WHITE PAPER: TECHNICAL

symantec.
Confidence in the connected world.

NetBackup Disk Based Data Protection Options

Alex Davies

Version 2.0 – Apr. 2009

Contents

1.0 Introduction

The move towards disk based data protection solutions is the single greatest market trend in data protection today. Disk based data protection solutions improve data protection performance, reliability, and disaster recovery capabilities. They lower cost, decrease backup windows, increase flexibility, and provide unique protection capabilities when combined with software technologies such as volume management, deduplication and replication.

NetBackup has supported simple disk based data protection using disk storage units and VTLs for several releases but with the release of NetBackup 6.5 several new and enhanced disk-based features were introduced, including the following:

- Intelligent disk management capabilities to enable more efficient use of server and storage resources.
- Load balancing across multiple NetBackup media servers to improve overall backup performance
- Pre-allocation of disk space to prevent backup failure due to lack of disk space.
- Deduplication features to eliminate redundant data from backup storage.
- Integration with intelligent disk storage devices from leading vendors to allow the user to get optimal value from these solutions.
- Support for 'off host' duplication with VTLs to provide tracking within NetBackup for backup duplication carried out within the VTL.

The value of these new and enhanced capabilities is much more than the sum of their individual contributions. The real benefit is in the way they are combined together to build powerful solutions to meet the challenges of protecting ever increasing amounts of data in IT environments where strained IT budgets, high levels of complexity, and strict compliance regulations are the norm.

There are five different types of disk storage devices supported by NetBackup 6.5.

Disk types available prior to the 6.5 release:

- **BasicDisk –** simple backup to a directory path on a single disk volume with the option of 'first in first out' staging to tape.
- **Virtual Tape Library (VTL)** – a disk storage device that uses a software layer to emulate a tape library. VTLs are an increasingly popular replacement for physical tape libraries as they offer simplified management and reduced operational costs.

New disk types introduced in the 6.5 release:

- **AdvancedDisk** – multiple disk volumes presented as a single storage pool on a media server providing dynamic storage allocation.
- **PureDisk Deduplication Option –** de-duplicated disk storage leveraging commodity disk.
- **OpenStorage Option –** API level integration with 3rd party intelligent disk storage devices.

This paper explains each of these disk options and their capabilities as well as the other new features in NetBackup 6.5 that make up the Disk Foundation. The paper also provides guidance in choosing the appropriate disk options for your environment.

1.1 Glossary

The following descriptive terms are used throughout this document:

Disk based data protection – the process by which backups are written to, and stored on, disk based data storage rather than tape media as is common in traditional backup models.

Storage unit – a logical target to which NetBackup writes backup data. Storage units may map to either disk or tape storage. The precise nature of the mapping depends on the type of storage unit.

Storage unit group – a collection of storage units that are addressed as a single entity by backup and duplication jobs. Individual storage units within the storage unit group that are used for particular operations depend of the rules associated with the storage unit group.

Disk pool – a collection of disk volumes that are presented as a single pool of storage that can be used by one or more storage units

Virtual Tape Library (VTL) – a SAN attached device combining array based disk storage with software that emulates the behavior of a tape library. Although the device is physically a disk array it appears to the backup software as a tape library with the storage divided into discreet 'cartridges' that are presented via a limited number of 'virtual' tape drives.

Storage server – the server component that controls the allocation of backup storage and directs the read and write traffic to it. In the case of tape, VTL and locally presented disk storage the storage server is always the NetBackup media server.

Intelligent disk storage device – a device that provides dedicated, intelligently managed storage that is more than just as simple disk array. In this paper the term 'intelligent disk storage device' is applied to intelligent devices that store backup data. Intelligent disk storage devices act as storage servers where they are deployed, taking over that role from the NetBackup media server.

Deduplication – the process of eliminating redundant data stored for backup and recovery purposes. Deduplication generally involves breaking down the data into segments, recording the segments for tracking purposes and writing unique segments to storage. Deduplication may take place during the initial write or during post-processing, depending on the device.

Storage lifecycle or lifecycle – a term describing the aging process of backed up data in which multiple copies stored on different types of storage expire at different times, usually resulting in a progressive increase in the time required to restore the data.

Storage lifecycle policy (SLP) – a mechanism which controls the creation and retention of multiple copies of the same backup data providing a plan or map of where backup data will be stored and for how long.

RPO (recovery point objective) – a service level figure defining the maximum allowable time, prior to a failure, to which an application must be recovered. For example, a recovery point objective of 24 hours means recovered data should no more than 24 hours older than the data at the time of failure. Actual recovery points may differ from established objectives.

RTO (recovery time objective) – a service level figure defining the maximum time allowed to recover an application, server or series of application or servers following a failure. Again it should be noted that the actual recovery time may differ from any established objective. Proper planning and testing needs to be carried out in order to assure that the actual recovery time aligns with the desired recovery time objective.

1.2 Additional resources

The following documents provide more background on the subjects discussed in this paper:

- The NetBackup Hardware Compatibility List and be found at http://support.veritas.com/docs/284599
- The white papers "Implementing Highly Available Disaster Recovery with NetBackup" and "Implementing storage lifecycle policies" can be found on the NetBackup whitepapers site at http://www.symantec.com/business/products/whitepapers.jsp?pcid=2244&pvid=2_1

2.0 NetBackup Disk Foundation

The Disk Foundation introduced in NetBackup 6.5 consists of a collection of new disk types and features designed to optimize the way backups are created using these disk types. Figure 1 below shows the four disk storage types that make up the Disk Foundation and some of the key features of each of them together with some common features which apply to all disk types. These features are explored on more detail in the subsequent parts of this section.

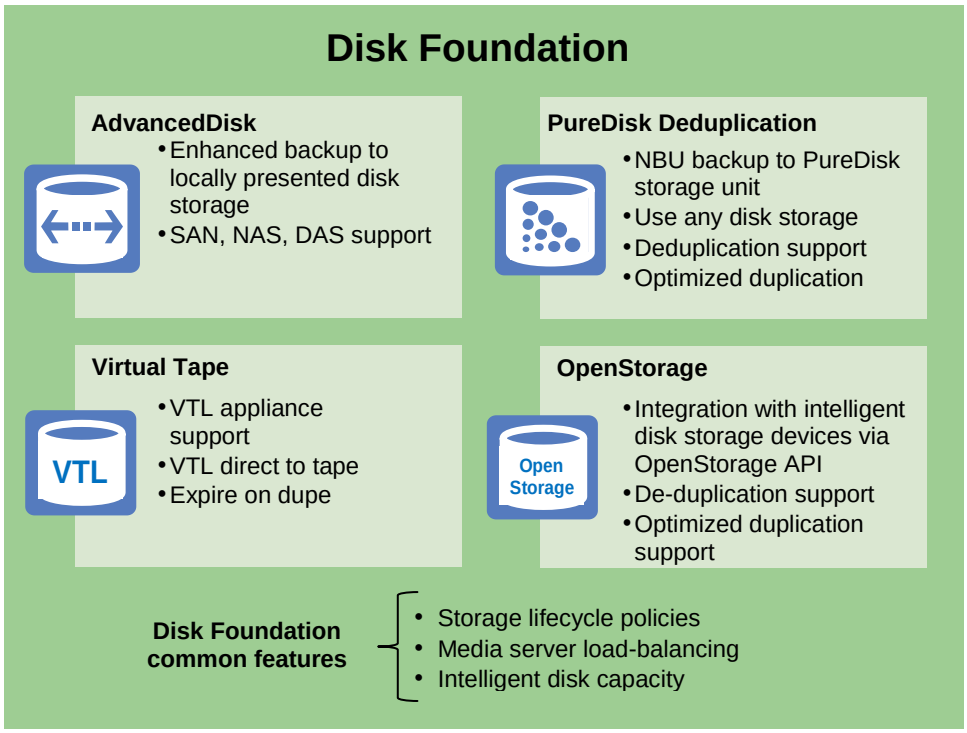

Figure 1 – Overview of the Disk Foundation Features

2.1 Media server load balancing and failover

Media server load balancing (MSLB) extends the concept of storage unit groups, originally introduced in NetBackup 5.1, to create a proactive system for routing client backups to the least heavily-loaded media server in a storage unit group. MSLB ensures that no single media server becomes a bottleneck and avoids the single point of failure of a 'downed' media server.

MSLB increases server utilization and availability by enabling NetBackup media servers to be grouped so that backup, restore, and duplication jobs are load-balanced across the group of media servers automatically and dynamically. The algorithm for selecting the best media server will assess the CPU and memory utilization, and number of jobs currently in progress on each media server to select the "best" media server for the next NetBackup job. If a media server fails for any reason, new jobs will be assigned to other media servers in its group until the failed media server is repaired so no new backup jobs will fail. Where checkpoint restart is possible, backup jobs running at the time of a media server failure are check pointed and will resume automatically on another media server within the same group.

The 'grouping' of media servers is achieved in one of two ways:

1. Media servers are defined as associated with a pool of disks (this may be either as 'members' of a disk pool or by means of access credentials for an intelligent disk storage device).
2. Media servers are associated with a storage unit group with the "load balance" radio button checked in the definition configured through the NetBackup Administration GUI.

media server load balancing looks at all media servers in a storage unit group or disk pool and directs the backup to the least heavily loaded server. It uses a load checking algorithm that measures CPU and memory utilization on each media server (ranking the server from 0, heavily loaded, to 3, idle) and the number of jobs running on the server.

2.2 Intelligent disk capacity management

Intelligent disk capacity management increases backup job success rates by automatically selecting an available disk volume for the backup target based on available capacity, estimated job size and the pre-allocating or "committing" the space on that volume so that other backup jobs do not attempt to use the same space.

NetBackup places a priority on writing the backup image to a single volume if possible; if no volume is large enough to store the entire backup image, multiple volumes can be spanned as necessary.

NetBackup will treat a disk as "full" if the sum of both the committed space and the used space exceeds the high-water mark. If a disk is determined to be full with this method, NetBackup will not use the disk until space is released on it. This feature eliminates the majority of backup failures due to full disk conditions.

Intelligent Disk Capacity Management only works with AdvancedDisk, OpenStorage, and PureDisk, which allow the amount of available space to be accurately determined, and cannot be used with BasicDisk storage units or with VTLs (which appear to NetBackup as tape).

2.3 Storage lifecycle policies

A storage lifecycle policy allows administrators to automatically specify where backup data will be stored and for how long. The storage lifecycle policy determines the locations where the backup is initially written and the destinations where it is subsequently duplicated to as well as the period of time that each copy of the backup will be retained. It also automates the duplication process making storage lifecycle policies an ideal technology for implementing standardized protection policies.

A storage lifecycle policy primarily consists of two core components: a list of storage destinations where copies of the backup images will be stored, and the retention period for each copy. A storage lifecycle policy can replace both the conventional duplication and staging processes with a single automated processing sequence by introducing a series of storage locations, or destinations, using different types of storage with different retention periods. Storage lifecycle policies ensure backup data always exists at the appropriate locations at the appropriate phases of the lifecycle.

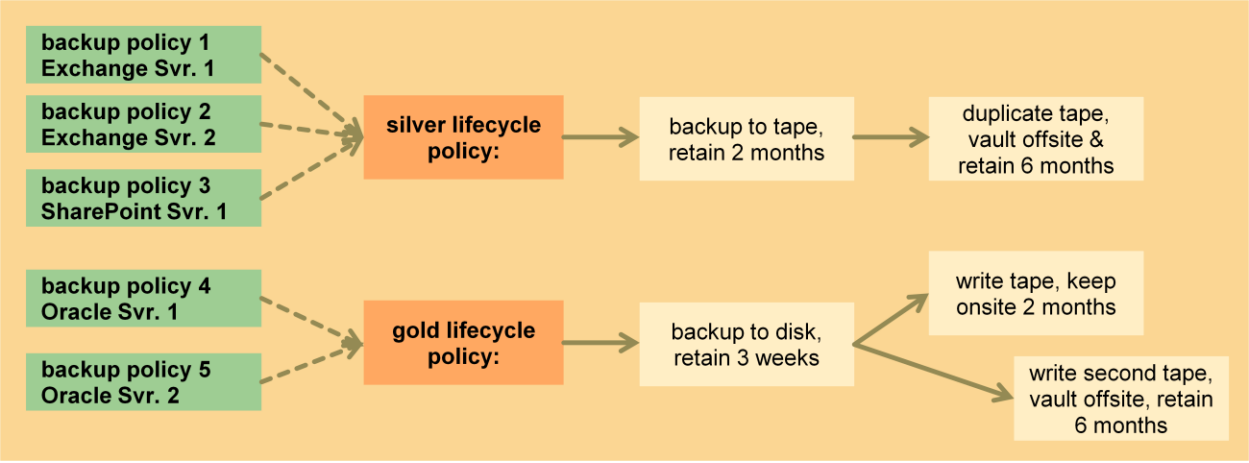

Figure 2 - Retentions within Storage Lifecycle Polices

Figure 2 above shows how two storage lifecycle policies can be applied to several backup policies to ensure that the correct number of copies of each backup is created and retained for the correct period of time.

A storage lifecycle policy is reusable by many NetBackup Backup Policies. If a storage plan changes (e.g. a new regulation is imposed on your business requiring changes to retention periods or the number of copies created), you simply need to change a small number of storage lifecycle policies and all of the backups will automatically employ the changes.

Storage lifecycle policies do not migrate data from one location to another progressively over a period of time. Instead, they ensure that all the image copies are created as soon as possible and apply different retention periods to each copy. Copies held on higher cost, higher performance storage typically have shorter retention periods than copies held on lower cost, slower access storage and space on this premium storage is released sooner. After the original backup completes, the storage lifecycle policy process creates copies of the image, retrying as necessary to ensure that all copies are successfully created.

2.3.1 Capacity managed retention and disk staging

Storage lifecycle policies allow the use of a new feature called capacity managed retention. Capacity managed retention has been introduced to support smart cleanup of backup images staged to disk pools. With capacity managed retention, the retention period is the time the image should ideally be kept on the storage device. This time period is known as the “try-to-keep time”. NetBackup will keep the image copy until such time as the space is required for other backups. This may be a longer or shorter period than the try-to-keep time depending on the demand for space. Images are expired and removed from disk when the disk high water mark is met with the less critical images being removed first, starting with the oldest ones, which allows the more important images to be retained as long as possible.

capacity managed retention can be applied to all disk based backup copies in a storage lifecycle policy, however the longest retention period used by any copy must be fixed rather than capacity managed. Consequently, if copies are only created on disk, at least one copy must have fixed retention and that retention period must be longer that any of the capacity managed retention periods. Copies of backups written to tape must have fixed retention periods and these must also be longer than any capacity managed retention periods defined for disk based copies. If a storage device cannot be configured in NetBackup to have high and low water marks (e.g. VTLs or other devices that present disk storage as an emulation of tape storage) it cannot be used with capacity-based retention.

2.3.2 Data classifications

Some types of data are more important than others (for example, medical records or financial data are probably more important than the pictures from the office party) and these types of data often have tighter RPO and RTO requirements than less critical data. Data Classifications have been introduced in NetBackup 6.5 as a mechanism for classifying backups by importance. Data Classifications can be applied to both Backup Policies and storage lifecycle policies and, when used in conjunction with capacity managed retention, enable administrators to get higher utilization out of their storage hardware while providing potentially faster restore times on their most important data.

Data Classifications are used by storage lifecycle policies to apply a rank to backup images written to the same capacity managed storage units. Backup Images are then retained for different periods of time based on their importance, overriding the traditional ‘first in first out’ model associated with Disk Staging storage units. Data Classifications ensure that the more critical data (as identified by administrators) remains on the fast recovery storage hardware in preference to less critical data that may be sharing the same storage.

Data Classifications can also be used in NetBackup 6.5 as a simple way of identifying, grouping and reporting on Backup Policies with a common level of importance. Data Classifications can be used even where Storage Lifecycle policies are not implemented. (Reports based on Data Classifications are available through Veritas Backup Reporter.)

Storage lifecycle policies and data Classifications are discussed in more detail in a separate whitepaper titled *“Implementing storage lifecycle policies”.*

2.4 Disk pools

Disk pools form one of the key concepts underlying AdvancedDisk. It significantly changes the way in which available disk space is used. A disk pool groups a set of disks together to form a single block of storage that can be shared among multiple storage units and, in some cases, multiple media servers.

With AdvancedDisk the disk pool provides a pool of storage for use by the storage units, replacing the more conventional one-to-one mapping between disk and storage unit provided by the BasicDisk model. In BasicDisk configurations each storage unit has access to a single disk volume or part thereof. With AdvancedDisk, the storage units can access all the disks in a disk pool and the disk used for a particular backup is selected based on the amount of space available. In effect the entire disk pool appears as a single disk to reach storage units.

The following table summarizes the advantages of using AdvancedDisk with disk pools over BasicDisk volumes.

Feature	BasicDisk	AdvancedDisk	Benefit of Disk Pools
Storage sharing	Fixed capacity based on a single disk. Multiple storage units can be created on a Single media server but each one requires a dedicated disk.	Multiple storage units on the same media server can share a disk pool. Space is allocated dynamically as required.	Automatic space allocation reduces the risk of backup failure.
Intelligent capacity management	No checks on available space before backup starts	"Will Fit" selection selects the storage unit based on 'available' space.	Pre-selection of storage ensures efficient backup operation.
Simplified growth	No built in ability to increase the size of individual storage units when disks fill up	Disk pool capacity is increased by adding additional disk volumes into the pool.	Minimal reconfiguration required as environment grows. Improved utilization of disk.
Media server load balancing	Does not support load balancing in storage unit groups	Supports load balancing – directing the backup to the least busy media server	Eliminates media server bottlenecks to increase the performance and success of backup jobs.
Alternate media server restore	Does not support access by more than one media server – if the media server is down the data is not accessible	SAN attached storage configured for AdvancedDisk sharing can use any media server associated with the disk pool for restore. If a media server is down another media server can do the restore.	Supports the presentation of storage to multiple media servers for added resilience
Storage lifecycle policies	Staging achieved using individual storage units with staging schedules	Data can be classified and managed differently throughout its life based on the importance of the data. After it has been duplicated, data can be selectively expired based on the relative importance to the business.	'Important' backups are available for rapid restore for longer periods, thereby improving RTOs.

Table 1 – AdvancedDisk disk pools vs. BasicDisk storage

Figure 3 below shows the difference between a media server with access to BasicDisk storage units and a media server with access to an AdvancedDisk disk pool.

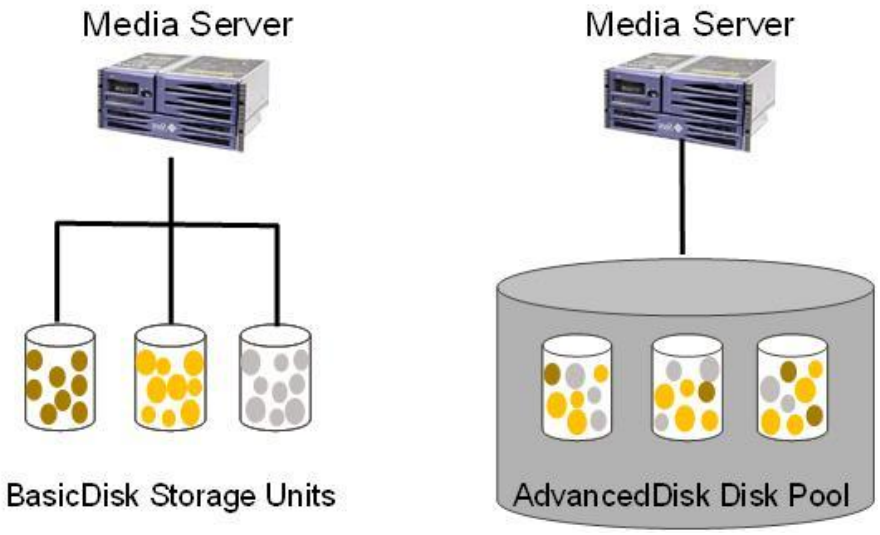

Figure 3 - BasicDisk vs. AdvancedDisk

In the BasicDisk configuration the three disks act as separated storage units and the backups are written to specific disks. Backup jobs do not use all three disks unless they are configured to form a storage unit group and there is no guarantee that the disk space is used efficiently among the three independent disks.

AdvancedDisk utilizes space more efficiently than BasicDisk. In the AdvancedDisk configuration the three disks form a single pool and backups may be written to any of the disks. When a backup runs, Intelligent Disk Capacity Management will automatically estimate the size of the backup and select the disk with the greatest amount of free space and will reserve space to match the estimated size of the backup image so that other backup jobs do not over commit beyond the disk's available space. Note how the different colored backup images reside on different disks in the BasicDisk configuration but are distributed across all of the disks in the AdvancedDisk disk pool.

Storage units that use AdvancedDisk disk pools on different media servers can be configured to form 'load balanced' storage unit groups. When a backup is run using the storage unit group, the storage unit on the least busy media server is automatically selected and a suitable disk within the AdvancedDisk disk pool on that media server is used for the backup.

An AdvancedDisk disk pool can be formed from any disks that can be seen by, and are mounted to, the media server. However it is recommended that all the disks in a particular disk pool should have common characteristics (size, speed, connection etc.) to ensure consistent behavior when different disk are selected for use.

In the initial NetBackup 6.5 release AdvancedDisk disk pools could only make use of disks presented locally to a single media server, requiring multiple disk pools and storage unit groups to achieve media server load balancing. With the introduction of 'AdvancedDisk sharing' in NetBackup 6.5.2 AdvancedDisk disk pools have been able to make use of disks presented to multiple media servers at the same time over NFS and file system clustering technologies, such as Veritas Cluster File System, to 'share' disk pools between multiple media servers. Configuring in this way requires only a single storage unit to achieve load balancing across a number of media servers.

2.5 Intelligent disk storage devices

An intelligent disk storage device combines disk storage with software and is presented to NetBackup media servers as a device rather than a simple collection of disks. Prior to the release of NetBackup 6.0 the only devices of this nature that could be used with NetBackup were virtual tape libraries (VTLs) which use software to present the disk storage as if it was a tape library containing a number of discrete tapes and tape drives. NetBackup 6.5 extends this capability through integration with PureDisk and the OpenStorage API which allows

the use of a wide range of OpenStorage compliant intelligent disk storage devices. OpenStorage compliant intelligent disk storage devices generally provide deduplicating capabilities and other features such as device level replication and energy efficient storage. These 'second generation' intelligent disk devices can offer significant advantages over VTLs because they present the disk storage as disk rather than emulating tape storage and can thus make full use of the Disk Foundation features. Table 2 below summarizes the advantages of these intelligent disk storage devices over VTLs.

Feature	Virtual Tape Library	Intelligent Disk Storage Device	Benefit
Parallel processing	Only one read or write operation allowed at a time per virtual tape significantly extending the time required to effectively protect data.	Concurrent read or write operations can occur per volume potentially completing all jobs in a much shorter time.	Shorter backup windows.
Resource sharing	Only one media server can read or write to a virtual tape at any given time	Multiple media servers can read or write to the same volume concurrently.	More efficient resource sharing between media servers.
Granularity – duplication	The virtual tape is smallest unit of capacity. Images written to a virtual tape cannot be copied until all read or write operations to the virtual tape are complete.	The backup image is the unit of capacity. Images on a disk volume can be copied while other images on the same volume are being read or written.	More efficient duplication cycles.
Granularity – expiration	The virtual tape is smallest unit of capacity. All images on the virtual tape must expire before any space is released.	The backup image is unit of capacity. Free space is available as soon as a backup image is expired.	More efficient utilization of storage.
Duplication	NetBackup is unaware of any replication that takes place at the VTL level.	Optimized duplication ensures NetBackup is aware of replication activity.	Copies created by replication can be used for restore by NetBackup.
Capacity Management – selection	Virtual tapes must be filled in sequence.	Volumes can be selected based on available space.	Intelligent capacity management for better storage utilization.
Capacity Management – staging	Virtual tapes cannot be used for disk staging as NetBackup does not know how full each virtual tape is.	Volume high and low watermarks allow staging operations.	Capacity managed expiration.

Table 2 – Virtual tape library vs. Intelligent disk storage device

3.0 NetBackup disk types

This section examines each of the six disk types identified in the introduction, explaining the key features of each type and their relative strengths and weaknesses.

3.1 BasicDisk

BasicDisk is any simple disk volume that is made available to NetBackup as a target. It could be a mounted volume from a NAS share, direct attached storage within a NetBackup media server (e.g. the C:\ drive) or a SAN attached volume.

BasicDisk storage units can be configured as 'staging' areas from which backups are duplicated to alternative storage but this process is limited to a 'first in first out' behavior requiring individual storage units to be configured for backups of differing importance.

3.1.1 What's good about BasicDisk?

- **Low capital cost** – BasicDisk is included in the standard NetBackup server license and standard infrastructure license and can also be used with low-end disk so it has the lowest purchase costs of any disk type. However the absence of intelligent automated management features means that the total cost of ownership of BasicDisk may be higher than some of the other disk options discussed in this paper.
- **Wide hardware support** – BasicDisk can be used with any disk storage that can be presented to a media server with a file system that can be written to, irrespective of the connection method or disk characteristics.

3.1.2 What are the downsides of BasicDisk?

- **Increased management overhead:**
  - BasicDisk does not support Disk Foundation features such as storage lifecycle policies, Intelligent Disk Capacity Management and media server load balancing which provide automated resource management
  - Unlike other disk types described in this section, BasicDisk storage units are confined to the individual disk volumes. BasicDisk storage is not presented or managed as pools of disk volume where extra capacity can be added by simply adding additional volumes into the pool. Instead, creating extra capacity requires configuring additional BasicDisk storage units and changing backup policy settings or storage unit group policies.
- **Staging limitations** – while BasicDisk storage can be used as a 'staging' area prior to writing to tape, BasicDisk storage units operate on a 'first in first out' model and do not support the more advanced staging through lifecycles.

3.1.3 When should I consider using BasicDisk?

BasicDisk can be used for simple disk staging or smaller disk backup environments but may not be appropriate for wide spread use in larger environments where performance, resource utilization, automation, and high availability are important. BasicDisk is especially well suited to support NetBackup catalog backups due to the simplicity of the disk storage presentation to NetBackup, which simply uses a server name and a path. NetBackup's EMM database can be easily rebuilt (or remapped to see the replicated volume instead) in order to restore the catalog.

The lack of support for Disk Foundation features limits the capabilities of BasicDisk. BasicDisk should not be used in any configuration where load-balancing across servers and disk volumes is required or where advanced staging capabilities such as capacity managed retention are needed.

3.2 AdvancedDisk

AdvancedDisk increases the utilization and performance of disk by combining disk resources together as a single pool rather than treating them as smaller, separate islands of disk. AdvancedDisk also extends the capability of BasicDisk through the Disk Foundation features which increase media server utilization and availability while improving backup success rates and performance. Instead of creating storage units from single disk volumes, AdvancedDisk allows you create storage units containing multiple disk volumes defined within a disk pool.

3.2.1 What are the benefits of AdvancedDisk?

- **Disk Foundation feature support** – AdvancedDisk can make use of features such as intelligent capacity management, storage lifecycle policies and media server load balancing
- **Wide hardware support** – like BasicDisk, AdvancedDisk can support most types of disk storage presented to a media server.

- **Easy upgrade path from BasicDisk** – AdvancedDisk uses the same hardware as BasicDisk and BasicDisk storage units can easily be upgraded to AdvancedDisk storage units provided the storage unit already exclusively owns the entire file system.
- **Reduced management overhead** – automated infrastructure reduces the level of administrator attention required by distributing job load and intelligently managing disk capacity to ensure that jobs do not fail because of out of space conditions until all the available disks are full.
- **Dynamic storage allocation –** because the storage units access a pool of disks, more capacity on other disks is automatically 'added' to a storage unit when a disk fills up The disk pool model also makes it simple to increase overall storage capacity as additional disks can be added to the disk pool without needing to reconfigure the storage units which use it.
- **Shareable storage –** with effect from NetBackup 6.5.2 it is possible to share an AdvancedDisk disk pool between multiple media servers. This capability depends on the storage being presented via a shareable technology such as NFS or Veritas Cluster File System (CFS)

3.2.2 When should I consider using AdvancedDisk?

AdvancedDisk is well suited for use where fast and flexible backup to disk is required, especially for short term retention, synthetic full backup and staging prior to writing data off to other storage media such as tape. AdvancedDisk can work with "whatever disk you have laying around" and assemble it into a highly utilized disk target it. This makes it an attractive option for customers that may have spare disk storage which is no longer required for primary storage and can be re-purposed as backup storage. The load balancing and capacity management features of Disk Foundation mean that AdvancedDisk offers clear advantages over BasicDisk in terms of more efficient disk and server utilization.

Using AdvancedDisk with a sharable file system such as NFS or CFS removes the single point of failure at the media server level that exists with BasicDisk.

3.3 Virtual Tape Library (VTL)

VTLs use disk arrays to emulate the behavior of tape drives and tapes in tape libraries using 'virtual tapes'. Their main advantages are that they are more reliable, require less user intervention and occupy less space than their physical tape counterparts. One reason they are more reliable is that virtual tape devices have no minimum streaming speed so they do not suffer damage as a result of slow write rates in the way that physical tape devices do. VTLs eliminate the management problems surrounding physical tape handling, thereby improving the reliability of backups.

VTLs are easy to operate as they behave like physical tape libraries assigning and deassigning space in blocks equivalent to the capacity of a virtual tapes. Virtual tapes appear to operations staff exactly the same as the tapes they are familiar with. Many VTLs also offer capabilities to replicate data or write it to physical tape but as this usually happens outside of NetBackup control, the resulting copy is not known to, and often not directly accessible by, NetBackup.

3.3.1 What's good about a VTL?

- **Ease of integration** – as VTLs appear exactly like tape libraries in NetBackup there are no requirements to change operational practices.
- **Platform independent sharing** – VTLs are generally platform independent and can be connected to multiple media servers of different types.
- **High performance with parallel write** – most VTLs are optimized for writing, allow large numbers of dedicated 'virtual tape drives' to be configured to allow fast writing to multiple virtual tapes in parallel. In practice, restrictions on the read rate of the source data means that these high transfer rates cannot be achieved.
- **Intelligent storage options** – many VTLs offer additional capabilities such as deduplication, write through to physical tape and replication between VTLs.

- **Limited Disk Foundation feature support** – VTLs can make use of storage lifecycle policies and media server load balancing but do not support any form of intelligent capacity management.

3.3.2 What are the downsides of a VTL?

- **Inefficient use of storage space** – virtual tapes suffer from the same limitation as physical tapes, until the backups on a virtual tape have expired none of the space is released. Even if a virtual tape has potential space available (the images have expired), the space cannot be re-used due to the constraints imposed by sequential tape access, all the images must expire before the virtual tape can be 'scratched' and re-used.
- **No intelligent capacity management** – as VTLs emulate tapes they cannot take advantage of Disk Foundation's ability to direct backups to the disk with the most available space. VTLs follow the same rules as physical tapes and backups are appended to partially full 'virtual tapes' without checking how much free space remains on at tape.
- **Restricted access to tapes and resident images** – pre-written images are inaccessible while tape is in use, blocking access for restores and delaying duplication until backups complete. Parallel access is not possible – only one image can be accessed at a time, consequently holding all other images hostage.
- **Content replication is not always tracked in NetBackup** – if virtual tape is replicated to a physical tape or another VTL using the VTLs internal functions, the replicated copy is not known to NetBackup and can only be use if the original copy is deleted or masked so that it is not visible. NetBackup also has no ability to limit which images are replicated should you wish to conserve space and follow SLAs where only a subset of the most important images need to be copied. Some VTLs make use of embedded NetBackup media servers or the NetBackup 'VTL-direct-to-tape' feature to avoid this problem by duplicating to virtual or physical tapes with different media identifiers. In these cases NetBackup is aware that the duplication has taken place even though the operation did not involve a media server directly.

3.3.3 When should I consider using a VTL?

VTLs represent an easy option when moving from a tape based solution to a disk based solution as they behave in the same way as tapes. VTLs make a good choice for improving backup performance and reliability when there is a limited ability to architect a true disk backup solution. Customer's looking to maximize performance and efficiency that have the flexibility to modify their architecture may consider other disk options.

3.4 OpenStorage

The OpenStorage initiative is an exciting disk-based innovation which allows NetBackup to take advantage of the new technological advances of intelligent disk storage devices, including: storage reduction, backup image duplication, synthetic backups, replication and energy efficiency. The tight integration with third party intelligent disk storage devices offers NetBackup users improved management of their backup images and added functionality without the limitations of tape emulation.

Symantec has developed the OpenStorage API that lets intelligent disk storage device vendors write plug-ins to their device for NetBackup media servers. This provides NetBackup with visibility into the properties and capabilities of the intelligent disk storage device, and control of the backup images stored in the device. The intelligent disk storage device can then be treated by NetBackup as disk devices rather than tape devices, as in the case of VTLs. Through the OpenStorage API, NetBackup controls when backup images are created, duplicated, and deleted. The Intelligent disk storage devices control how the images are stored in and copied between devices. In this manner, the OpenStorage vendors add unique business value to the overall solution via specialized technological innovations such as backup image deduplication, WAN-optimized backup image replication for disaster recovery purposes, power management, and a wealth of other innovations.

With OpenStorage customers can utilize a variety of qualified, best-of breed, disk-based solutions sold by a number of device vendors. As with VTLs, customers benefit by having an easy-to-deploy, purpose-built, disk-based protection solution. Unlike virtual tape solutions, OpenStorage intelligent disk storage devices present disk storage to NetBackup rather than presenting disk as tape. The result is that a number of limitations inherent to tape are avoided.

Integration with NetApp NearStore has been available since the release of NetBackup 6.0 and is based on an earlier implementation of the OpenStorage API than the one implemented in NetBackup 6.5. The NetApp devices that integrate with the newer OpenStorage API, taking advantage of Disk Foundation features such as media server load balancing that are not available in the existing NearStore integration, are expected to be available in 2008. To obtain list of supported OpenStorage vendors refer to *the NetBackup Hardware Compatibility List*.

3.4.1 What's good about OpenStorage?

- **Disk Foundation feature support** – through OpenStorage intelligent disk storage devices can make use of NetBackup features such as intelligent capacity management, storage lifecycle policies and media server load balancing
- **Leverages the capabilities of the intelligent disk storage device** – the capabilities vary between different devices.
    - **Reduced storage requirements** – devices that support deduplication require significantly less back end storage than the conventional disk types supported by BasicDisk and AdvancedDisk.
    - **Deduplicated replication for improved disaster recovery –** NetBackup can take advantage of the replication capabilities of the OpenStorage devices to "off-host" duplicate NetBackup images to secondary devices. Space optimized replication capabilities allow 'off-site over the wire' capability. The cost of bandwidth, tape shipping, and storage associated with vaulting data to another data center is drastically reduced.
    - **Energy/power savings –** when used with devices that support energy saving MAID (massive array of idle disk) technologies.
- **Platform independent storage sharing –** the Intelligent disk storage device can be presented to multiple media servers running on different operating systems.
- **Single point of control –** the intelligent disk storage devices are controlled through NetBackup. The migration or copying (the storage lifecycle) of the data within the intelligent disk storage device is controlled through NetBackup.

3.4.2 What are the downsides of OpenStorage?

- **Deduplicating devices challenge management predictability:**
    - **Unpredictable storage demands** – space saving within the device is only as good as the vendor's deduplicating algorithm and some may not be as efficient as others.
    - **Unpredictable space reclamation** – when backup images are removed from conventional disk storage devices the amount of space freed up is directly proportional to the size of the backup image. For deduplicating devices there is no proportional relationship between the backup image and space occupied. This problem is compounded by some deduplicating devices that do not reclaim individual unique blocks as they are released. For this reason deduplicating devices are not suitable for use as 'staging' areas and may 'under report' available space to NetBackup's Intelligent Disk Capacity Management system.

3.4.3 When should I consider using OpenStorage?

Intelligent disk storage devices offer various benefits over conventional disk devices including deduplication, optimized duplication and reduced power consumption. They also offer many benefits over VTLs in terms of space utilization and operational efficiency. The decision on when to use an intelligent disk storage device, and which intelligent disk storage device to use, will depend on the user requirements.

3.5 PureDisk Deduplication Option

The PureDisk Deduplication Option (PDDO) allows a media server to store ordinary NetBackup data in a PureDisk Storage Pool leveraging PureDisk global deduplication technology. As with OpenStorage, NetBackup

maintains visibility and control of the replication and expiration process and thus the entire lifecycle of the backup image.

NetBackup clients can transmit backup data over the LAN or SAN to a collection of NetBackup media servers. These media servers then deduplicate the backup stream before storing it to the PureDisk 6.5 storage server. Existing or new PureDisk clients can also backup to the PureDisk Storage Pool.

3.5.1 What's good about PureDisk Deduplication Option?

- **Global deduplicated storage integrated directly into NetBackup –** With deduplication factors typically between 10 and 50, PDDO provides significant savings in storage costs. Its integration directly into NetBackup brings ease of management and a very low cost of ownership.
- **PureDisk capabilities for all NetBackup clients –** Any existing NetBackup client or NetBackup agent can store its data on the PureDisk storage server through the NetBackup media servers.
- **Disk Foundation feature support** – PureDisk can make use of storage lifecycle policies and media server load balancing. Intelligent capacity management does not apply to PureDisk devices as the entire storage pool appears as a single volume to NetBackup.
- **Encryption built in –** PDDO can automatically encrypt all data stored on in the PureDisk Storage Pool.
- **Central point of control –** The client backup data stored in the PureDisk storage server is fully managed by NetBackup.
- **Deduplicated replication –** NetBackup can take advantage of the WAN-optimized replication capabilities of PureDisk Storage Pools to "off-host" duplicate NetBackup images to secondary storage pools. Duplicated images can be stored with different retention levels, treated independently from the original copy, and be verified and cataloged by NetBackup.
- **Scalable performance and capacity –** A PureDisk Storage Pool is a modular system that can be scaled in performance and capacity by adding 'content router' nodes. Global deduplication happens across all content router nodes in the storage pool, maintaining deduplication efficiency even when storage nodes are added to the system. Data deduplication is performed on the NBU media servers. Multiple streams can be deduplicated in parallel on the media server and multiple media servers can deduplicate data in parallel as global deduplication optimizes the data across the whole dataset. Because the deduplication occurs on the NetBackup media server rather than the storage device the maximum number of media servers using a single PureDisk Storage Pool can be greater than the number using an equivalent OpenStorage device or deduplicating VTL.
- **Commodity disk –** PureDisk works with a broad list of disk systems (DAS, SAN, NAS, iSCSI) which makes it easily deployable in an existing storage environment.

3.5.2 What are the downsides of PureDisk Deduplication Option?

- **Possible higher peak CPU load on the NetBackup media server** – In order to stream only unique data to the PureDisk storage pool, the media server breaks the image into segments and checks to see if segments are currently stored within the storage pool. As only the single instanced data is sent to the PureDisk storage pool the outbound I/O overhead is reduced. This means the media server can handle more inbound I/O but it also means that there is more processing activity going on in the media server. To handle the deduplication processing in this situation the media server may require more peak CPU power than a similar server backing up to conventional disk. To reduce the CPU impact of deduplication, backup streams can be spread over multiple media servers while still maintaining deduplication optimization.
- **Additional management overhead** – because it is not a dedicated appliance in the way that VTLs and OpenStorage devices are, a PureDisk Storage Pool is more complex to configure and manage than these types of storage.
- **Commodity disk performance limitations** – because the PureDisk Storage Pool uses commodity disk rather than the dedicated tier 1 storage used in VTLs and OpenStorage devices the performance of the

storage may be less that for an equivalent VTL or OpenStorage device. This can translate into a slower restore speed and consideration should be given to the restore SLA when selecting the disks for use in the storage pool.

3.5.3 When should I consider using PureDisk Deduplication Option?

PureDisk Deduplication Option is ideally suited for longer term retention of data on disk because it offers dramatic space savings. Its built in replication also makes it ideal for customers who may look to replace tape warehousing with electronic vaulting to their own data centers or third party service providers. PureDisk is a software solution that allows using your preferred storage hardware. The PureDisk architecture scales in performance and capacity by adding media servers, and content router nodes. Customers who are considering VTLs or hardware based deduplication devices may want to consider PureDisk Deduplication Option. PureDisk provides excellent space reduction and optimized duplication through its deduplication and replication capabilities but is generally slower for both backup and restore than conventional disk storage.

4.0 Implementing a disk based data protection solution

The previous sections have shown there are now several different disk types available, each of which has its own unique set of capabilities – so which disk type is right for you? In many cases the 'right' choice of disk based data protection is a mixture of different disk technologies; AdvancedDisk for load balancing and performance, OpenStorage and PDDO for space and power efficient long term storage. This section compares the different disk types and how they can be used to best advantage and presents some examples of where combinations of different disk types are used to meet different requirements.

4.1 Choosing the right disk type for the job.

Choosing the right disk type to meet a specific requirement is crucial. There are three factors that will influence the decision process:

- Capability – is the type of disk suitable for the intended purpose?
- Cost of ownership – does this choice represent the best value for money?
- Compatibility – is this choice compatible with my existing infrastructure?

4.1.1 Capability

As we have seen, each disk type has its strong and weak points, which means that some disk types are more suitable than others for particular tasks.

Table 3 summarizes the capabilities of the five main disk types and illustrates their relative strengths in five key areas.

	VTL	OpenStorage	AdvancedDisk	PureDisk	BasicDisk
Backup Performance[1]	●●●●●	●●●○○	●●●●●	●●●●○	●●●●●
Recovery Performance[1]	●●●●●	●●●○○	●●●●●	●●●●○	●●●●●
Lifecycle management capabilities[2]	●●●○○	●●●●○	●●●○○	●●●●○	○○○○○
Storage Efficiency[3]	○○○○○	●●●●○	●●●○○	●●●●●	○○○○○
Deduplication capabilities[4]	●○○○○	●●●●○	○○○○○	●●●●●	○○○○○
Optimized replication/ duplication capabilities[5 & 6]	●○○○○	●●●●○	○○○○○	●●●●●	○○○○○
	○○○○○ = Least			●●●●● = Most	

Table 3 - Summary of Capabilities

Note: Both VTL and OpenStorage ratings will be highly dependent on the hardware selection.

1. Backup Performance and Recovery Performance will typically be impacted by deduplication or compression being utilized. If you are using an intelligent disk storage device check with your device provider for information on the typical performance impact. These figures will also be impacted by many factors elsewhere in the infrastructure such as network bandwidth and source data performance characteristics and the backup storage is seldom the overall performance bottleneck in a data protection system.
2. Some VTLs have “direct to tape” functionality for use with NetBackup which may increase their performance.
3. VTLs must mimic tape cartridges which limits recovering storage space.
4. PureDisk, most OpenStorage devices and some VTLs support deduplication. The deduplication capabilities of OpenStorage devices and VTLs vary widely from device to device. Check with your device provider for typical deduplication capabilities. AdvancedDisk does not support deduplication.
5. PureDisk and most OpenStorage devices support optimized duplication enabling ‘replicated’ data to be presented in two locations in the same NetBackup domain. Although some VTLs support replication it does not occur under NetBackup control and thus the replicated copy cannot be presented in the same domain.
6. Although the OpenStorage API supports optimized duplication the OpenStorage device may not support the feature. Check with your OpenStorage device provider for optimized duplication support.

4.1.2 Cost of ownership

The total cost of ownership of different disk types can be broken down into two components, the capital cost of the initial hardware and the on-going operational cost of running the backup environment.

The capital cost of appliance based solutions is generally higher than that of commodity disk based solutions as additional hardware and software components are required. However it is important when considering deduplicating devices such as PureDisk and intelligent disk storage devices to remember that the deduplicating effect means that ‘cost per gigabyte’ of storage may be significantly less than it appears as the actual space used on the device is less than on a equivalent conventional disk or VTL.

Operational costs of deduplicating solutions are also lower as the reduced storage capacity translates into reduced power and cooling requirements. The ‘self managing’ nature of the appliances mean they tend to be more reliable and require less operator intervention than commodity disk solutions and, by the same token, the Disk Foundation features mean that AdvancedDisk implementations also have a lower operating cost than BasicDisk implementations. Thus while a BasicDisk solution is likely to have one of the lowest capital costs per gigabyte it also likely to have the highest operational cost as there is no automatic management capability.

4.1.3 Compatibility

It is important to remember the following points when considering the choice of disk type:

1. Not all disk types are supported with all media server platforms and versions of NetBackup 6.5.x.
2. Intelligent disk storage devices and PureDisk are delivered outside of NetBackup’s packages and on different release schedules. Integration timescales with NetBackup are dependent on multiple development groups, including third-party vendors.

Reference should be made to the latest NetBackup compatibility lists to ensure that the proposed solution is supported with the existing or proposed infrastructure.

4.2 Disk based data protection solution examples

The following examples show how a combination of disk types can be used to meet the short term and long term backup requirements.

4.2.1 Balanced backup for optimal throughput

The use of a shared AdvancedDisk disk pool with storage lifecycle policies provides an efficient mechanism for making initial backups that can subsequently be duplicated to tape or deduplicating storage. Figure 4 shows clients backing up to a number of media servers which are operating in a load balanced configuration with a shared AdvancedDisk disk pool and a SAN attached OpenStorage device. Storage lifecycle policies are used to control the initial backups and the duplication of those backups requiring longer term retention to the OpenStorage device. The advantage of 'staging' the backups in this way is that the space can be very efficiently reclaimed from the AdvancedDisk disk pool as the backup expire.

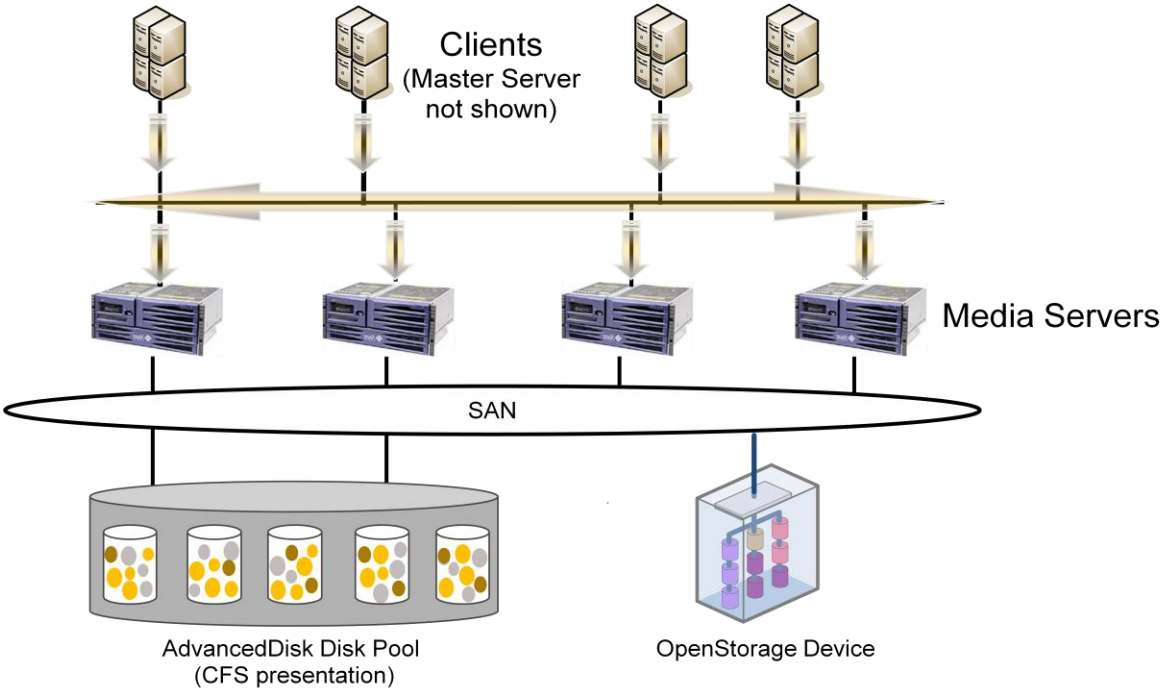

Figure 4 - Load balancing with a shared disk pool and OpenStorage

The down side of this configuration is that, because a shared disk pool presented using Cluster File System is used, it requires all of the media severs to be the same platform. **Error! Reference source not found.** below shows a similar configuration using local disk pools with a LAN attached intelligent disk storage device. In this case storage unit groups are used to achieve the initial load balancing across the media servers. Although this allows for a mix of media server types it does tie the backup to the media server and if a media server is unavailable for any reason the backup and duplication jobs that use it will fail. As storage lifecycle policies are used to control the duplication process the duplication jobs will continue to retry automatically and will catch up when the media server comes back on-line.

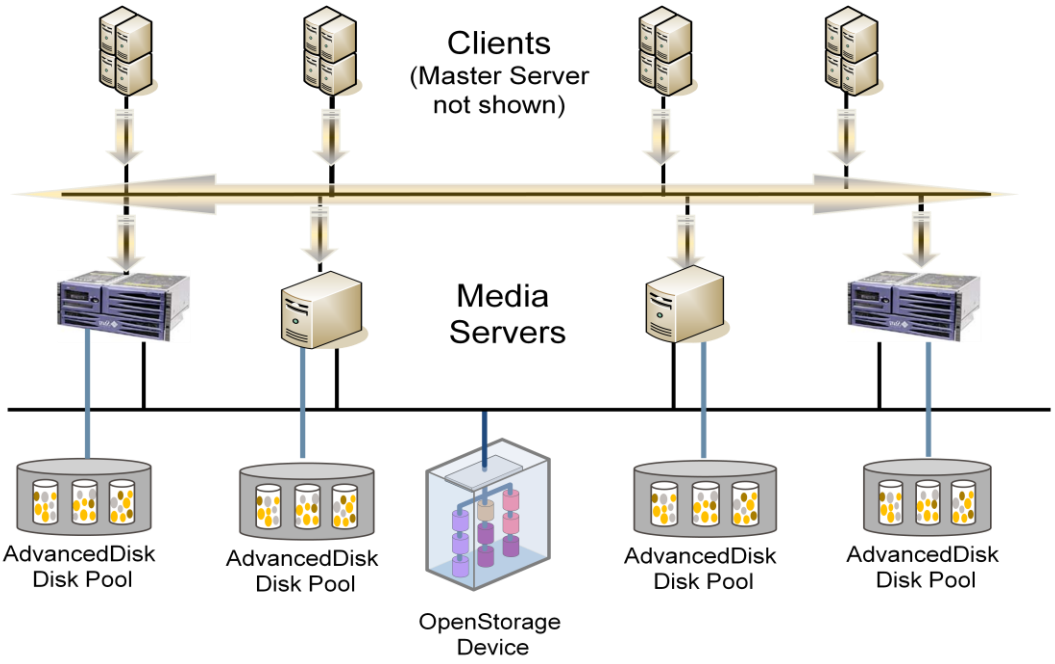

Figure 5 - Load balancing with local disk pools and OpenStorage

4.2.2 Combined data center and remote office protection using PureDisk

PureDisk Deduplication Option uses the same PureDisk infrastructure as the PureDisk Remote Office backup system. This means that a PureDisk storage pool located at a main data center can act as a remote replication server for a regular PureDisk storage pool located at a satellite data center. Figure 6 shows a configuration of this nature. In the main data center backups from clients are directed to a PureDisk storage pool by storage lifecycle policies.

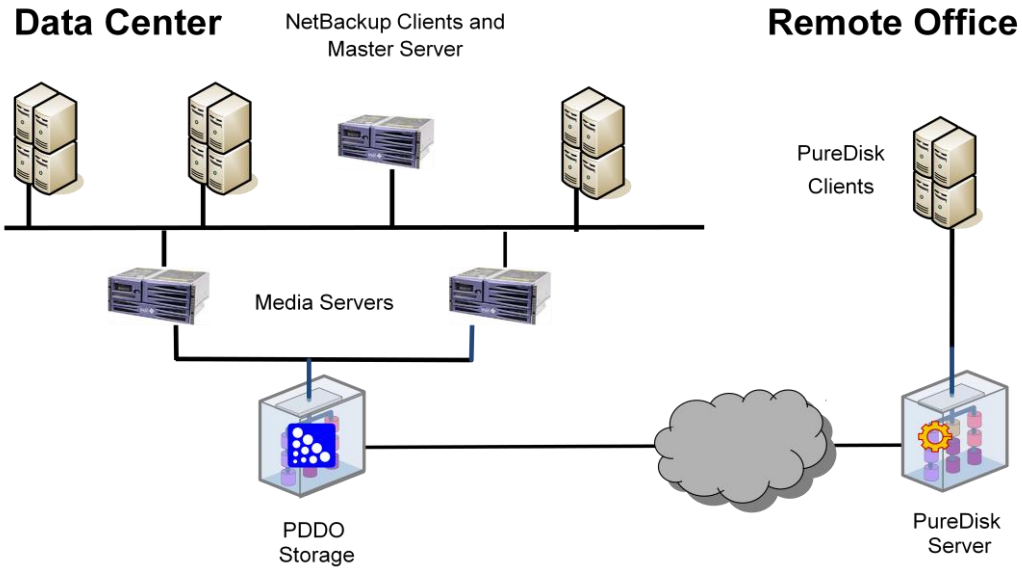

Figure 6 – Data center and remote office backup combination

At the satellite data center clients are backed up to a PureDisk storage pool using PureDisk clients. The PureDisk storage pool at the satellite data center is replicated to the PureDisk Storage Pool at the main data center, providing off-site protection to the backup data. At the main data center these backups from the satellite location

can, if required, be exported to NetBackup to allow rapid recovery of a failed client. This feature is particularly useful if the satellite site has no resident IT staff, as a failed client can simply be 'rebuilt' at the main site and shipped to the satellite site.

The backups from the main data center held on the PureDisk storage pool can also be replicated to the satellite data center to provide off-site protect. However as this occurs without NetBackup's knowledge backups replicated to the satellite site cannot be accessed directly by NetBackup and in the event of the PureDisk storage pool failing completely would have to replicated back to the main data center before they can be used for restore.

4.2.3 Highly available dual site/single domain configuration

The dual site/single domain model provides the ultimate in cost effective protection against site loss by combining two production sites into a single backup environment and ensuring that all critical backups exist in both locations. In a dual site/single domain configuration a single NetBackup domain spans two geographically remote sites, each of which effectively becomes the disaster recovery site for the other. Copies of the backups exist at both sites and the master server is clustered in such a way that control of the domain can be switched from site A to site B if necessary (this concept is discussed in more detail in a separate whitepaper, "*Implementing Highly Available Data Protection with Veritas NetBackup"*).

Figure 7 shows a Dual Site/Single Domain configuration with PDDO storage. Backups are written to the PDDO storage using storage lifecycle policies. The PDDO devices are configured to replicate to each other and the optimized duplication feature is used to update the NetBackup catalogs to include the copy in the PDDO device at the remote site when the replication has completed.

In this configuration the PDDO storage is not used simply to provide long term space optimized storage, it is also used to provide an automatic 'off site' capability by leveraging the optimized duplication capability of PDDO.

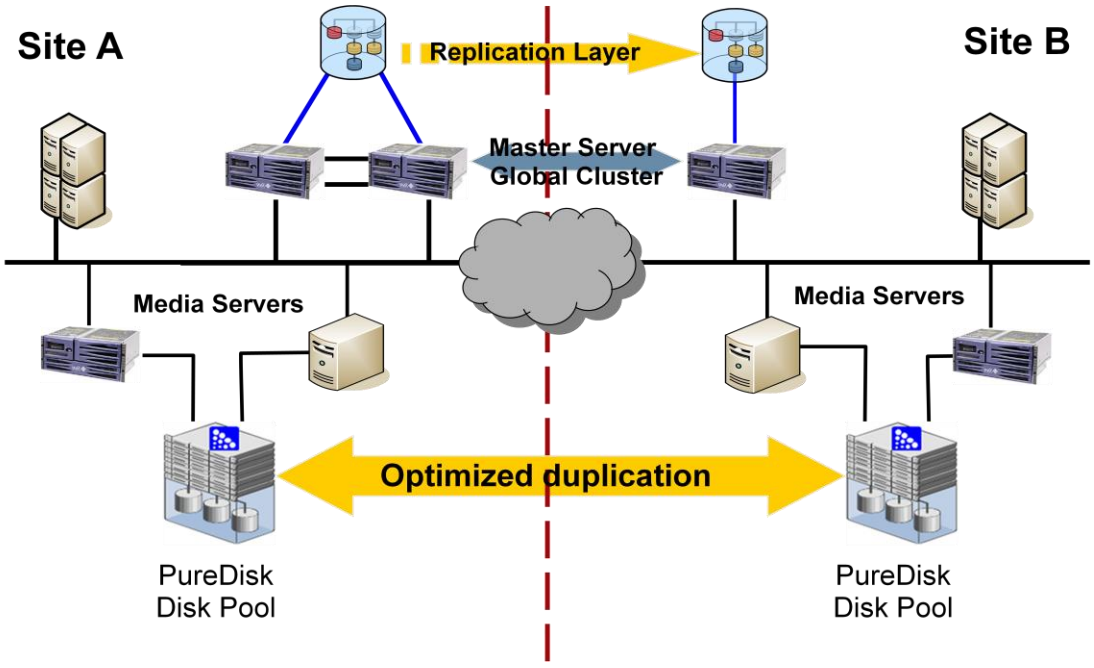

Figure 7 - Dual Site/Single Domain with PureDisk

Under normal operation the master server runs on one of the cluster nodes on Site A. In the event of a site failure at Site A, the master server cluster is failed over to Site B. As the backups of the critical applications at site A have been duplicated to site B, recovery of those applications at site B can begin immediately. A full off-site disaster recovery of critical applications is achieved with no requirement to move any physical media between sites or to recover any NetBackup component before starting the application recovery.

About Symantec

Symantec is a global leader in infrastructure software, enabling businesses and consumers to have confidence in a connected world. The company helps customers protect their infrastructure, information, and interactions by delivering software and services that address risks to security, availability, compliance, and performance. Headquartered in Cupertino, Calif., Symantec has operations in 40 countries. More information is available at www.symantec.com.

For specific country offices and contact numbers, please visit our Web site. For product information in the U.S., call toll-free 1 (800) 745 6054.

Symantec Corporation
World Headquarters
20330 Stevens Creek Boulevard
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934
www.symantec.com

10/07 xxxxxxxx