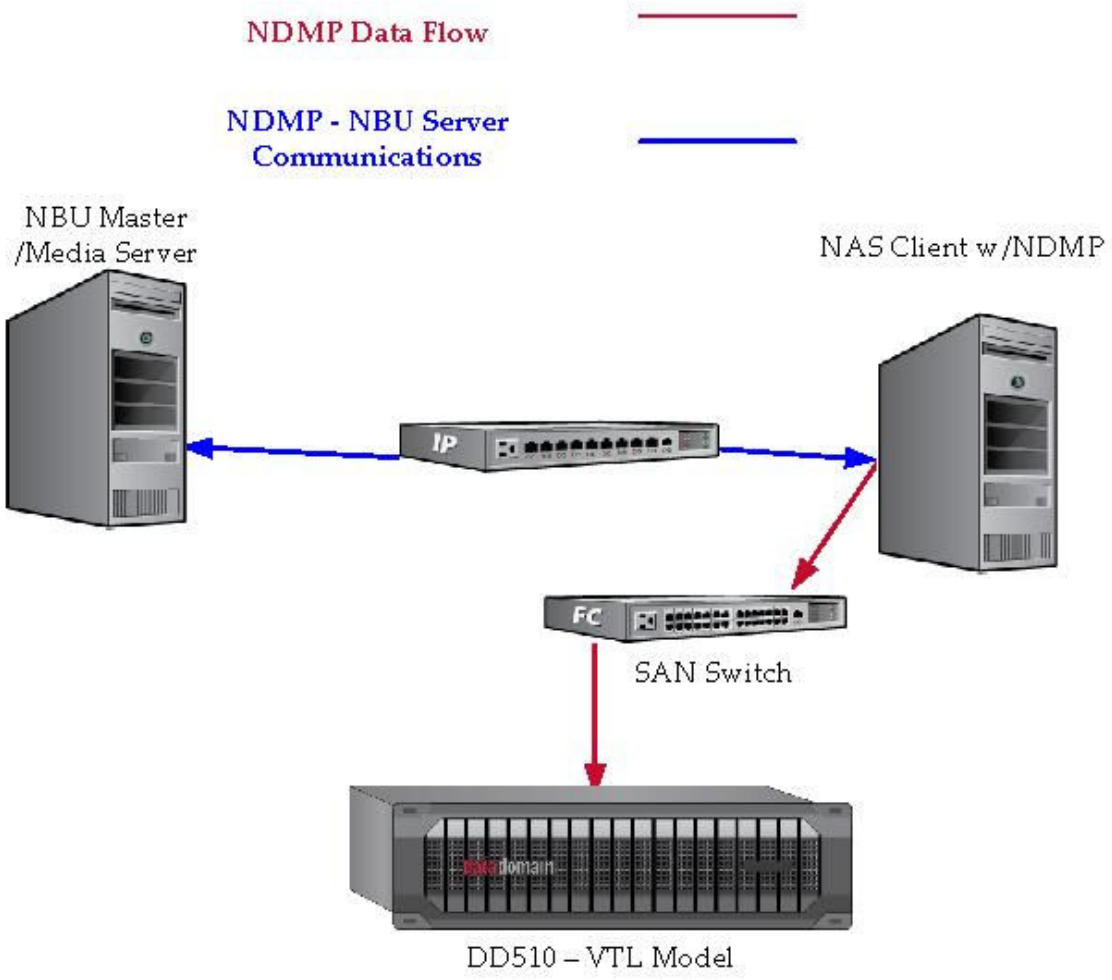

Figure 12: NDMP Integration with NBU and Data Domain System

Archiving Integration

Long term data retention in NBU is accomplished either via traditional backup or archive functionality with Enterprise Vault. A NBU Vault based long term backup solution automatically manages the duplication of data across tiers of archive storage, each with a set retention. The Enterprise Vault solution provides for a fully functional archiving solution.

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

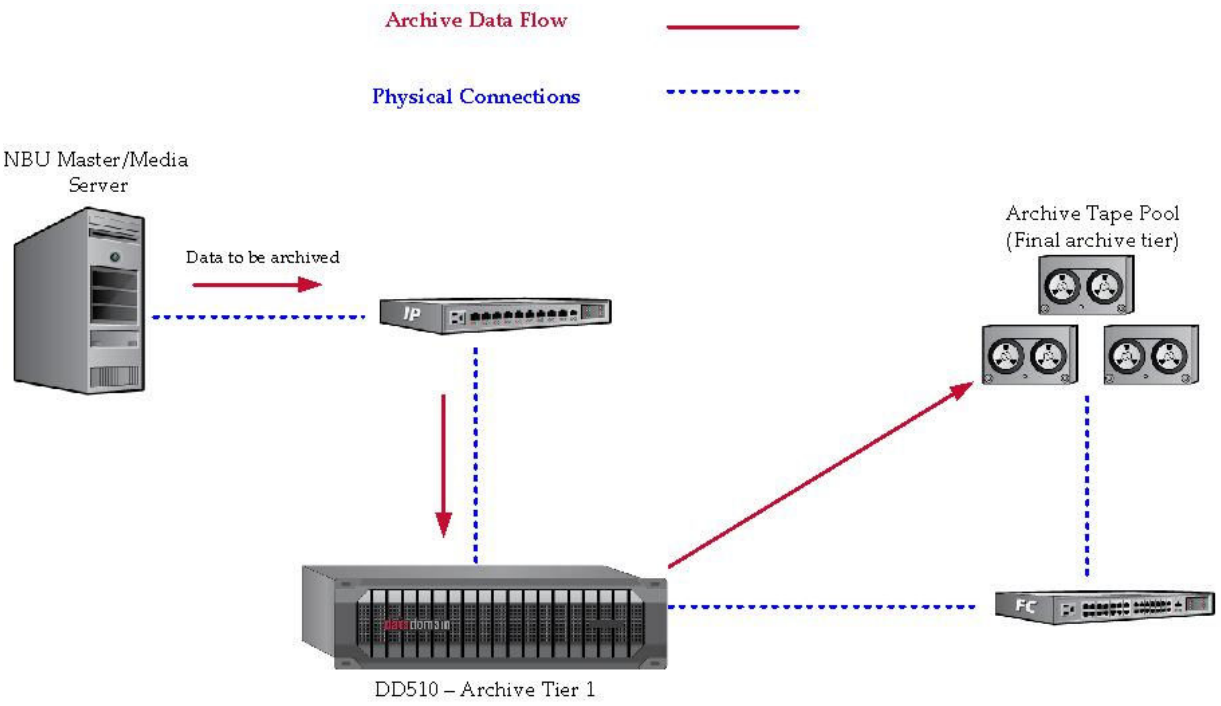

Figure 13: NBU Long Term Retention to Tape with a Data Domain System for Tier 1

NBU data with extended retentions can also be moved directly into the archive environment from backup images stored within the same Data Domain system. This data can also coexist with existing active backup data from a NetBackup Master/Media Server.

NBU Catalog Backups

NBU catalog backups are critical to the recovery of a NBU environment. A NBU catalog backup can be full or incremental just as any other backup, and captures all new image record changes currently committed since the last backup. Catalog backups are usually written to tape and run on a schedule during which, typically, no other backups are running.

A drawback of NBU backup to tape is that once a tape is designated for catalog backups it is no longer available for regular backups unless a manual process is run to change this designation. This leads to inefficient use of tape resources especially with today's larger tape capacities. With a typical LTO3 cartridge holding 1 TB, a 100 GB NBU catalog will leave 90% of the tape unutilized.

With the introduction of a Data Domain system to the NBU environment, NBU catalog backups can be written directly to the Data Domain system. When NBU catalog backups are sent to the Data Domain system, storage space is not wasted; this is true whether the Data Domain system is used in VTL mode or file mode. A Data Domain system as the primary backup destination storage unit can cut the number of tapes used for NBU catalog backups in half. If a second Data Domain system is introduced at an alternate site, the NBU catalog can be replicated using the Data Domain system. This setup will eliminate all of the inefficiently used catalog backup tapes. It will also reduce the amount of NBU server time required to duplicate backups daily by leveraging Data Domain replication for the offsite database backup copy.

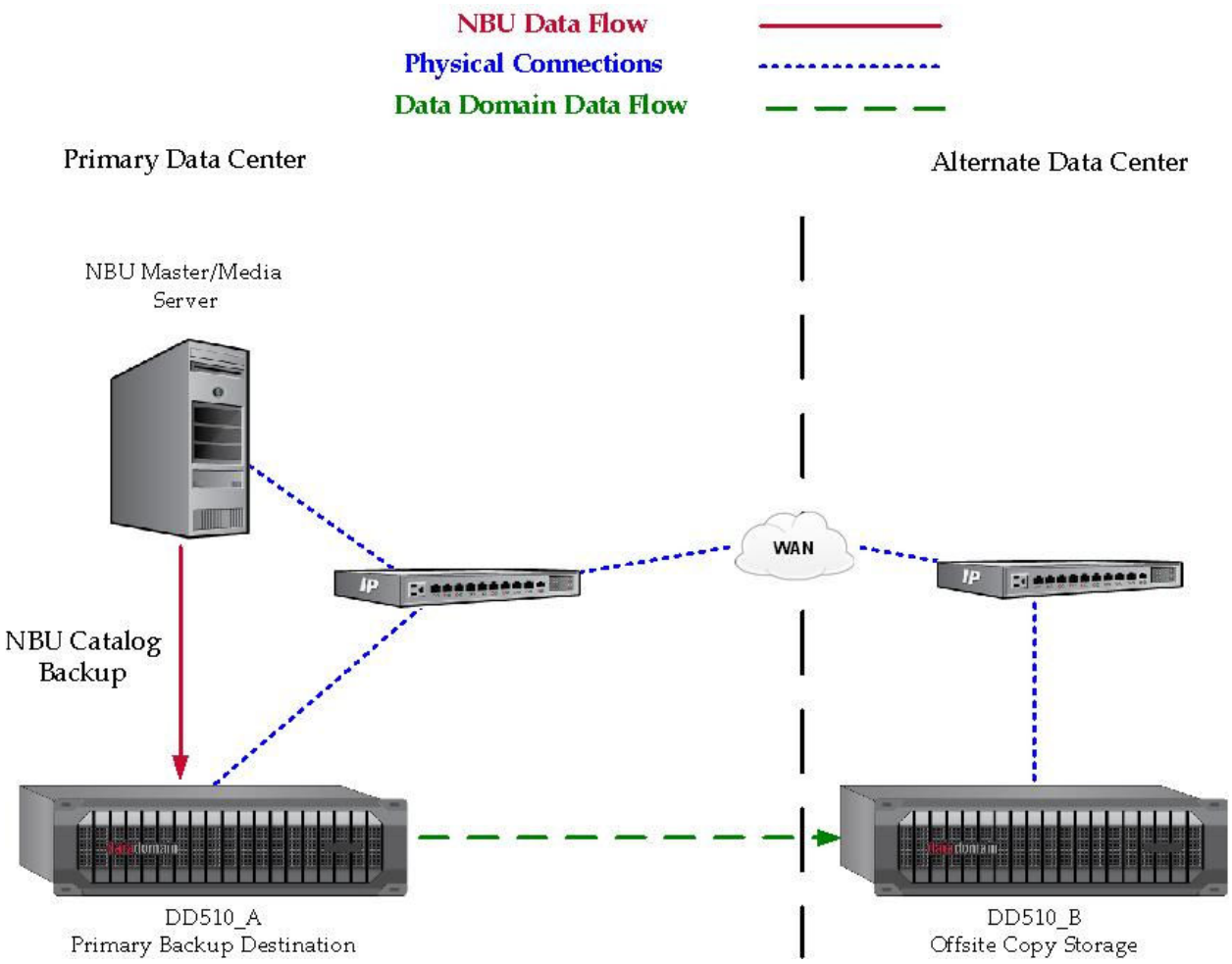

Figure 14: NBU Catalog Backup using Data Domain Systems

NBU Disaster Recovery

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

The NBU Vault option is available to automate NBU disaster recovery operations with duplication and tape management profiles. The general sequence of events to recover a NBU server instance includes the following steps.

Note: To start the recover process you will need a new server running the same OS, patch level, file system configuration, and the same short and fully qualified host name as the production NBU master server. It will also need to be connected to the Data Domain system at the DR site.

1. Install the NBU master server Code with complete licensing information
2. Ensure that the disk where you are restoring the catalog contains the directory where the catalog previously resided. Recreate any symbolic links to catalog locations that may have existed
3. Stop required services on the new master server and any media servers that are attached to the master environment

4. Execute the proper "bprecover –r –dpath *device_path*" to select the items that are needed to recover
5. Restart the NBU services on all servers

Before running any disaster recovery plan it is best to refer to the relevant NBU 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 Disk Storage Unit with Manual Tape Vaulting

In the scenario illustrated in the following figure, a Data Domain system instance is used as a traditional disk storage unit.

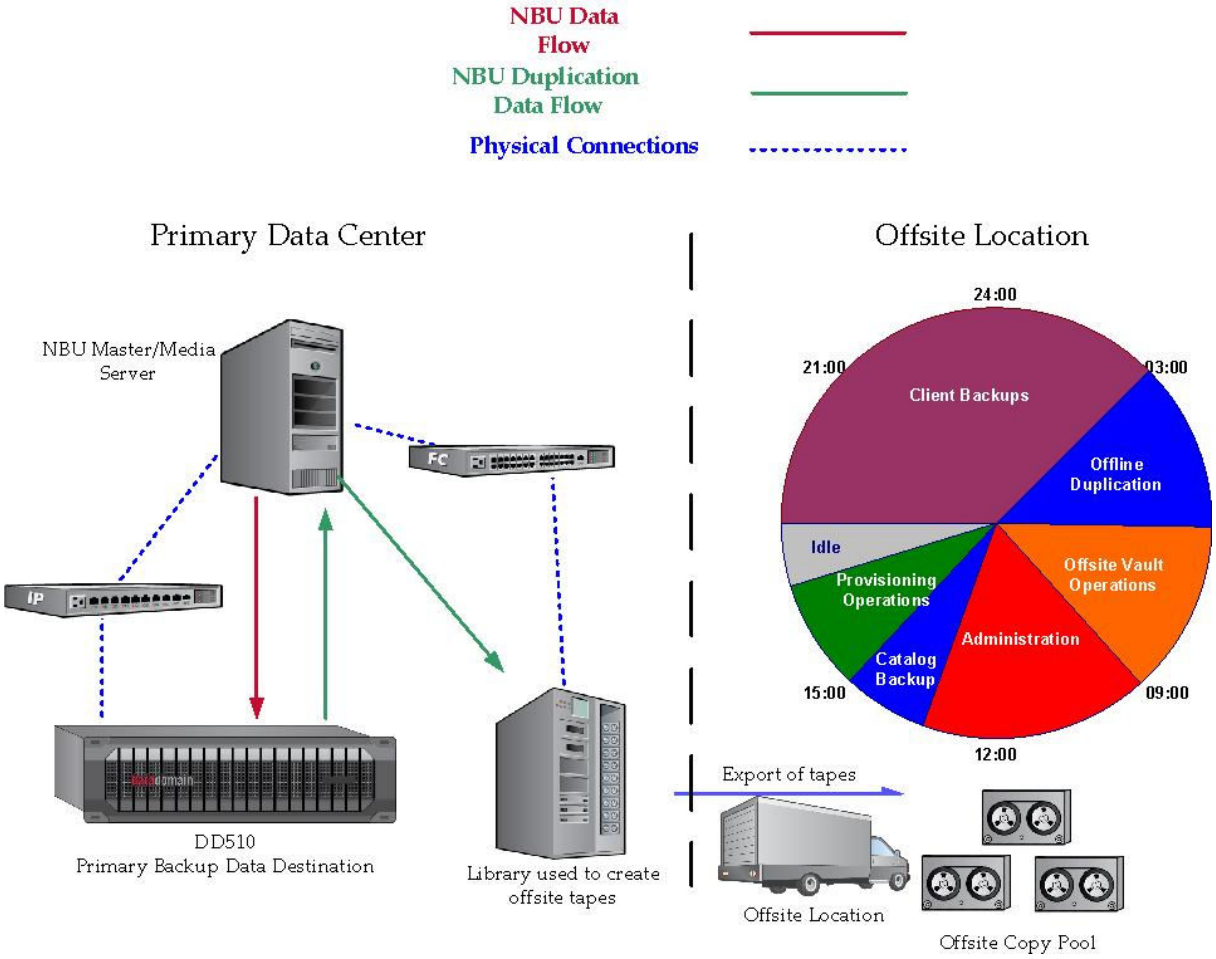

Figure 15: Local Data Domain Disk Storage Unit with Manual Tape Vaulting

A tape library is used to create physical tapes for offsite duplicates and the tapes are manually transported offsite via courier. Due to the use of a Data Domain system as the primary storage unit for backups, inline copy and any manual duplication efforts have been eliminated, shortening the daily operational workload.

Local Data Domain Disk Storage with Electronic Tape Vaulting

In the scenario illustrated in the following figure, a Data Domain system instance is used as a traditional disk-based storage unit. A tape library is accessible via extended SAN at the alternate site, and is used for daily electronic tape vaulting.

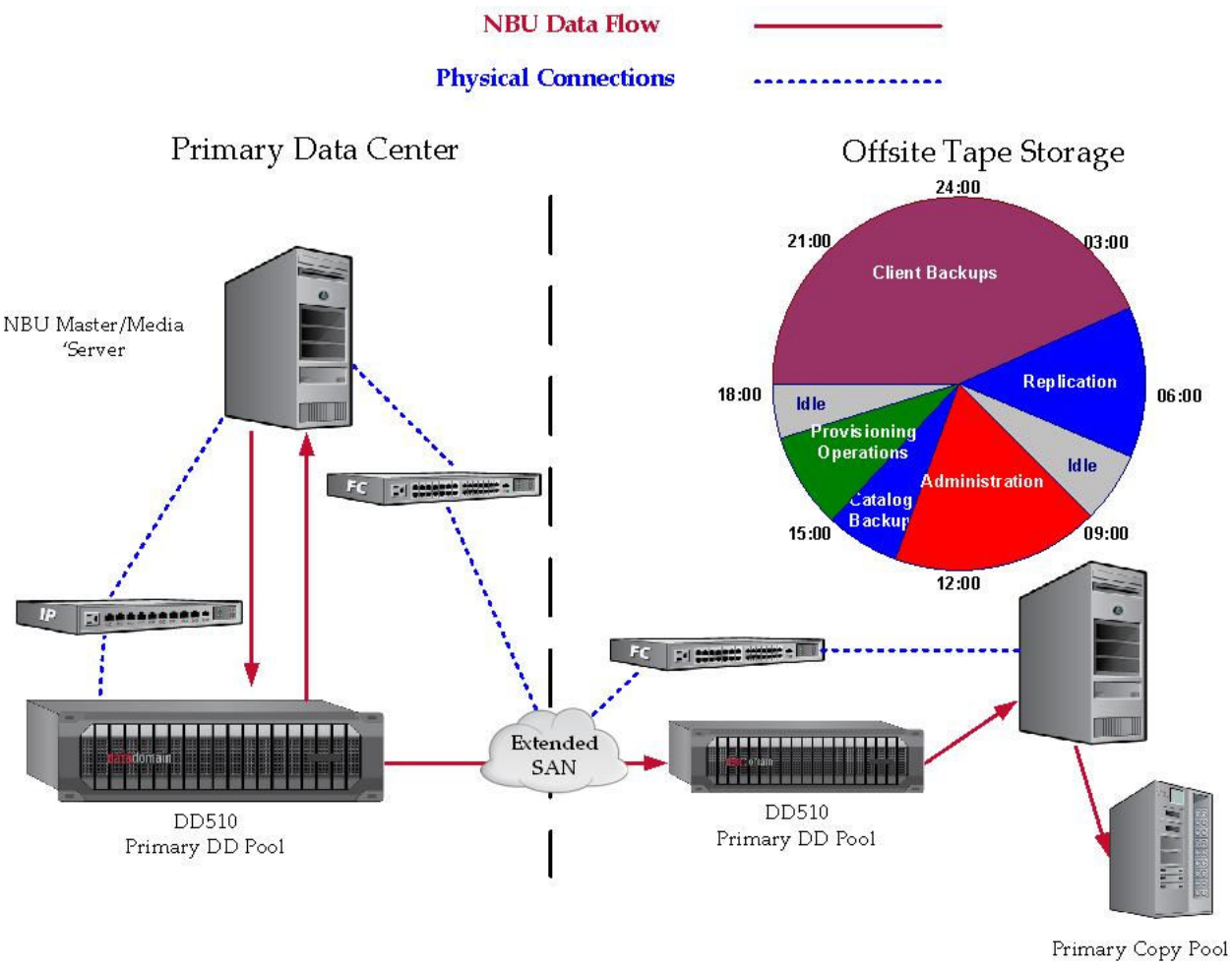

Figure 16: Local Data Domain System Disk Storage with Electronic Tape Vaulting

Again, inline tape copy and manual duplication efforts have been removed from the daily tasks, shortening the daily operational workload. This scenario benefits sites with existing electronic vaulting and offsite tape infrastructure in place.

> **Note:** All compression and deduplication benefits achieved on the Data Domain system are not realized on tape copy, since all tape copy operations are performed at the NBU application level.

Local Data Domain Primary Storage with Data Domain Restorer 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 backup and archive storage pools. Another Data Domain system is placed at the alternate datacenter and data is replicated between sites using the Data Domain system replication. The use of the Data Domain Restorer replication reduced the network bandwidth needed for electronic vaulting of data.

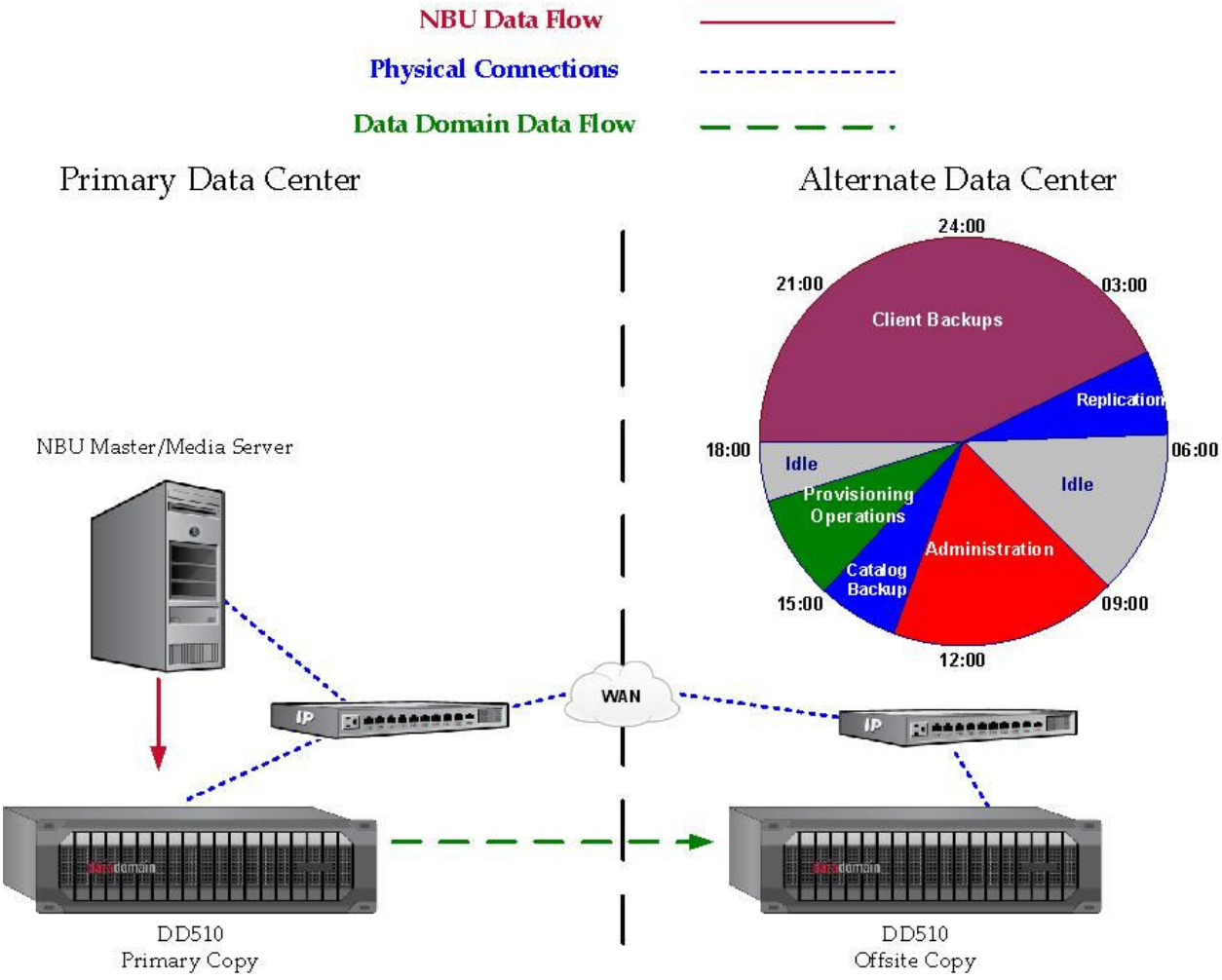

Figure 17: Local Data Domain System Primary Backup and Archive Storage with Replication to Offsite Data Domain System

By utilizing the Data Domain systems at both the primary and alternate data center, the daily operational tasks for backup storage pools are reduced. Creation of physical tapes for vaulting purposes can occur at either location in this scenario.

Mixed Production / DR with Data Domain Restorer Replication

In the scenario illustrated in the following figure, multiple NBU media server production instances electronically vault data to the Data Domain systems at the opposite site.

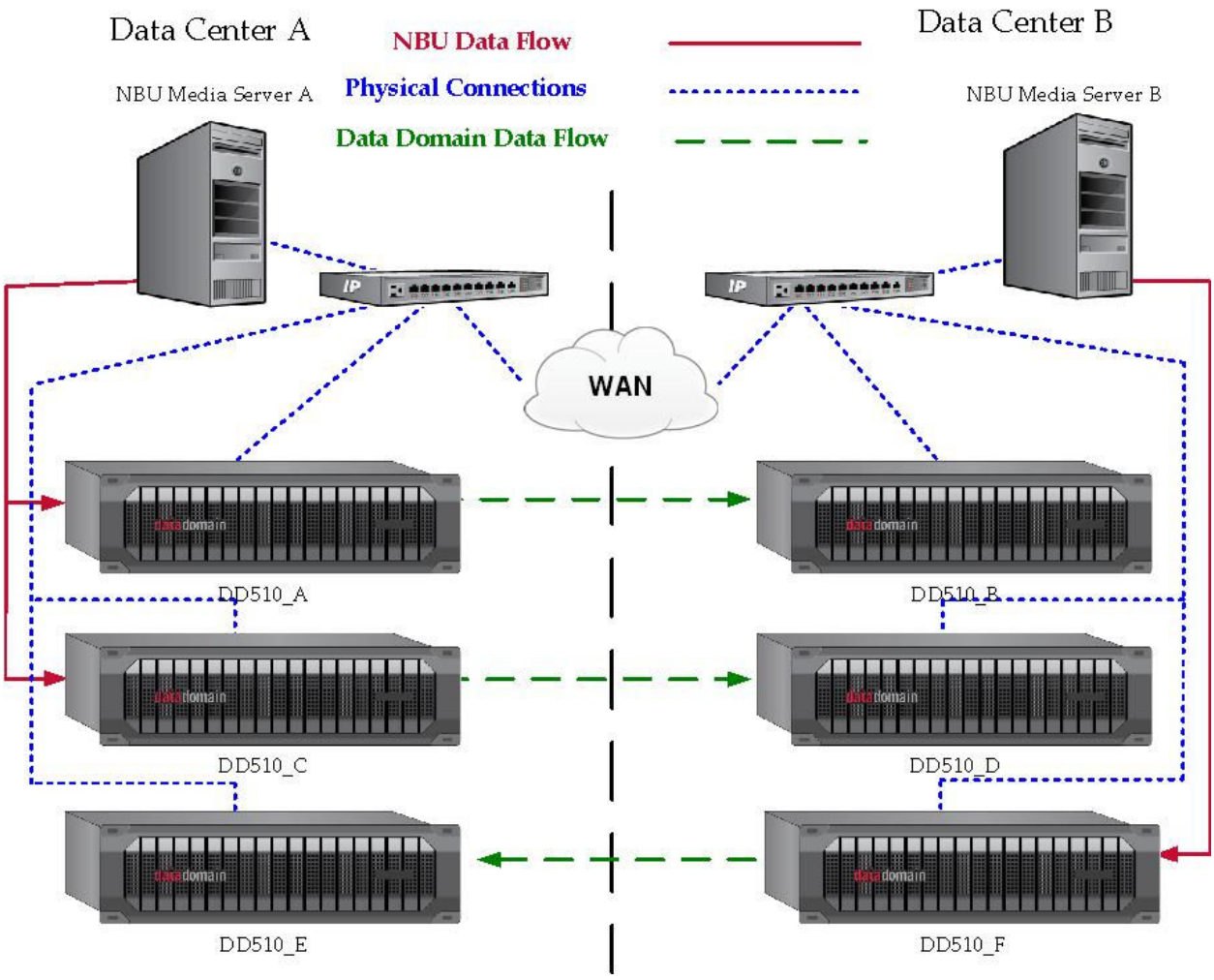

Figure 18: Mixed Production / DR with Data Domain Replication

Each data center supports a mixture of production and disaster recovery capacity, and the NBU master server vaults 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 acting as a repository for NBU backup data replicated to the DR site via Data Domain replication. In this example each data center site is controlled by its own media server instance with the NBU master server at the centralized DR location.

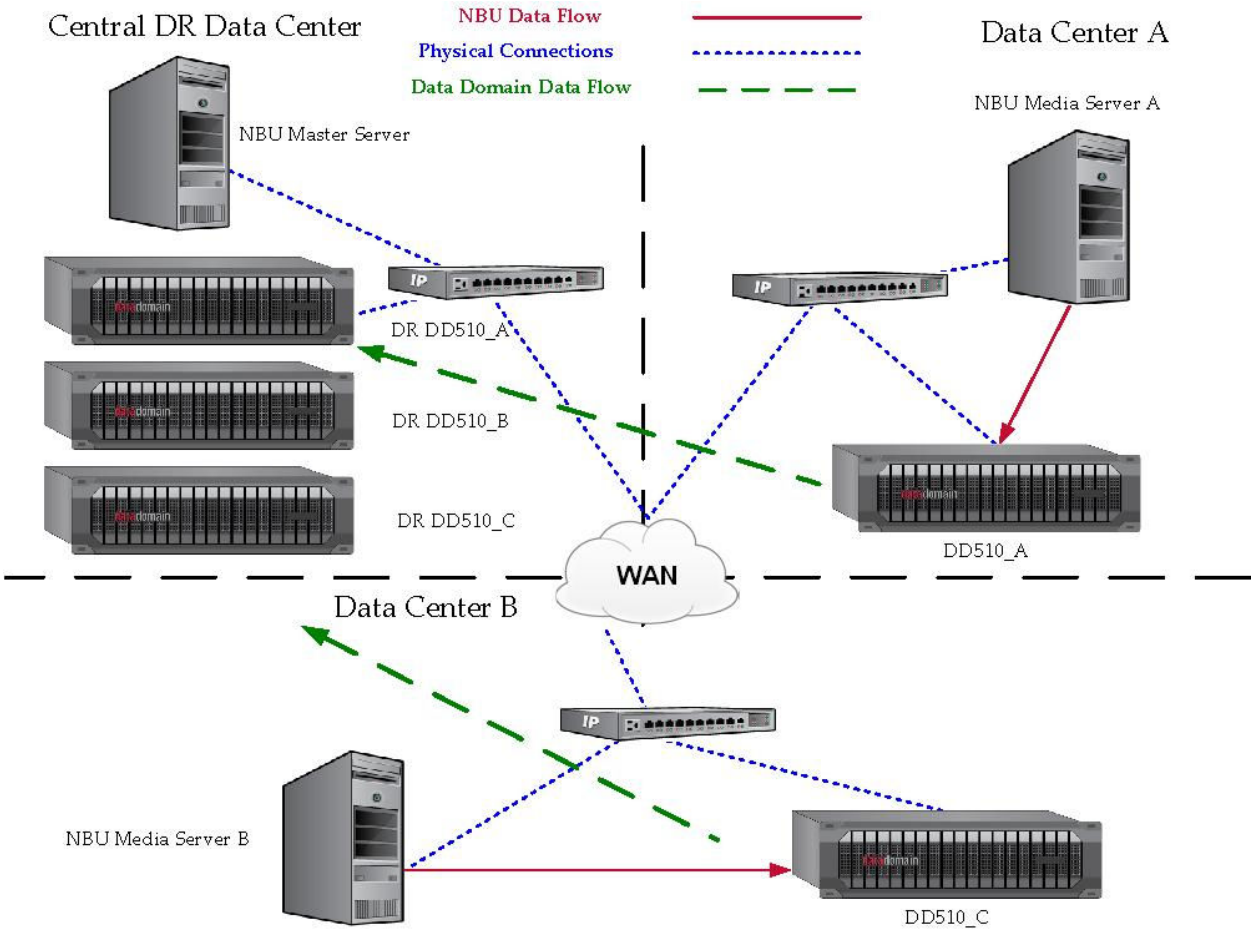

Figure 19: Central Disaster Recovery Data Center Configuration

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 NBU backup data to a single Data Domain system instance the DR site using Data Domain replication. A customer utilizing this architecture would realize an additional benefit from deduplication of data from all sources at the remote site.

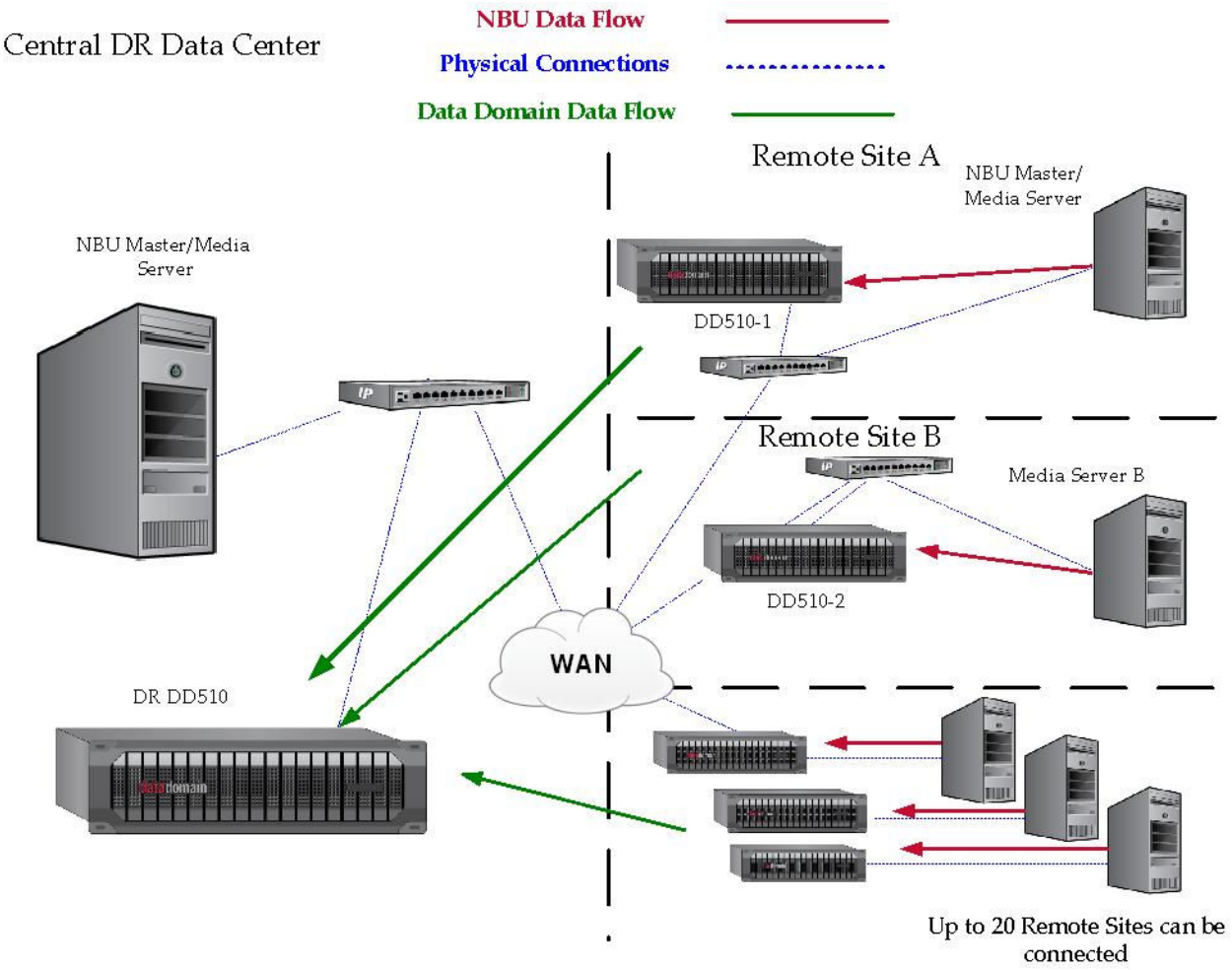

Figure 20: Remote Office Local and Disaster Recovery Configuration

Conclusion

Complete protection, cost effectiveness, and flexibility are key qualifiers for many IT organizations when considering new storage technologies. Implementing Data Domain inline deduplication storage in new, or existing, architectures addresses each of those requirements for NBU customers. Data Domain offers important redundant protection and a fast and reliable, replacement for traditional physical tape storage, while greatly reducing the outside costs incurred from those traditional tape resources. NBU customers will have the advantage of the availability of a variety of new storage architectures, each able to save man hours and limit incurred storage costs faced by aging infrastructure technologies.

The ease of integration into NBU environments only strengthens the already compelling case for Data Domain nearline storage to revolutionize the way data centers and remote sites manage physical storage

resources and replication for disaster recovery. For extremely large NBU instances, as with any storage implementation, appropriate planning, sizing and integration strategies will help make the Data Domain system's already efficient solution, excel at enterprise backups.

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.

Disclaimer

While GlassHouse has conducted extensive research and relies upon resources that it believes are trustworthy, it cannot guarantee their accuracy or the resulting representations or opinions of GlassHouse set forth herein; and, accordingly, GlassHouse disclaims all warranties.