

EMC Data Domain Boost for Symantec NetBackup OpenStorage

Best Practices Planning

Abstract

Conceptually understanding EMC® Data Domain® Boost for Symantec NetBackup provides a clear view of the business value and technical merits of integrating these two solutions in an IT infrastructure. This white paper moves past the conceptual stage to solution planning and deployment. Best practice guidelines are presented with the goal of eliminating implementation challenges. Knowledge and experience gained from a large number of customer deployments are logically presented for the overall benefit of those using a DD Boost solution with Symantec NetBackup OpenStorage.

July 2010

Copyright © 2010 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

Part Number h7296

Table of Contents

Executive summary	4
Introduction	4
Audience	5
DD Boost for Symantec NetBackup OpenStorage background	5
Terminology and naming conventions	6
Data Domain hostname	6
Storage server	6
Logical Storage Unit	6
Disk pool	7
Network planning	8
Example network topologies	8
Recommended network configurations	11
DD Boost advanced load balancing and link failover	13
DD Boost distributed segment processing	16
Deployment options	18
Optimized duplication	18
Storage units and storage server access	18
Network considerations	20
Throttling optimized duplication	20
DD Replicator low-bandwidth optimization	20
Optimized duplication failures	20
Seeding remote Data Domain systems	21
Data Domain total job streams	22
Duplication job configuration options	23
Duplication to tape	24
Hierarchical duplication	25
Multiple NetBackup domain configurations	26
Catalog backup and recovery	28
NetBackup versions 6.5 through 6.5.3	28
NetBackup version 6.5.4	28
NetBackup version 7.0	29
Conclusion	29
References	29

Executive summary

EMC® Data Domain® deduplication storage with Data Domain Boost (DD Boost) for Symantec NetBackup OpenStorage software is straightforward to install and use in most environments; however, deployments involving multiple sites and a complex environment may experience issues with naming conventions or network infrastructure. Therefore, all implementations should be well planned and documented so that they can be deployed more quickly with fewer challenges when compared to ad-hoc techniques.

Deployment is often followed by a series of trials or a period of testing intended to prove that the solution functions as planned. In this white paper, OpenStorage best practices are examined and discussed to assist in eliminating the bottlenecks associated with deployment and functional testing of the solution.

Introduction

This paper focuses on backups and the creation of duplicate backup copies with OpenStorage. Network configurations, optimized duplication, centralized tape operations, and recovery are examined.

Recommended best practices, as well as alternate strategies, are covered with the goal of enhancing OpenStorage solution planning and deployment.

Table 1 shows a summary of best practice recommendations for those already familiar with Data Domain systems, NetBackup, and OpenStorage. Detailed discussion, explanation, and reasoning for these recommendations are covered in subsequent sections.

Table 1. Recommended options and settings

Option or setting	Recommendation
Name resolution	Use host files to convert hostnames to IP addresses.
Storage server	Use a hostname when configuring a storage server.
	Configure one storage server per Data Domain system.
	If a Data Domain system is shared with multiple NetBackup domains, create one storage server per NetBackup domain.
Logical Storage Unit (LSU)	Create one LSU per Data Domain system.
	If a Data Domain system is shared with multiple NetBackup domains, create one LSU on the Data Domain system per NetBackup domain. <i>Note: LSUs must not be shared across multiple NetBackup domains.</i>
	An LSU should be named with 50 or fewer ASCII characters including "-" (minus sign) and "_" (underscore).
Disk pool	Create one disk pool per LSU.
	Limit I/O streams per volume based on the "Total Data Streams" value as derived from the appropriate <i>EMC Data Domain Operating System (DD OS) Administration Guide</i> .
Media server load balancing	Architect load balanced OpenStorage storage units such that they utilize only NetBackup media server platforms with the same operating system and server platform performance attributes.
Backup network topology	Use a dedicated backup network between the NetBackup media server(s) and the target Data Domain system(s) when possible.
	Enable and use an OpenStorage interface group employing multiple network connections between NetBackup media servers and Data Domain systems when possible.
	Use a virtual interface configured as a failover pair on the Data Domain system for the network connection registered with NetBackup.

Option or setting	Recommendation
	Configure interface groups to contain only interfaces with the same characteristics.
	Use IP addresses within a single subnet when configuring an OST interface group.
Distributed segment processing	Enable the DD Boost distributed segment processing feature.
Storage unit	Limit maximum concurrent jobs for backup and inbound optimized duplication jobs after considering required streams for outbound optimized duplication jobs and duplication jobs to tape.
	If multiple storage units use the same target disk pool, limit the combined maximum concurrent jobs to the value recommended in the <i>EMC Data Domain Operating System (DD OS) Administration Guide</i> .
Optimized duplication	Segregate optimized duplication traffic onto a dedicated replication network when possible.
	If using NetBackup version 6.5.4 or later, use the default NetBackup “RESUME_ORIG_DUP_ON_OPT_DUP_FAIL = FALSE” setting such that failed optimized duplication jobs are not transferred fully inflated through NetBackup media servers and over WAN connections.
	Use the low bandwidth optimization mode on Data Domain systems in cases where WAN bandwidth is less than 6 Mb/s.
	Use NetBackup storage lifecycle policies to control optimized duplication.
	Use fixed retention periods.
Hierarchical duplication to tape	Source new duplication copies from the geographically closest prior backup image copy.
	Specify the use of a NetBackup alternate read server geographically close to the source and destination storage units.
Multiple NetBackup domain configurations	Use naming conventions that assist in simplifying the shared use of Data Domain systems in multiple NetBackup domains.
Catalog backup and recovery	Test and document the appropriate catalog backup and recovery solution based on the NetBackup version in use and site-specific requirements.

Audience

System administrators and vendor staff associated with and performing OpenStorage deployments are encouraged to use this paper to take advantage of substantial real-world knowledge gained by EMC from assisting other customers.

DD Boost for Symantec NetBackup OpenStorage background

OpenStorage software provides API-based integration between Data Domain storage systems and NetBackup. The API gives NetBackup visibility into the properties and capabilities of the Data Domain storage system, control of the backup images stored in the system, and wide area network (WAN) efficient replication to remote Data Domain storage systems.

DD Boost for Symantec NetBackup OpenStorage enhances the integration between NetBackup and Data Domain systems. It distributes part of the deduplication process to the media server, improving backup throughput up to 50 percent, reducing media server loads 80 percent to 90 percent, and decreasing LAN bandwidth requirements 20 percent to 40 percent. It also enables advanced load balancing and failover at the Ethernet link layer.

DD Boost for NetBackup has two components. The *DD Boost Library* is embedded in the OpenStorage plug-in that runs on NetBackup media servers. The *DD Boost Server* is built into DD OS 4.8 or later and runs on a Data Domain system. The two components integrate seamlessly across the IT infrastructure to enable these benefits. Supported with NetBackup 6.5 and later, OpenStorage-enabled Data Domain deduplication storage systems and the Symantec NetBackup OpenStorage Disk Option provide the following key enhancements for disk-based data protection strategies:

- NetBackup optimized duplication – Backup image duplication using EMC Data Domain Replicator, providing network-efficient replication and able to be controlled, monitored, and cataloged by NetBackup.
- Integrated NetBackup reporting of Data Domain replication job status.
- Recovery of replicated backup images in their entirety or at a granular level via the NetBackup user interface.
- Sharing of OpenStorage storage units among heterogeneous NetBackup media servers.
- NetBackup media server load balancing, eliminating the need to manually divide client backups across NetBackup media servers utilizing OpenStorage storage units.
- Tape consolidation – Backup images from remote locations and branch offices can be replicated to a centralized location where they can be duplicated to tape under the control of NetBackup.

Terminology and naming conventions

Nomenclature, the assigning of names to OpenStorage-specific components, is an important consideration. Best practice is to use a naming convention that will be easily understood by the user, system engineer, and potentially any support personnel involved with the OpenStorage solution. Key recommendations for this nomenclature follow.

Data Domain hostname

The Data Domain hostname identifies a system. Hostnames are converted to IP addresses via host files or a name service such as DNS. Hostnames are not specifically NetBackup objects, but when used as storage server names, they tell NetBackup media servers at the TCP/IP level how to connect to the storage server.

- Use the assigned fully-qualified hostname.
- Use host files to convert hostnames to IP addresses. While generally reliable, naming services such as DNS may unintentionally introduce processing delays or error conditions into the data protection infrastructure.
- When possible, avoid creating secondary hostnames to associate with alternate IP interfaces.

Storage server

The storage server is a logical object defined within NetBackup that “points” to a Data Domain system. NetBackup communicates with the storage server and uses “tpconfig” credentials to request use of a disk pool for backup and restore operations.

- Have only one storage server defined within a NetBackup domain per Data Domain system.
- Use hostnames when configuring storage servers. Do not use IP addresses in place of hostnames as this will limit the options to route optimized duplication traffic.
- Keep the name of the storage server unique across the enterprise.

Logical Storage Unit

A Logical Storage Unit (LSU) is a disk target within a storage server.

- Create one LSU per Data Domain system when deployed in a single NetBackup domain. This includes scenarios where the LSU will be used as both the source and destination of different optimized duplication jobs within a single NetBackup domain. Create additional LSUs as applicable

for any additional NetBackup domains. Never create more than 100 LSUs on any single Data Domain system.

- Name LSUs with 50 or fewer ASCII characters including the “-“ (minus sign) and “_” (underscore) characters.
- When possible name the object with a “-lsu” extension to allow for easy identification as an LSU object.

Multiple LSUs can be configured on a single Data Domain system. This can create conflicts for advanced NetBackup features such as media server load balancing, capacity reporting, and others, and is only a recommended best practice when the Data Domain system is being configured in multiple NetBackup domain installations.

Disk pool

A disk pool is a NetBackup object that correlates to an LSU and storage server combination (Figure 1).

- Have only one disk pool per LSU. When multiple storage servers have been defined relative to a single Data Domain system, create one disk pool per storage server.
- Name the object with a “-dp” extension for easy identification as a disk pool object.
- Limit I/O streams per volume when configuring disk pools on Data Domain systems.

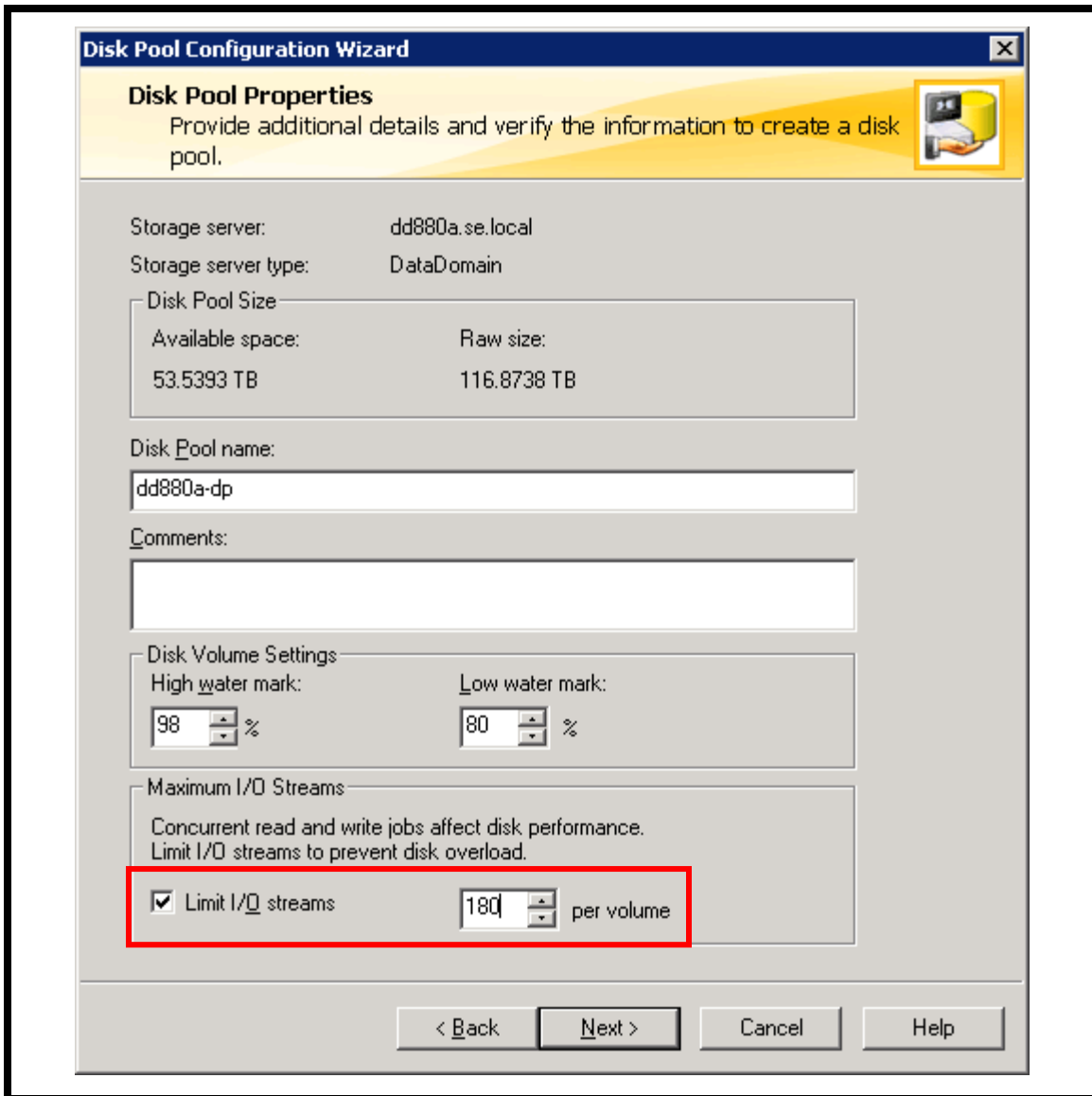


Figure 1. Disk Pool Configuration Wizard

In the figure, the disk pool is named with a “-dp” suffix. The **Limit I/O streams** checkbox is selected. A value of **180** streams per volume is supplied as derived from the appropriate *EMC Data Domain Operating System (DD OS) Administration Guide*.

For additional information about setting the limit I/O streams per volume, see the “Data Domain total job streams” section on page 22.

Network planning

There are varying degrees of network complexity associated with a given OpenStorage deployment. At a minimum, a single Data Domain system configured as a storage server is network-connected to a NetBackup media server. NetBackup-optimized duplication and media server load balancing add additional requirements.

Example network topologies

This section discusses examples of network topologies such as the following:

- NetBackup media server and Data Domain systems sharing a common LAN configured for optimized duplication (Figure 2)

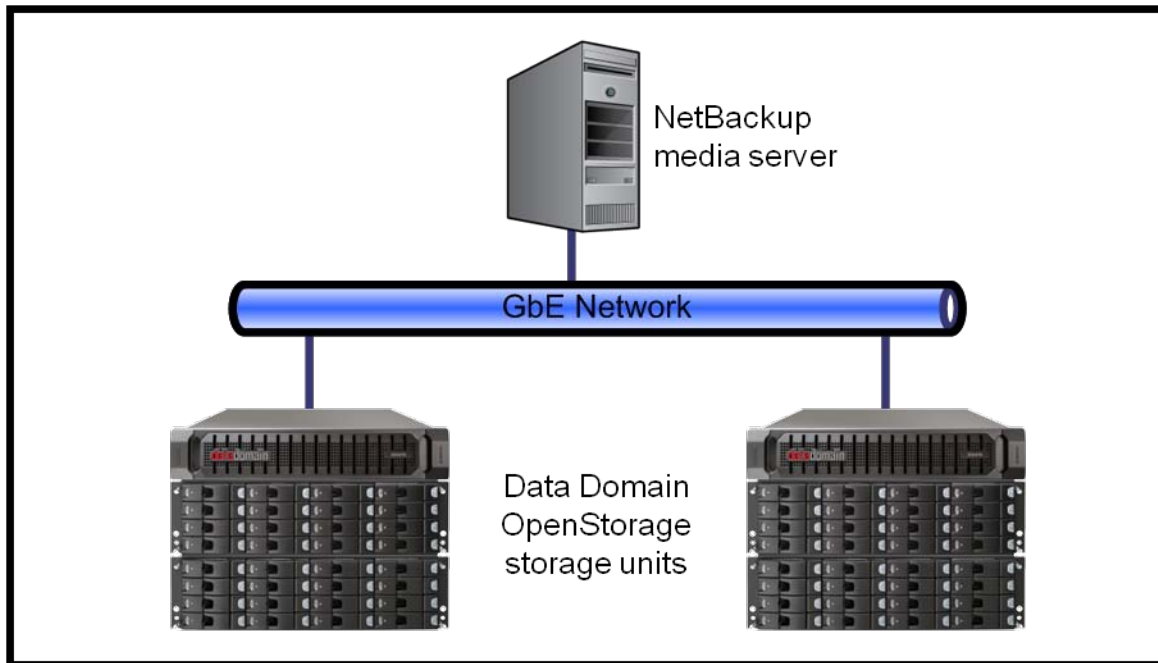


Figure 2. Optimized duplication with a common network

The figure shows an example of a NetBackup master/media server network connected to two Data Domain systems. In this use case, both backup and optimized duplication traffic use the same network interface card (NIC) on a given Data Domain system.

- NetBackup media servers and Data Domain systems configured to use separate backup and optimized duplication networks (Figure 3)

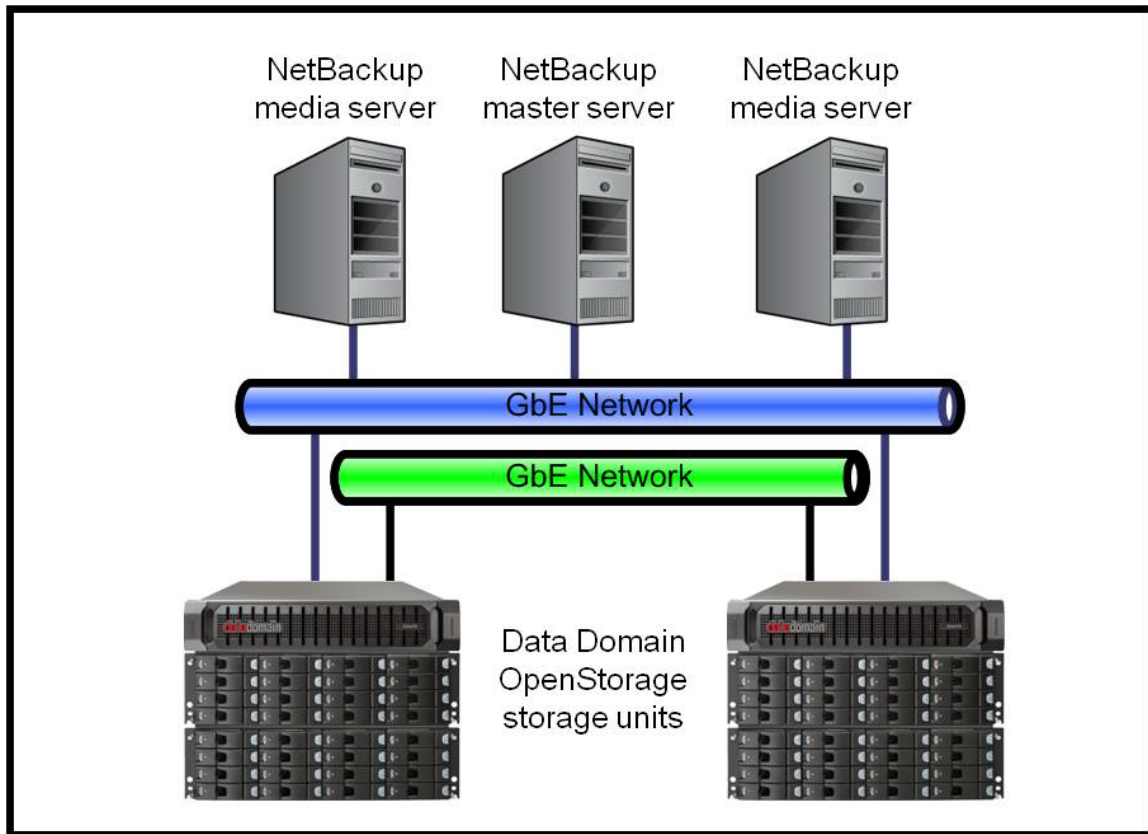


Figure 3. Optimized duplication with a separate network

The figure shows three NetBackup media servers and two Data Domain systems. Each NetBackup media server has a LAN connection (blue) to both Data Domain systems. A WAN (green) also connects the Data Domain systems. Backup traffic uses the LAN connection between a given NetBackup media server and a Data Domain system. Optimized duplication traffic uses a separate NIC on each Data Domain system. This configuration may be preferred in cases where backup and restore data transfer rates require a 10 GbE network, with a lower bandwidth network able to accommodate optimized duplication traffic.

- NetBackup media server load balancing with a Data Domain system (Figure 4)

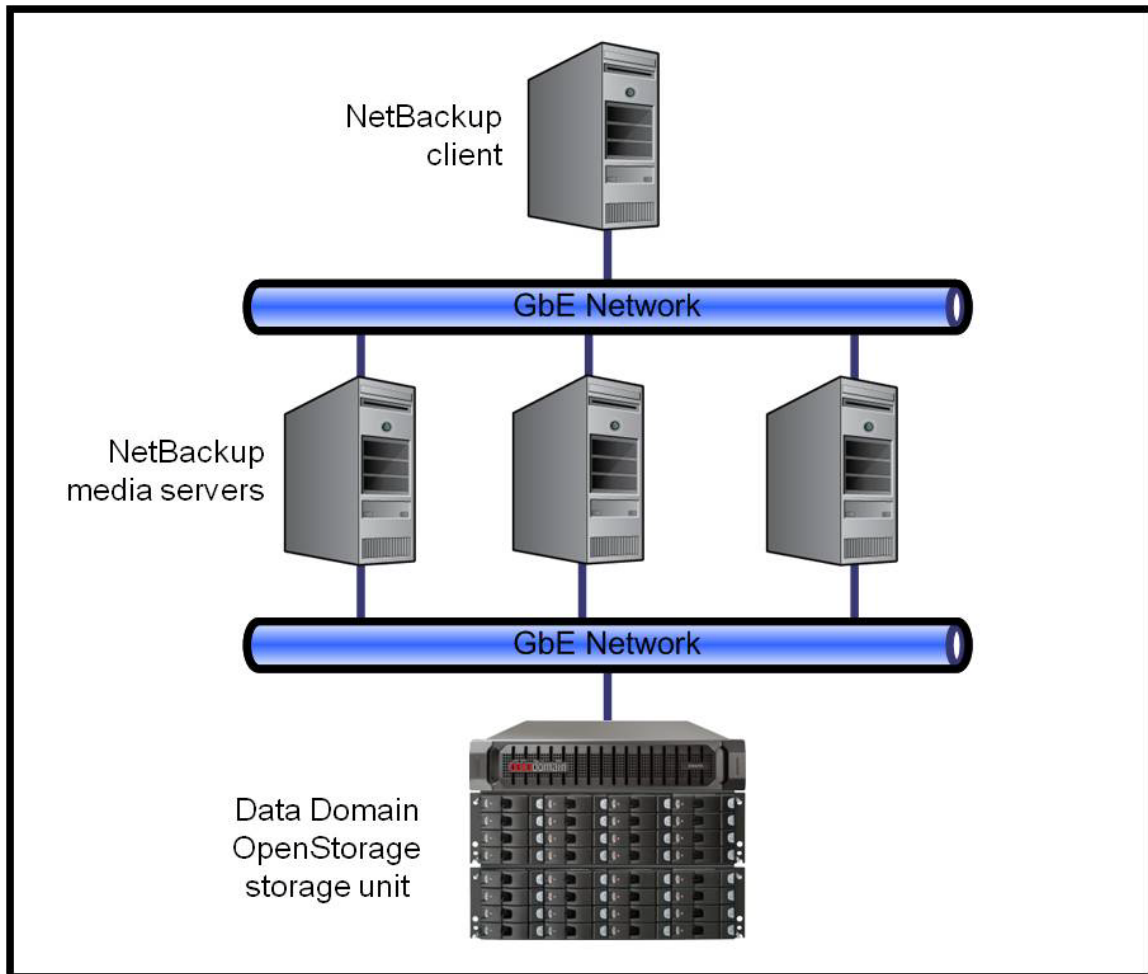


Figure 4. NetBackup media server load balancing

A NetBackup client can be backed up through a number of different NetBackup media servers. The Data Domain OpenStorage storage unit has been configured so that each NetBackup media server can access its resources. This enables NetBackup media server load balancing, where the least loaded media server is used to fulfill a backup request. Additionally, this configuration allows NetBackup to bypass an offline media server when fulfilling a backup or restore request.

Use media server load balancing and NetBackup media servers with similar characteristics within a given storage unit. Mixing NetBackup media servers with dissimilar hardware and operating systems within a storage unit may cause a given backup job to perform inconsistently depending on which resource is selected by NetBackup at runtime. Consistent service levels are more likely to be obtained when all NetBackup media server resources have similar attributes.

Recommended network configurations

The best practice recommendations throughout this paper are based on known reference deployments that exhibit desirable behavior and performance characteristics. Simplicity is preferred over complexity. Ease of deployment, simplified administration, and predictable results have yielded the general recommendations that follow.

- Interconnect NetBackup media servers and Data Domain systems using a dedicated backup area network. NetBackup clients and NetBackup media servers should ideally be interconnected through a separate front-end network (Figure 5).

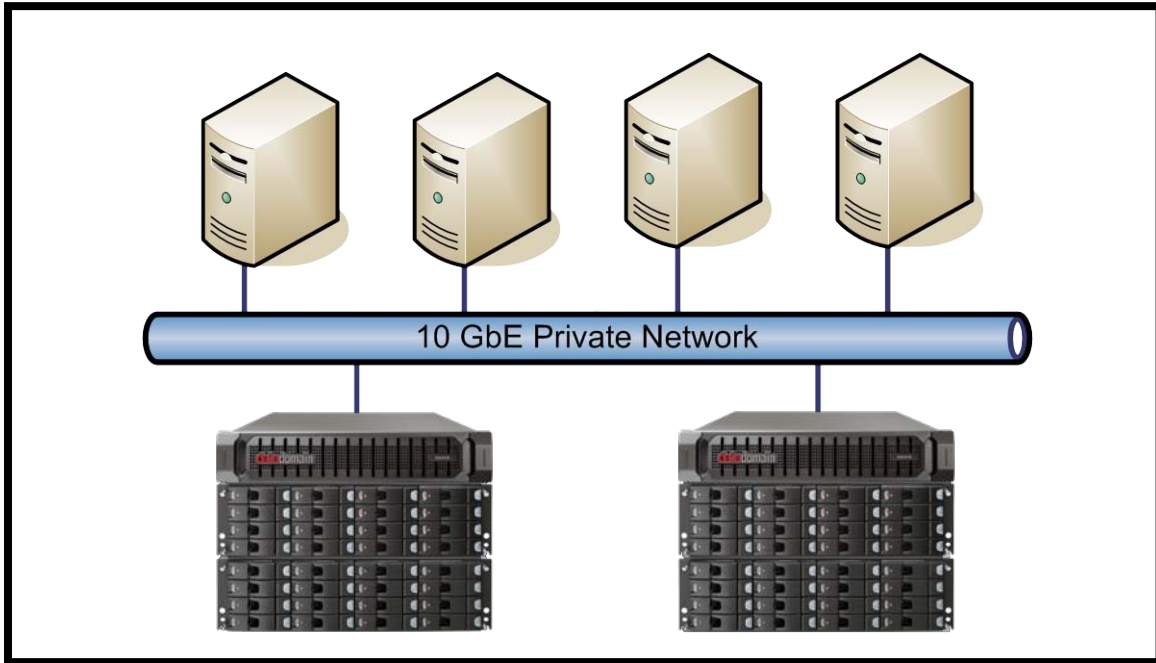


Figure 5. Dedicated backup network

This figure shows a dedicated backup network, which provides a number of tangible benefits:

- By segregating NetBackup media server and storage unit traffic from other network traffic, potential contention issues are limited to backup and recovery jobs. Known available bandwidth can be managed to achieve aggressive data protection and recovery service levels.
- A scalable infrastructure has been established in case data protection network bandwidth requirements change over time.

While not always possible based on customer requirements and pre-existing NetBackup media server and network infrastructure deployments, the use of a dedicated backup network is preferred when compared to mixed-use network configurations.

DD Boost advanced load balancing and link failover

Supported with DD OS version 4.8 or later when combined with OpenStorage plug-in version 2.0 or later, the DD Boost Library enables scalable link aggregation at the application layer to create a single interface group. With this DD Boost feature, data transfers between the OpenStorage plug-in and Data Domain system are automatically load balanced across the network connections in the interface group. If a network connection in the group becomes unavailable, subsequent data transfers utilize accessible remaining network connections as long as the unavailable connection is not the one registered within NetBackup. For additional information, please refer to the *EMC Data Domain OpenStorage (OST) Administration Guide*.

- Use interface groups to increase available network bandwidth between NetBackup media servers and Data Domain systems. Previously published recommendations detailing the use of multiple network connections where each NetBackup media server converted the hostname of a Data Domain system to a different IP address are no longer valid.

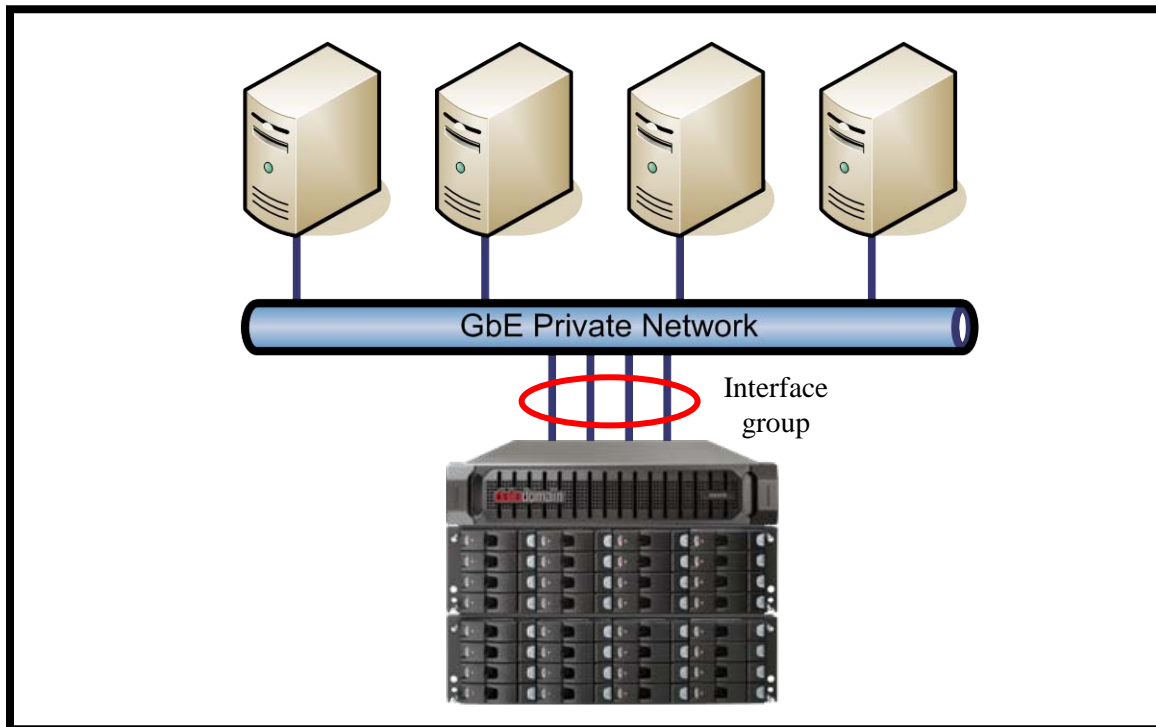


Figure 6. Recommended use of multiple GbE networks

In Figure 6, each of four NetBackup media servers connects to a Data Domain system configured as a single storage server. The Data Domain system has been configured with an interface group that consists of four network interfaces. Advanced load balancing and failover will automatically balance data transfers between the Data Domain system and NetBackup media servers across all interfaces in the interface group. If an interface fails, the backups will be automatically routed to other available interfaces in the interface group.

- Configure interface groups such that the storage server name registered within NetBackup corresponds to a virtual interface on the Data Domain system configured as a failover pair.

The Data Domain system interface corresponding to the storage server name as configured within NetBackup when executing the “nbdevconfig –createstds” command may be referred to as the registered interface. If the registered Data Domain system interface enters a non-operative state, all backup and restore jobs between the NetBackup media servers and the Data Domain system will fail. To mitigate the impact of a single interface failure, configure the registered Data Domain system interface as a failover pair (Figure 7).

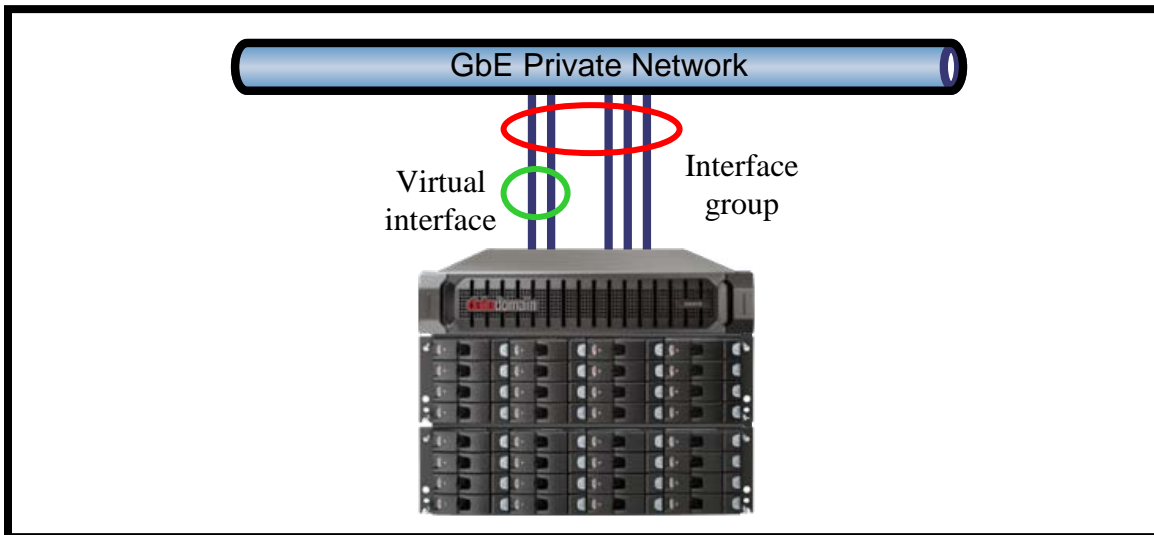


Figure 7. Interface group with a virtual interface configured as a failover pair

A Data Domain system is configured with an interface group where one of the network connections is a virtual interface configured as a failover pair. The virtual interface corresponds to the network connection registered within NetBackup.

- Configure interface groups to contain only interfaces with the same characteristics. Load balancing occurs across all available interfaces in an interface group and is based on the number of jobs currently executing when a new job starts. Adding interfaces with different performance characteristics into an interface group can lead to inconsistent service levels. For instance, an interface group containing both 1 GbE and 10 GbE interfaces may result in backup or restore job data transfer rate differences depending on which interface is selected for use. Note that the registered Data Domain system interface does not need to be added to the interface group. This enables the use of two 1 GbE interfaces configured as a failover pair as the registered interface, while using two 10 GbE interfaces within an interface group.

- Use IP addresses within a single subnet when configuring an interface group (Figure 8).

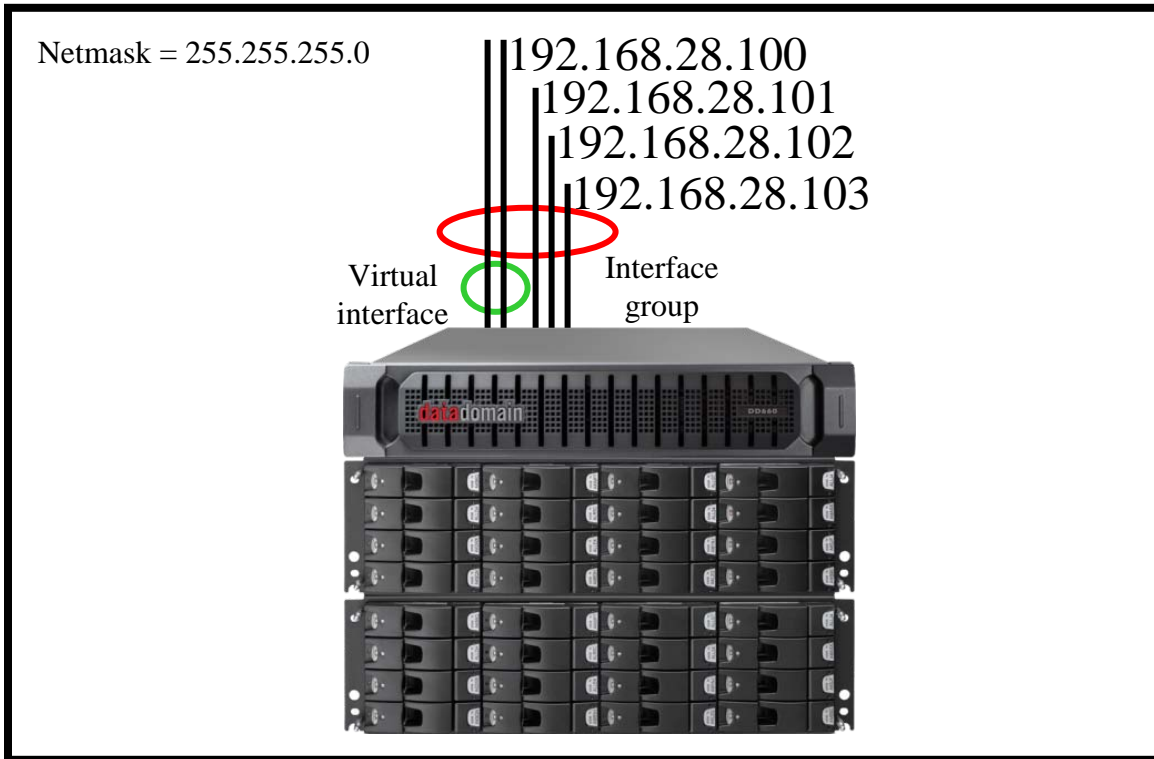


Figure 8. Interface group configured with IP addresses within a single subnet

In Figure 8, a Data Domain system is configured with an interface group consisting of four network addresses on the same subnet. This configuration allows load balancing and failover to occur between the Data Domain system and a network IP switch.

An alternate, but *not recommended*, configuration using multiple subnets within an interface group can be used to increase the aggregate network bandwidth available between a NetBackup media server and a Data Domain system (Figure 9). This requires configuring network interfaces on both the NetBackup media server and Data Domain system for each subnet that will be utilized. In this configuration, lab-based failure injection on any network connection between the NetBackup media server and IP switch resulted in backup job “network connection broken” failures across all other networks, including the network where the failure was induced. Users considering the viability of this configuration are encouraged to investigate the impact of a network failure between the NetBackup media server and IP switch prior to production deployment.

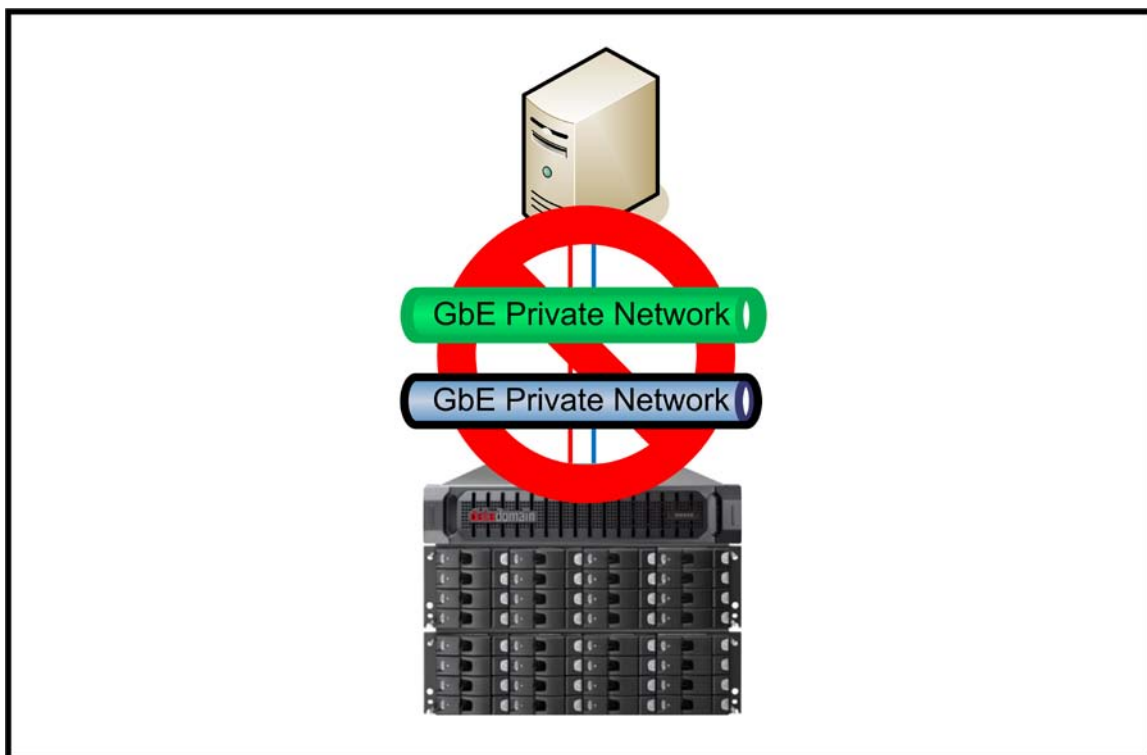


Figure 9. Not recommended: Interface group with multiple subnets

In this graphic, a Data Domain system is configured with an interface group consisting of two network addresses on different subnets. This configuration allows load balancing to occur between the Data Domain system and NetBackup media server. Failover functionality between the NetBackup media server and either switch falls outside the scope of DD Boost advanced load balancing and failover.

Consider one additional recommendation:

- Whenever possible, use a network connection that includes failover capabilities.

DD Boost distributed segment processing

DD Boost distributed segment processing allocates part of the deduplication process to NetBackup media servers, usually increasing the aggregate backup data transfer rate, and decreasing both NetBackup media server and Data Domain CPU loads. Additionally, LAN traffic is reduced between NetBackup media servers and the Data Domain system, as only unique data is transferred over the backup network. For additional information, please refer to the *EMC Data Domain OpenStorage (OST) Administration Guide*.

DD Boost distributed segment processing is enabled by default on Data Domain systems where the initial DD OS install occurred with version 4.8 or later. Data Domain systems upgraded to DD OS 4.8 from prior DD OS versions will not have the distributed segment processing option enabled by default.

- Use distributed segment processing in NetBackup OpenStorage environments. Enable this option by using the command line interface “ost option set boost enabled” command (Figure 10) or from the EMC Data Domain Enterprise Manager GUI.

```
dd690a.se.local - PuTTY
sysadmin@dd690a# sysadmin@dd690a# ost option show
Option      Value
-----
boost       disabled
-----
sysadmin@dd690a# ost option set boost enabled
OST option "boost" set to enabled.
sysadmin@dd690a# ost option show
Option      Value
-----
boost       enabled
-----
sysadmin@dd690a# sysadmin@dd690a#
```

Figure 10. Enabling the Boost option

DD Boost option status is queried with the “ost option show” command. The results of the command indicate that the Boost option was disabled. The subsequent command, “ost option set boost enabled”, is used to enable the DD Boost option.

Specifically *not recommended* are configurations where a NetBackup media server and Data Domain system are connected via a wide area network (Figure 11).

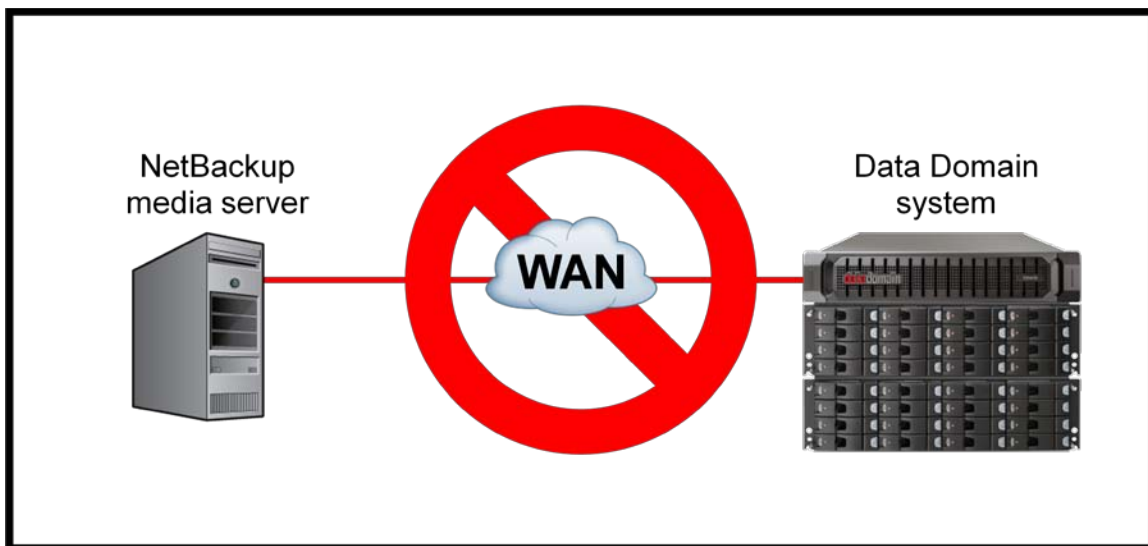


Figure 11. Non-recommended WAN configuration

Note:

- This configuration has not been tested by EMC or Symantec.
- The NetBackup NBRMMS service (NetBackup Remote Manager and Monitor Service) checks disk pool status frequently. High-latency WAN connections may cause NBRMMS to indicate that the disk pool is “down” or inoperative.

Deployment options

Optimized duplication

NetBackup optimized duplication replicates deduplicated data between source and destination systems, and typically requires only a fraction of the network bandwidth consumed by backup or recovery jobs. In addition, DD Boost-managed file replication enables NetBackup to manage Data Domain network-efficient replication through optimized duplication controls. Simple in principle, optimized duplication is also simple to configure once requirements are understood.

Storage units and storage server access

The first items to consider are source and destination storage units (for example, Data Domain systems). If both connect to the same NetBackup media server, that server needs credentials to access both the source and destination Data Domain systems. Alternatively, when the source and destination storage units reside on different NetBackup media servers, the server initiating the optimized duplication job requires credentials to access both the source and destination Data Domain systems. Thus even though a particular NetBackup media server may never directly back up to or recover from a particular Data Domain system, when initiating optimized duplication, it still needs access credentials for the Data Domain system acting as optimized duplication destination storage.

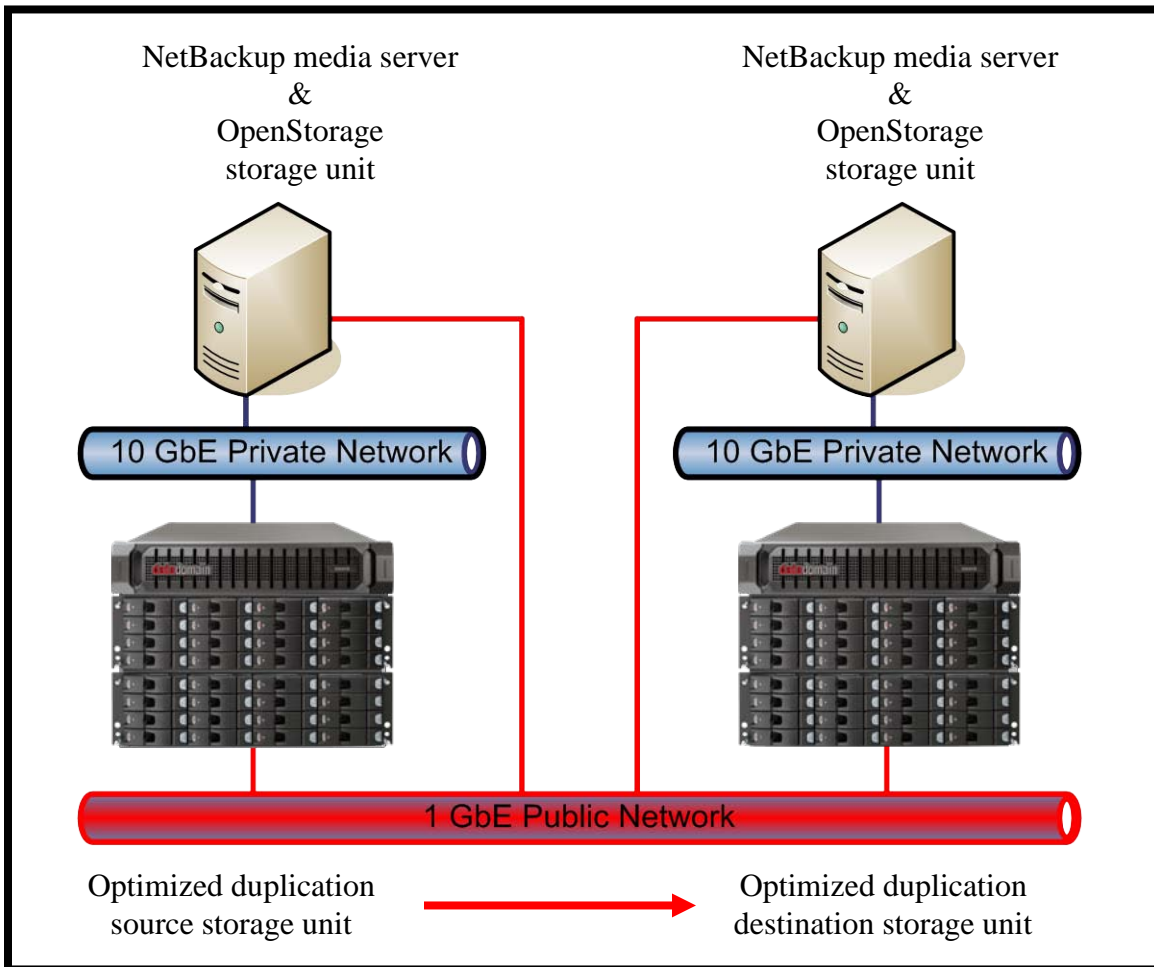


Figure 12. Separate source and destination NetBackup media servers

Figure 12 shows optimized duplication between two OpenStorage storage units. The NetBackup media server initiating an optimized duplication job needs to have credentials to access both the source and destination OpenStorage storage units.

Credentials can be set differently depending on which version of NetBackup is being used. Versions prior to NetBackup 7.0 require executing the “tpconfig” command on each NetBackup media server needing access to a given OpenStorage storage unit. NetBackup 7.0 and later versions also provide a GUI-driven storage server creation wizard that allows credentials to be specified. Regardless of how credentials are configured, they allow the NetBackup media server to use the OpenStorage storage unit for backup and recovery jobs, as well as for optimized duplication.

In cases where optimized duplication uses a destination OpenStorage storage unit that is geographically distant from the NetBackup media server initiating optimized duplication, the storage unit definition should not allow the geographically distant NetBackup media server to use the storage unit for backup jobs. This is easily accomplished from within the NetBackup Change Storage Unit dialog box as shown in Figure 13.

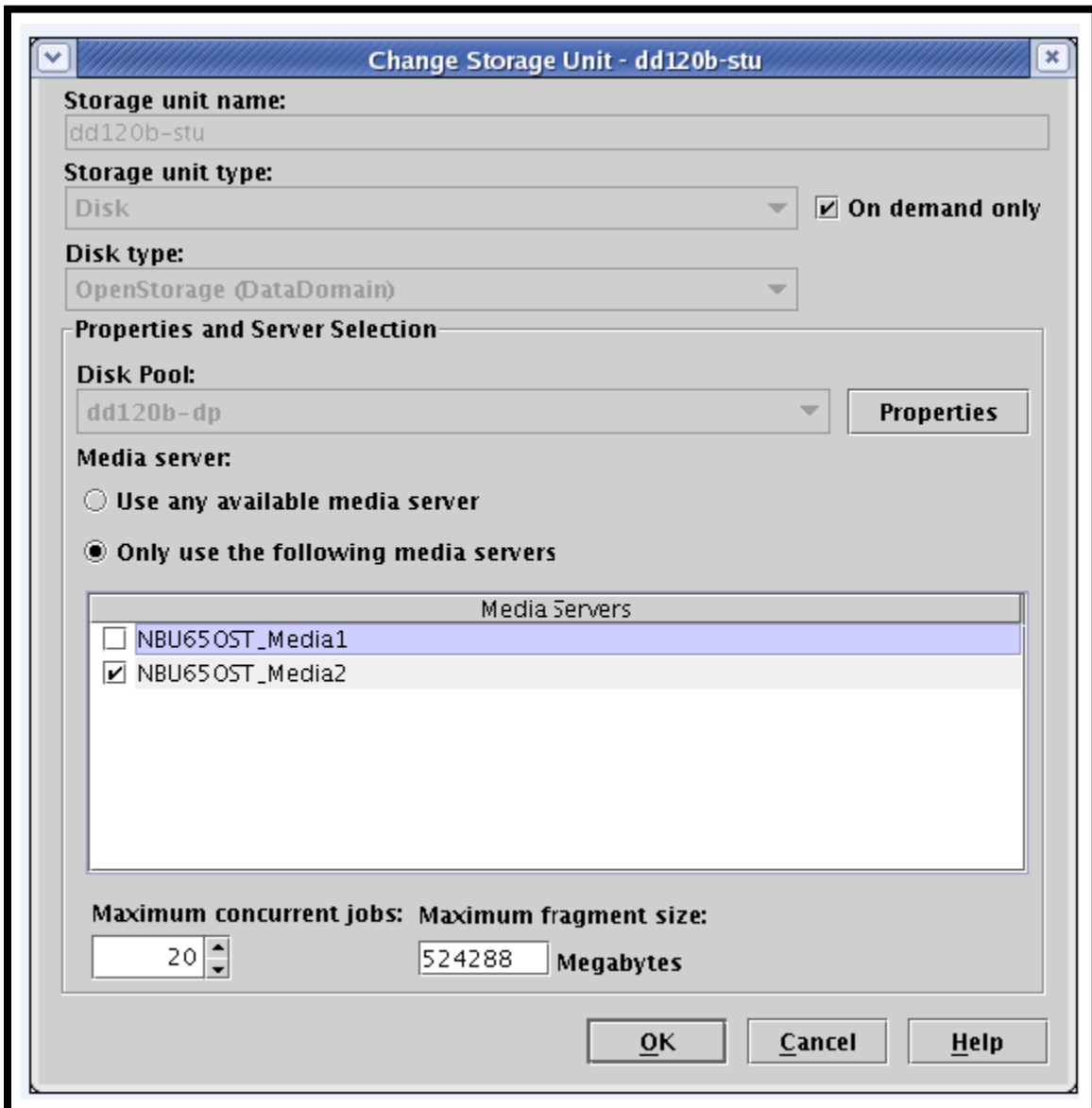


Figure 13. Change Storage Unit dialog box – use only specific media servers

In Figure 13 the NetBackup storage unit named “dd120b-stu” is configured to allow only the NetBackup media server named “NBU65OST_Media2” to use it for backup jobs. The NetBackup media server named “NBU65OST_Media1” has credentials to access the storage unit for the purpose of initiating optimized duplication jobs.

Network considerations

Replicating backup images under the control of NetBackup optimized duplication includes the ability to use the same network that is used for backup and restore operations or to use a different network. In cases where optimized duplication traffic flows between geographically different locations, some customers have chosen to use a separate dedicated network connection for optimized duplication. This connection links source and destination Data Domain systems specifically for the purpose of replication controlled by NetBackup-initiated optimized duplication. Users needing to track WAN link usage may also prefer this approach.

The network used for optimized duplication is based on network name resolution on the source Data Domain system. The destination Data Domain system is known to the source Data Domain system based on the IP address supplied by DNS, or by a local hosts file entry. Populating the source Data Domain systems hosts file with the desired IP address of the destination Data Domain system is all that is required to use a specific NIC and network. If this value is not present, NetBackup will perform optimized duplication using the same network it uses to access the source and destination Data Domain systems for backup and restore jobs.

Throttling optimized duplication

Throttling of optimized duplication is controlled at a global or system level on each Data Domain system. Limits on network bandwidth used by the replication process can be based on various criteria such as a scheduled or temporary rate.

Throttling network bandwidth consumption may elongate optimized duplication job runtimes and impact recovery point objectives for disaster recovery. Other effects might include the queuing of backup jobs if optimized duplication jobs push the storage unit beyond its maximum concurrent jobs parameter. Beyond throttling and based on service level requirements, it might be possible to reduce network bandwidth required by limiting the amount of data that needs to be replicated. One possible method is to perform optimized duplication of only full backups, where incremental backups are not duplicated. Different storage lifecycle policies can be employed for full and incremental backups should this methodology align with service level objectives.

DD Replicator low-bandwidth optimization

Supported with DD OS version 4.8 and later, DD Replicator low-bandwidth optimization provides increased efficiency when performing optimized duplication across WAN links of less than 6 Mb/s. Enabling this DD Replicator feature requires no changes on NetBackup media servers as NetBackup is unaware of this Data Domain system configuration option.

Optimized duplication failures

In the event that an optimized duplication job fails, the resulting action taken by NetBackup is dependent on the version of NetBackup being used. The *EMC Data Domain OpenStorage (OST) Administration Guide* provides deployment requirements and additional information.

- Prior to NetBackup version 6.5.4, if an optimized duplication job fails, duplication job retry will attempt to use conventional (not optimized) duplication. This equates to sending the fully hydrated backup image from the source OpenStorage storage unit through one (or possibly two) NetBackup media server(s) to the destination OpenStorage storage unit. Where this NetBackup behavior would negatively impact production data protection operations, users should consider running NetBackup 6.5.4 or later.
- NetBackup version 6.5.4 and later, by default, will not retry a failed optimized duplication job unless it was initiated by means of a storage lifecycle policy. When retried via a storage lifecycle policy, the job

will attempt to use optimized duplication only. While not generally recommended, this default behavior can be overridden by configuring NetBackup to retry with conventional duplication. Adding “RESUME_ORIG_DUP_ON_OPT_DUP_FAIL = TRUE” to the NetBackup “bp.conf” file will modify default product behavior. The *Symantec Veritas NetBackup Shared Storage Guide UNIX, Windows, Linux Release 6.5.4* (or later version) document provides detailed information.

Seeding remote Data Domain systems

New deployments looking to utilize optimized duplication may have network bandwidth limitations between sites that could cause the first week of duplication jobs to run for extended time periods. One solution is to seed the remote Data Domain system locally, and then after a week or so relocate it to the intended remote site.

The challenge this presents is that while the remote system is in transit, optimized duplication failures may occur. A solution is to adjust the appropriate NetBackup storage unit so that no failures occur. By setting the storage unit **Maximum concurrent jobs** parameter to a value equal to zero, optimized duplication jobs will enter a queued state instead of failing, as shown in Figure 14.

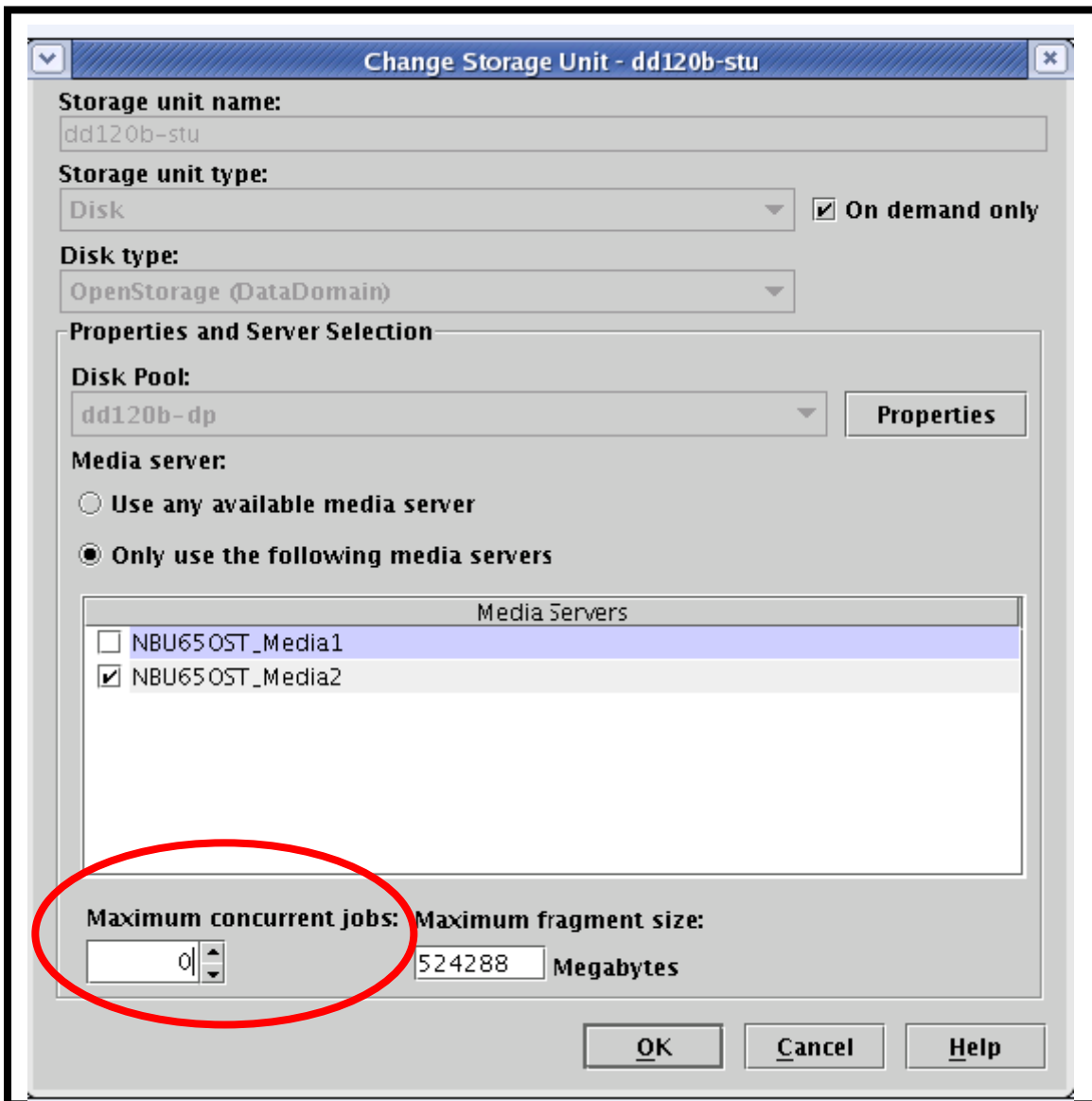


Figure 14. Change Storage Unit dialog box – maximum concurrent jobs

The storage unit **Maximum concurrent jobs** parameter is set to a value of **0**. Use this technique when relocating a Data Domain system from a local site to a final destination site so that related optimized duplication jobs will enter a queued state instead of failing.

Data Domain total job streams

Data Domain systems have model-, NVRAM-, memory-, and DD OS version-dependent recommendations regarding maximum write, read, and replication stream counts. Adhering to these recommendations during backup, optimized duplication, and duplication to tape assists in achieving predictable performance.

Stream count information can be found in the *EMC Data Domain Operating System (DD OS) Administration Guide*. A stream count recommendation example is shown in Table 2.

Table 2. Recommended stream counts examples

Platform	RAM	Total data streams	Maximum write streams	Maximum write streams with optimized duplication	Maximum read streams	Mixed read and write streams
DD880	64 GB	180	180	90	50	< = 180 Writes and < = 50 Reads

The facilities provided by NetBackup to limit total jobs streams consist of two configuration options:

- Limit I/O streams per volume
The first configuration option, “limit I/O streams per volume,” is set within a given disk pool. Data Domain systems configured according to the recommendations in this paper will present a single LSU representing the entire file system. As such, there is a single volume on which to limit I/O streams. This parameter can be used to limit the total number of read and write streams independent of the storage unit **Maximum concurrent jobs** parameter.
- Maximum concurrent jobs
The second configuration option, “maximum concurrent jobs,” is set within a given storage unit. NetBackup monitors the number of jobs running on a storage unit in order to enforce the maximum number of current jobs value. Monitored jobs include backup and inbound optimized duplication jobs. (Note that NetBackup does not distinguish between backup and inbound optimized duplication jobs. There is no granular ability within NetBackup to limit backup and inbound duplication jobs separately. Outbound optimized duplication jobs are not counted or controlled by the storage unit **Maximum concurrent jobs** parameter. Neither are reads that occur as part of a duplication to tape.)

Table 3 shows the two example use cases based on the data presented in Table 2:

Table 3. Example use cases

Use case	Description	Disk pool setting	Storage unit setting
1	Data Domain system used for backups and as the source for optimized duplication jobs	180	90
2	Data Domain system used as the destination for optimized duplication jobs and as the source for duplication jobs to tape	180	130

The Data Domain system has recommended stream count values. NetBackup is able to limit I/O streams on a given disk pool at a high level, and at more granular levels within a storage unit. Solution architects

should be cautious and monitor the number of simultaneously executing backup and duplication jobs based on prescribed limits to optimize performance.

- Limit I/O streams per volume with a setting that corresponds to the Total Data Streams value for the Data Domain system being configured. The storage unit **Maximum concurrent jobs** should be set based on backup and inbound optimized duplication jobs after considering streams required for any outbound optimized duplication jobs, or duplication jobs to tape.

The disk pool **Limit I/O streams per volume** function is available for use with NetBackup version 6.5.5; however the read side of optimized duplication operations is not counted as streams. Therefore, users should account for these uncounted streams when scheduling backup jobs and optimized duplication with this NetBackup version.

Additional information can be found in the *Symantec Veritas NetBackup Documentation Updates NetBackup 6.5.5* document, or in the *Symantec NetBackup Shared Storage Guide UNIX, Windows, Linux Release 7.0*, depending on which NetBackup version is deployed.

Duplication job configuration options

NetBackup storage lifecycle policies provide the ideal vehicle for initial backups as well as the ability to create duplicate backup images (Figure 15). Storage lifecycle policy duplication tasks initiate optimized duplication jobs on OpenStorage storage units. Consider the following:

- Use NetBackup storage lifecycle policies to control optimized duplication.
- Use storage lifecycle policies to facilitate setting different retention periods for backup and duplication jobs.
- Use **Fixed** retention periods versus **Staged capacity managed** and **Expire after duplication** retention period types.
- Use data classification in conjunction with storage lifecycle policies if desired, but doing so is not a requirement for optimized duplication.
- Use storage unit groups as backup destinations within storage lifecycle policies but not as duplication destinations.

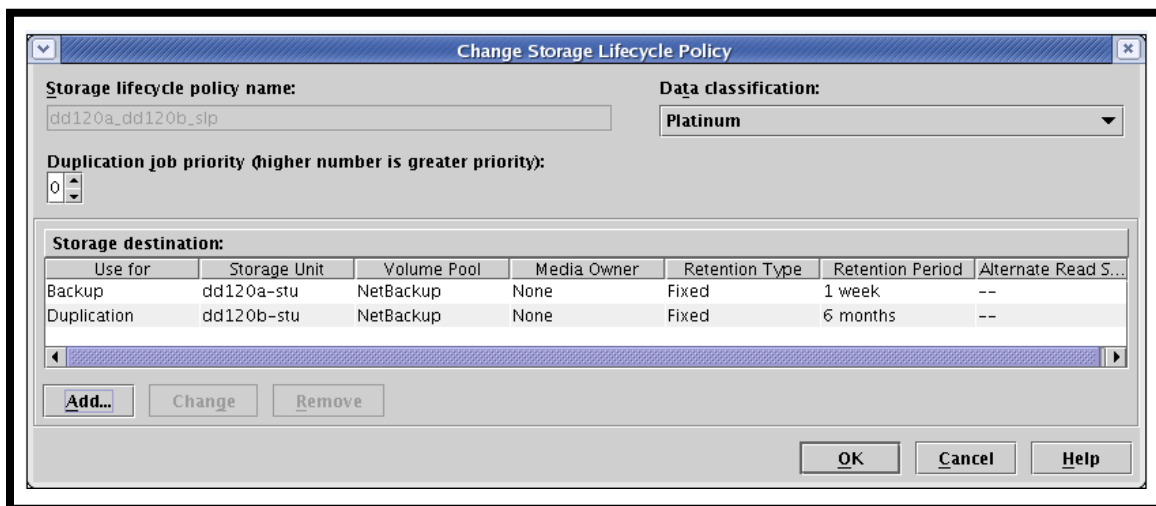


Figure 15. Change Storage Lifecycle Policy dialog box

Figure 15 contains a backup storage destination equal to storage unit “dd120a-stu” with a fixed retention period of one week. The example also includes a duplication destination equal to “dd120b-stu” with a fixed retention period of six months. The storage lifecycle policy has optionally been assigned a data classification value equal to **Platinum**. When the duplication task is executed it will result in an optimized duplication job that appears in the NetBackup activity monitor.

Storage lifecycle policy duplication relies upon certain default settings that control the point at which a duplication job launches. An optional configuration file can be created that customize lifecycles to run duplication jobs based on customer requirements. In some environments, default settings may be sufficient, but if necessary, they can be adjusted by creating a LIFECYCLE_PARAMETERS file. The *Symantec Veritas NetBackup Administrator's Guide, Volume I* should be consulted for additional information before adjusting these values.

Duplication to tape

Data Domain storage systems using DD Boost for NetBackup easily integrates with OpenStorage solutions to support retaining long-term copies of deduplicated backup images on removable tape media.

- NetBackup supports the duplication of backup images, not media or specific tape cartridges.
- NetBackup supports the creation, cataloging, and tracking of up to 10 copies of a particular backup image.

The default value for **Maximum backup copies** is **2**. This value specifies the total number of backup copies of a specific backup image that can reside in the NetBackup catalog. Deployments requiring more than two copies of a particular backup image will need to adjust this parameter accordingly.

One use case has an initial backup to a Data Domain system followed by optimized duplication to a remote Data Domain system and subsequent duplication to tape for very-long-term retention. This would require that the **Maximum backup copies** parameter be set to a value of **3**.

The **Maximum backup copies** parameter can be adjusted with the NetBackup administrative GUI via Host Properties > Master Servers > Global Attributes (Figure 16).

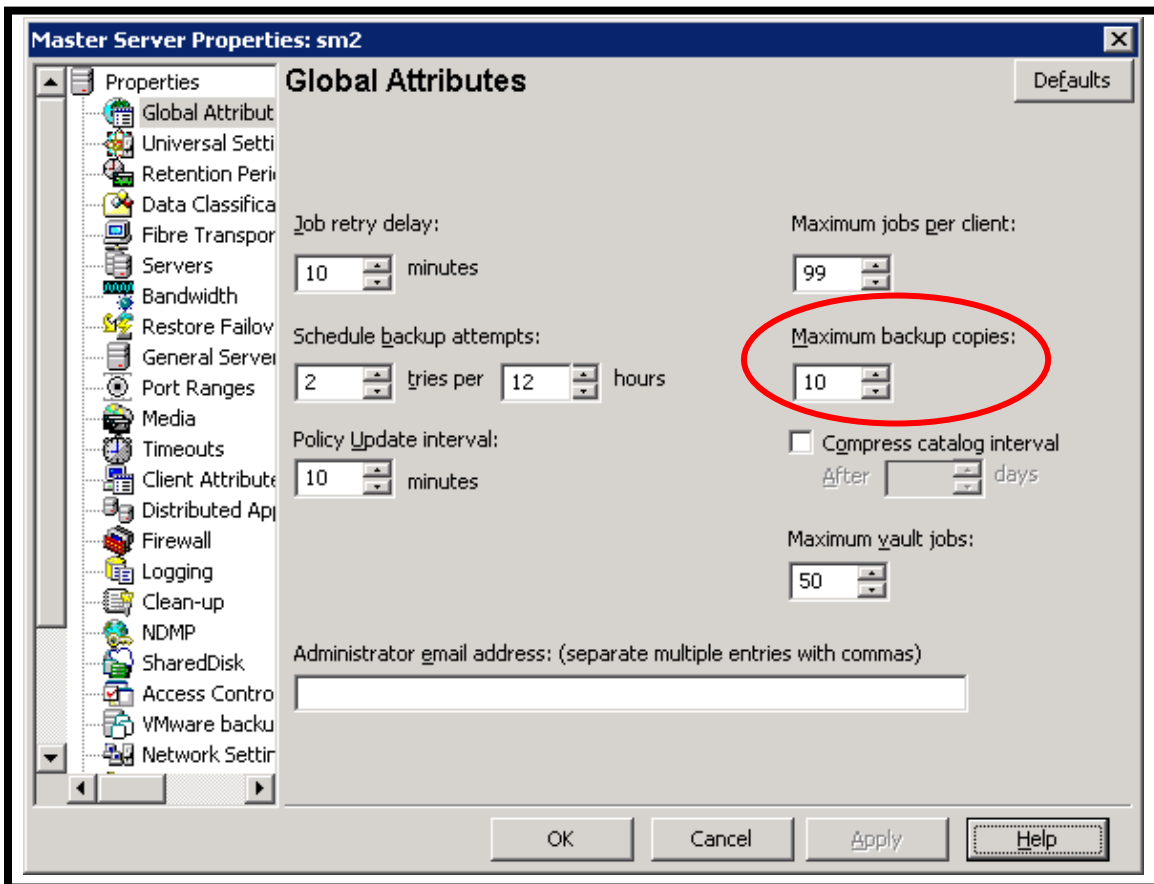


Figure 16. Maximum backup copies

By default the NetBackup global attribute **Maximum backup copies** is set to a value of **2**. Altering the value to accommodate additional copies is easily performed via the administrative GUI.

Hierarchical duplication

NetBackup versions 6.5.4 and later support a storage lifecycle policy capability referred to as “hierarchical duplication” (Figure 17). Hierarchical duplication provides the ability to create new duplicate backup image copies from a specific prior backup image copy.

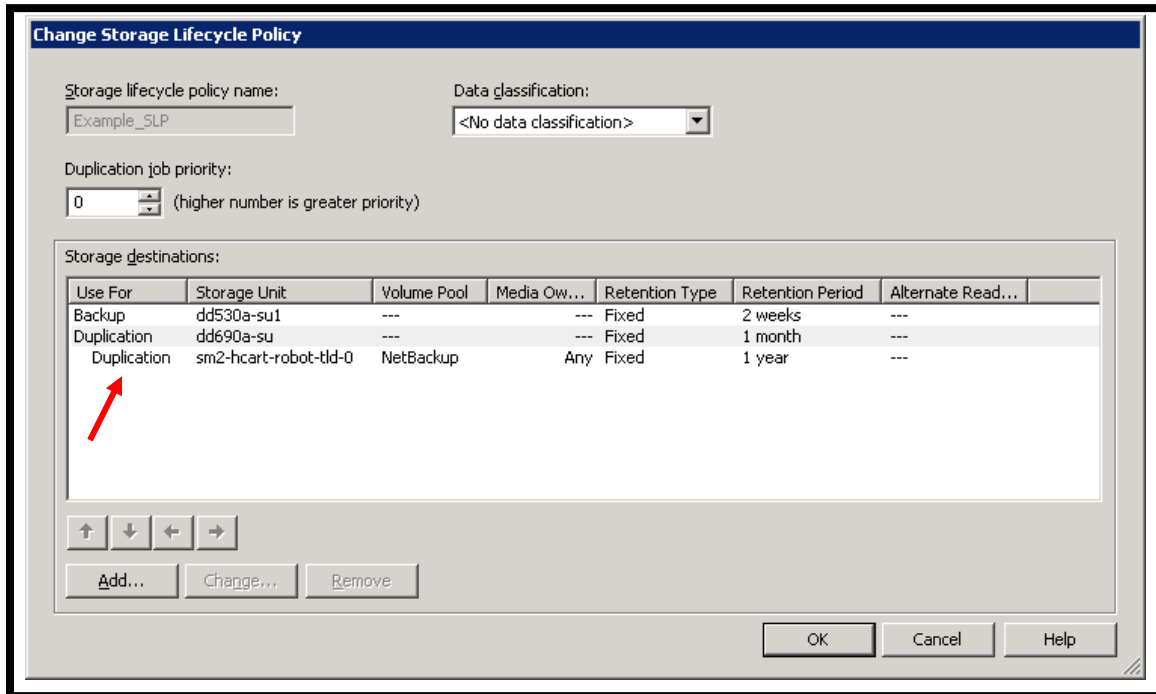


Figure 17. Storage lifecycle policy with hierarchical duplication

The storage lifecycle policy example in Figure 17 includes an initial backup destination followed by two duplication destinations. The first duplication destination will be created from the initial backup copy. The second duplication destination (indented) will be created from the previous duplicated copy. Copy 1 is the initial backup, copy 2 is created from copy 1, and copy 3 is created from copy 2.

Hierarchical duplication can be used for tape-based duplication of existing backup images. Duplicating from a geographically local copy of a backup image is more efficient than using a remote copy of the same image. In this case, the corresponding NetBackup storage lifecycle policy duplication destination should be configured so a local media server is specified as the alternate read server (Figure 18).

- Specify the use of a local NetBackup media server resource when creating tape-based duplicate backup images.

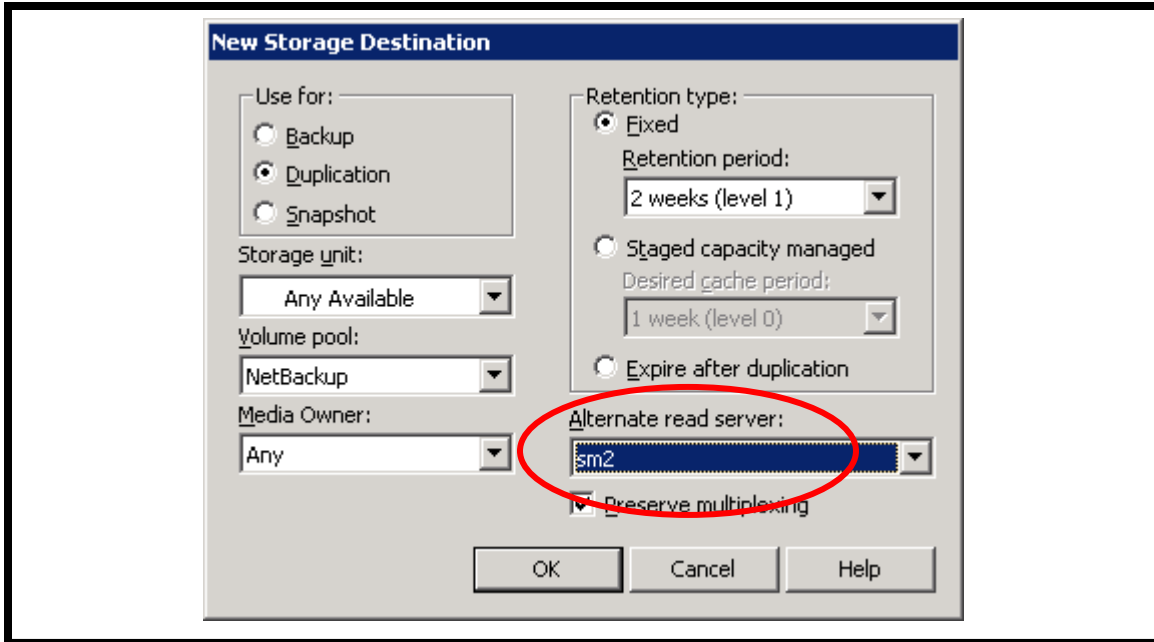


Figure 18. Alternate read server selection with hierarchical duplication

The **Alternate read server** pull-down menu is used to select a specific NetBackup media server for duplication operations. Selecting a NetBackup media server in close geographic proximity to source and destination storage units avoids the use of distant resources. Selection of a local NetBackup media server assists in achieving predictable and efficient duplication performance.

Multiple NetBackup domain configurations

By definition a “NetBackup domain” refers to a unique NetBackup installation or instance. A NetBackup domain consists of one master server, media servers, and clients. This section examines the use of Data Domain systems that are configured for use in multiple NetBackup domains, where two or more NetBackup master servers connect to a single physical Data Domain system to store their backups (Figure 19).

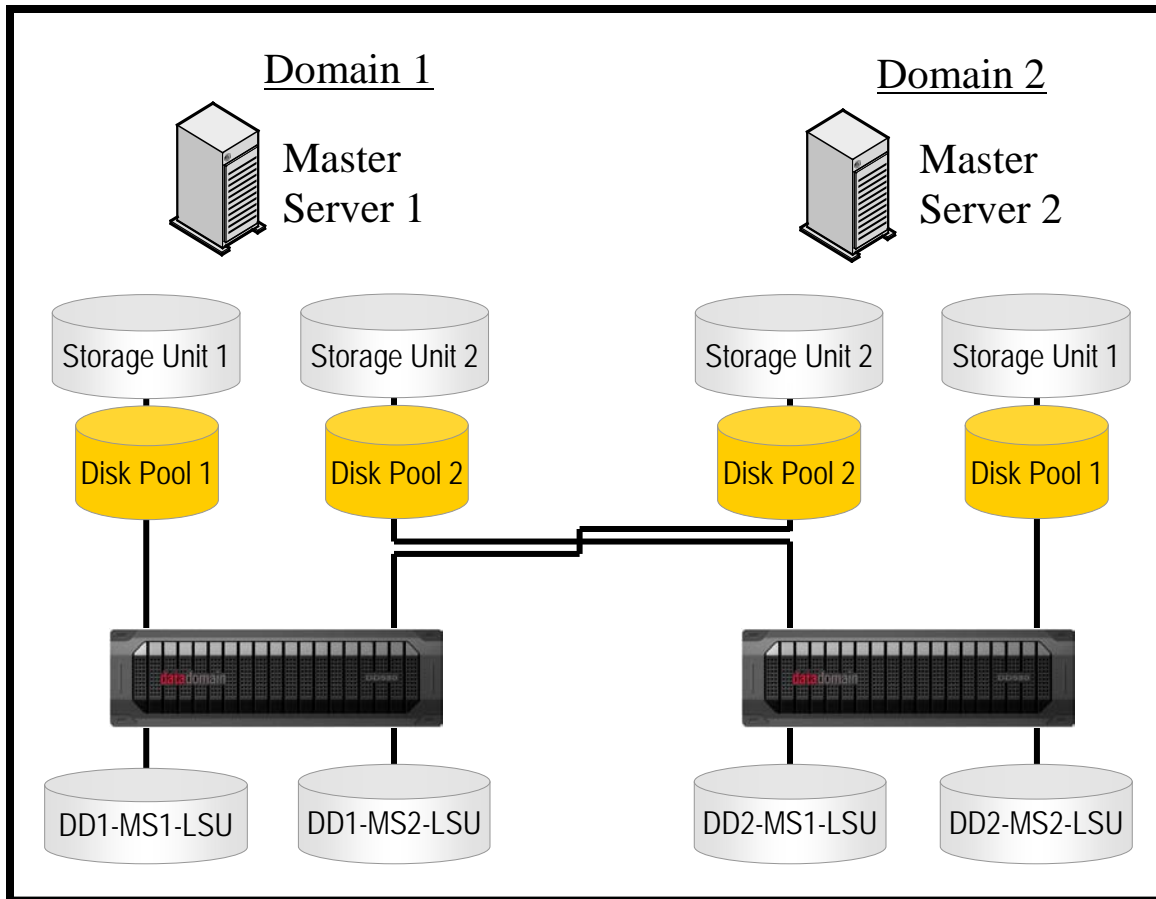


Figure 19. Data Domain systems in multiple NetBackup domains

Figure 19 shows two Data Domain systems, each of which is shared between two different NetBackup domains. Note that each Data Domain system has been configured with two LSUs, one for each NetBackup master server. Highly recommended in this type of deployment are naming conventions that assist in eliminating any potential confusion over what resources are used by each NetBackup domain.

Deploying a single Data Domain system in multiple NetBackup domains provides cost advantages compared to using different Data Domain systems for each NetBackup domain. One example is where there are two NetBackup domains and each domain is in a different data center. Data center 1 performs local backups to a Data Domain system and then uses optimized duplication to create a copy of the backup images at data center 2. Likewise, all data center 2 local backups are performed to a Data Domain system and optimized duplicates are created at data center 1. If the Data Domain systems were not shared with multiple NetBackup domains, this configuration would require four separate systems. The shared model is cost-effective in that it only requires two separate systems.

- Use naming conventions that assist in simplifying the shared use of Data Domain systems in multiple NetBackup domains.

For instance, a single Data Domain system shared by two NetBackup domains would have two LSUs configured. Both LSUs will be visible to each different NetBackup domain. Naming the LSUs such that they include the Data Domain system name, NetBackup master server name, and the “-lsu” suffix assists in properly configuring each LSU as a disk pool within NetBackup.

- Plan for the fact that complexity increases as the number of NetBackup domains that share a Data Domain system increases.

Consideration should be given to the fact that adding NetBackup domains to a Data Domain system adds LSUs, storage servers, disk pools, and storage units, creating acute management challenges.

While multiple-domain configurations provide obvious cost benefits, there are some caveats:

- NetBackup capacity reporting may be skewed as each NetBackup domain sees the Data Domain system as a unique storage pool and responds as if it has complete and total usage of the available capacity.
- The maximum number of concurrent jobs supported by a single Data Domain system is effectively shared across more than a single NetBackup domain. There is no mechanism available to arbitrate the use of available backup, restore, or optimized duplication streams between different NetBackup domains. These values are manually set by the user and may result in sub-optimal resource utilization.
- A given LSU configured on a Data Domain system intended for use by one NetBackup domain is both visible and available for configuring a disk pool on any other, perhaps unintended, NetBackup domain.

Catalog backup and recovery

Different NetBackup versions exhibit different characteristics with regard to catalog backup and recovery in conjunction with OpenStorage.

Deploy, test, and document the appropriate catalog backup and recovery solution based on the NetBackup version in use and site-specific requirements.

NetBackup versions 6.5 through 6.5.3

These versions do not support performing catalog backups to OpenStorage storage units, including Data Domain systems using DD Boost for NetBackup. Users attempting to configure catalog backup policies with OpenStorage storage units may receive a warning dialog box indicating that the storage unit type is not supported for catalog backups.

With these NetBackup versions, add a basic disk storage unit to the configuration. In this configuration, the Data Domain system functions as both an OpenStorage storage unit and a basic disk storage unit. Catalog backups should be configured to use the basic disk storage unit. In cases where a second copy of the catalog backup is required, DD Replicator directory replication should be used to replicate the catalog between Data Domain systems. The directory being replicated should be the catalog, not the LSU configured for use with OpenStorage.

NetBackup version 6.5.4

This version initiated support for catalog backups to OpenStorage storage units; however, *Symantec Veritas NetBackup Shared Storage Guide UNIX, Windows, Linux Release 6.5.4* recommends against sending catalog backups to OpenStorage storage units. This Symantec recommendation is based on the extra tasks in configuring an OpenStorage storage server at catalog recovery time. Despite the Symantec recommendation, catalog backup and recovery using OpenStorage storage units is functional with this NetBackup version.

Users are still advised to avoid using storage lifecycle policies to duplicate catalog backups with this NetBackup version. Known issues involving catalog recovery from duplicate copies exist with advanced disk, tape, and OpenStorage in NetBackup version 6.5.4. The issues have been documented by Symantec. This URL provides additional information:

<http://seer.entsupport.symantec.com/docs/327439.htm>

Users of this NetBackup version needing a second copy of catalog backups should add a basic disk storage unit to the configuration. In this configuration, the Data Domain system functions as both an OpenStorage storage unit and a basic disk storage unit. Catalog backups should be configured to use the basic disk storage unit. DD Replicator directory replication should be used to replicate the catalog between Data

Domain systems. The directory being replicated should be the catalog, not the LSU configured for use with OpenStorage.

NetBackup version 7.0

NetBackup version 7.0 sustains support for catalog backups to OpenStorage storage units. The prior Symantec recommendation against sending catalog backups to OpenStorage storage units has been removed from the *Symantec NetBackup Shared Storage Guide UNIX, Windows, Linux Release 7.0*. With NetBackup version 7.0, a wizard-based utility assists with the tasks related to configuring an OpenStorage storage server at catalog recovery time.

With NetBackup version 7, the prior restriction associated with storage-lifecycle-policy-based duplication of catalog backups has also been resolved by Symantec. Users are advised to reference this URL for additional information when recovering the NetBackup catalog from a non-primary copy:

<http://seer.entsupport.symantec.com/docs/345669.htm>

Conclusion

DD Boost for Symantec NetBackup advances the OpenStorage integration between Data Domain systems and Symantec NetBackup. Data Domain systems used with OpenStorage advance the ability to use disk as disk, allow you to store more data on disk with inline deduplication, and simplify the creation of backup copies with optimized duplication. With the addition of DD Boost, OpenStorage users improve backup throughput, reduce LAN bandwidth requirements, and decrease media server loads.

With DD Boost, NetBackup can manage DD Replicator to create duplicate backup copies at remote sites with optimized duplication, enabling advanced disaster recovery strategies. With Data Domain network-efficient replication, disaster recovery copies of backup images are created faster, and are available at the disaster recovery location sooner when compared to tape-based solutions.

Creating any required tape-based duplicate backup copies, with duplicates sourced from either the initial backup or the optimized duplicate copy, facilitates very-long-term retention of backup images. Consolidated tape operations, where fewer data centers create tape-based backup copies, is made possible via the flexibility to duplicate to tape at either the source or destination data centers.

References

The following documents can be found on the EMC Data Domain secure access customer support site at <https://my.datadomain.com/>.

- *EMC Data Domain OpenStorage (OST) Administration Guide*
- *EMC Data Domain OpenStorage (OST) Quick Start*
- *EMC Data Domain Operating System (DD OS) Administration Guide*