NDMP Backup of Dell FS Series NAS using CommVault Simpana

Dell EMC Engineering
January 2017

Dell EMC Best Practices

Revisions

Date	Description
January 2013	Initial Release
June 2013	Results added for FS7600 and FS7610 platforms
January 2017	Updated to include new branding and formatting

Acknowledgements

This best practice white paper was produced by the following members of the Dell Storage team:

Engineering: Chidambara Shashikiran

Technical Marketing: Raj Hosamani

Editing: Camille Daily

Additional contributors: Jacob Cherian, Suresh Jasrasaria, Puneet Dhawan, Gabby Lavy, Mark Welker, Andrei Ivanov, and Mike Kosacek

Table of contents

Executive summary

The exponential growth of data presents several challenges for IT administrators tasked with protecting data. Two basic considerations come into play, determining how quickly data must be recovered and how to minimize the extent of data loss. The resulting solution must balance budget considerations while:

- Meeting backup window constraints
- Minimizing required network bandwidth
- Meeting recovery service level agreements (SLAs)
- Mitigating the risk of data loss

For effective data protection, system administrators implement strategies that balance recovery point objective (RPO) and recovery time objective (RTO) considerations such as:

- Implementing enterprise class NAS systems with built in high availability and data integrity features
- Snapshots for short-term, quick, checkpoint copy and recovery of important user files
- Replication for protection of data on NAS appliances, particularly for failover in a disaster recovery scenario
- Full backups using NDMP (an application for complete backup protection that meets disaster recovery, compliance, and off-site storage requirements)

1 Introduction

The storage industry is seeing an exponential increase in the growth rate of unstructured data. Analysts agree that the growth rate of unstructured data will continue to exceed that of other data types.

This paper discusses best practices for protecting file data on Dell™ FS Series NAS Appliances using NDMP. It begins with a review of the data protection and integrity features built into the FluidFS architecture, followed by an in-depth discussion of the NDMP feature.

Extensive testing for this solution was performed using Dell FS7600 and FS7610 NAS appliances as the NDMP host and a Dell DL disk based backup and recovery appliance with CommVault® Simpana® software used as the backup server (NDMP client). It is important to note that the principles and best practices detailed in this paper could easily be applied to backup deployments involving other FS Series NAS appliances (such as FS8600 and FS8610 NAS appliances with Dell EMC SC Series storage arrays) and any other data management application (DMA) choices (such as Symantec™ provided features similar to the CommVault data interface pairs are supported).

1.1 Key findings

The key findings from the tests performed to characterize NDMP backup are listed below.

- The Fluid File System scale out architecture delivers near line rate network throughput for backup.
- The virtualized architecture of FluidFS allows system architects to blend pay-as-you-go convenience with scalable performance considerations.
- The CommVault Simpana backup solution features (such as Data Interface Pairs) and Fluid File System (FluidFS) network load balancing features (such as Virtual IP architecture) enable customers to build a scalable backup architecture. These optimizations lower the RTO by enabling up to 300% improvement in backup and restore performance compared to default NDMP configuration.
- Direct Access Recovery (DAR) helps to achieve a better RPO by enabling granular recovery options with minimal storage overhead for most practical use cases.
- Backup and restore rates are dependent on the size and layout of the files on a NAS.
- A one-size-fits-all backup strategy does not exist and an appropriate backup and recovery policy needs to be implemented based on the SLAs and requirements. The most common use case scenarios and how to address those challenges are discussed in this paper.
- The size of files and file systems must be considered while choosing a backup strategy.
- The 10 GbE based FS Series NAS appliances are most suitable for throughput hungry applications such as media files (typically large sized files such as videos and images), which usually require higher backup and restore throughput.
- A divide-and-conquer approach more effectively manages backup considerations of very large file systems.

1.2 Audience

The paper is intended for solution architects, application and storage engineers, system administrators, and IT managers who need to understand how to design, properly size, and deploy a backup solution for the FS Series based NAS appliance. It is expected that the reader has a working knowledge of NDMP architecture, FS Series NAS system administration and iSCSI SAN network design.

1.3 Terminology

The following terms are used throughout this document.

Backup server: A server responsible for backing up and restoring data.

Backup target: Any storage device such as tape, disk, or NAS connected to the backup server.

Common Internet File System (CIFS): The file sharing protocol used in Windows.

Data Management Application (DMA): A backup application that controls NDMP backup or restore session.

DL2200: Dell DL Disk-Based Backup Appliance; an integrated hardware, software, and service solution (by CommVault Simpana) that provides simplified and efficient backup and restore operations.

FluidFS: FluidFS or Dell Fluid File System is the Dell proprietary scale-out distributed file system and adds file services to Dell storage product lines. FluidFS running on a FS Series NAS appliance is also referred to as FS Series Firmware.

FS7600: The Dell NAS appliance providing 1Gb Ethernet connectivity for client and SAN.

FS7610: The Dell NAS appliance providing 10Gb Ethernet connectivity for client and SAN.

NAS Containers: To provision NAS storage, containers are created in a NAS cluster. Multiple CIFS and NFS shares can be created in these containers for user access. In the storage industry, these NAS containers are also referred as file systems.

NDMP differential backup: In the case of differential backups, the incremental backup dump level is always 1. This indicates that all changes since the last full backup (dump 0) are copied.

NDMP full backup: NDMP incremental backups are handled by setting the dump level on the NAS filer to 0. This setting indicates full backup so that the entire container is backed up.

NDMP incremental backup: Controlled by the dump level parameter (Range: 1 to 9), it copies only the changes since the last dump level.

NDMP synthetic full backup: It's a synthesized backup created from the most recent full backup and subsequent incremental and/or differential backups.

NDMP token-based incremental backup: DMA maintains the timestamp database and controlled by time token used during each incremental backup. This method does not rely on NDMP level based incremental backups.

Network Attached Storage (NAS): A self-contained computer or appliance which provides file-based data storage services to other devices on the network.

Network Data Management Protocol (NDMP): NDMP is an open-standard protocol for performing backup and restore of heterogeneous NAS appliances. NDMP provides a common interface between backup application and heterogeneous NAS devices without installing any third-party software on NAS server.

Network File System (NFS): The file sharing protocol used in a Unix network.

Recovery point objective (RPO): is the amount of data loss that's acceptable and defined by application in case of disaster.

Recovery time objective (RTO): is the amount of time it takes to recover the lost or corrupted data from backup.

2 Fluid file system architecture

The FluidFS architecture shown in Figure 2 is highly available through an underlying cluster technology that consists of multiple controllers working together, monitoring each other, and providing automatic failover capabilities. The basic implementation is a pair of controllers (FluidFS Appliance) in a cluster that can be scaled by adding additional NAS appliance depending on client workload characteristics. To achieve data distribution and maintain high availability, each controller in a cluster has access to the other controllers in the cluster through a private dedicated and redundant interconnect network.

The strengths of this architecture include:

Facilitating continued connectivity: All critical system components, including hardware and software, are redundant. Multiple network paths to each controller shield against network failure.

Mirrored Write Cache: Write data is mirrored between controllers to provide availability and prevent data loss.

Automatic recovery: FluidFS continuously monitors all hardware and software components and, in the event of failure, maintains data availability without manual intervention.

Self-healing: A cluster enables each member controller to monitor its peer. If a controller detects a service failure on a peer controller, it tries to restart the controller before initiating a failover.

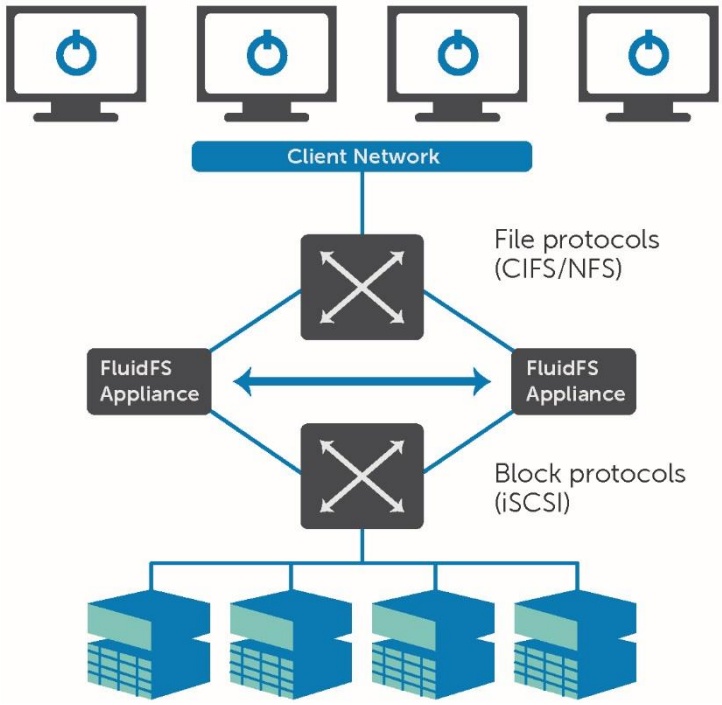


Figure 1 FluidFS Logical Architecture

3 NDMP

NDMP is an open standard protocol for enterprise-wide backup of NAS devices on a client network. The main objective of NDMP is to address issues faced by Data Management Application (DMA) vendors such as Symantec and CommVault when attempting to backup networks of heterogeneous NAS devices.

3.1 Overview and benefits

If mission-critical business data cannot be restored after a system failure, the entire business is put at risk. The system failures might include hard failures such as complete hardware malfunction or soft failures such as data corruptions due to virus infections. For this reason, most data protection strategies include backups of NAS devices using an NDMP-compliant backup application.

NDMP allows NAS device vendors to focus on maintaining compatibility with a single protocol, rather than maintaining support for multiple backup software products. Similarly, NDMP allows backup software vendors to focus on supporting the NDMP protocol, rather than multiple NAS device platforms. The main advantages of NDMP are:

- Standard, dedicated protocol optimized for NAS backup and restore operations
- Open protocol that enables flexible backup and recovery options, including: (1) full, incremental, and differential backups, and (2) quick, granular recovery of a single file or directory using the DAR feature
- Separate control and data paths for more efficient backup and restores over the client network
- A wider choice of backup software applications because there is no need for a custom software agent for each NAS vendor architecture

The illustration below summarizes the functional components of an NDMP backup solution.

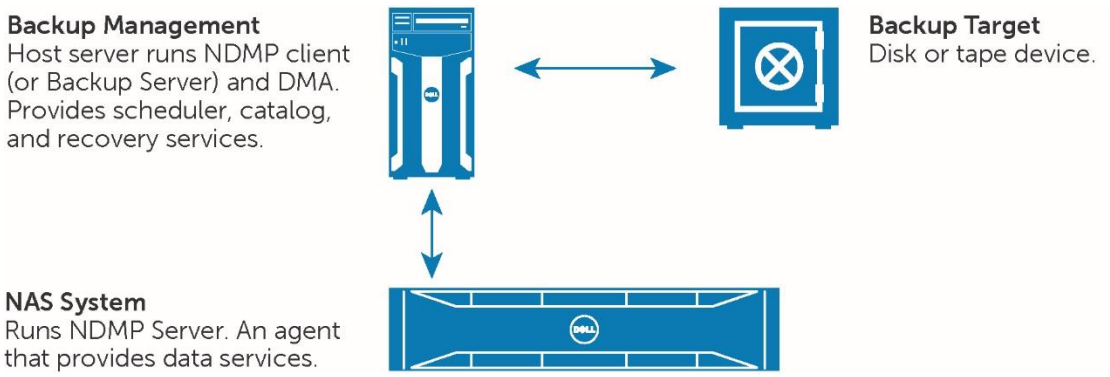


Figure 2 Functional Components of an NDMP Backup Solution

3.2 NDMP architecture

NDMP supports three methods of backup over the local area network.

Local NDMP backup: The backup target is directly attached to the NAS. The backup data is transferred directly from block storage to the attached backup target without traveling across the LAN. Only the backup control data travels across the LAN from the NDMP client running the backup software.

Remote NDMP backup: The backup target is attached to the NDMP client with backup software. The backup data is transferred from the NAS server over the LAN to the backup client, and then from the backup client to the attached backup target.

Three-Way NDMP: The backup data is sent from one NDMP server using a network connection to a remote NDMP server. The remote NDMP server will have access to the backup target.

Note: Local NDMP is not yet supported on FluidFS, but the backup target can be connected to the NAS device using a switch so that both controller ports will have access to the backup device.

3.3 NDMP backup types

NDMP backups may be full, incremental, or differential. A full backup includes all the files on a NAS device. It is the most time consuming and space intensive backup type, and is usually run no more than once a week for large data sets. Because there is a relatively long interval between backups, typical backup strategies include daily incremental or differential backups of data that has changed since the last full backup. An incremental backup includes only the data that changed that day (or since the last incremental backup). In contrast, differential backups include all the data that has changed since the last full backup. Additionally, advanced functionality of backup software, such as incremental forever and synthetic backup, can be used to significantly reduce the bandwidth and time requirements of performing a full backup.

NDMP refers to these backups as dumps, which range from dump level 0 to level 9. Table 1 shows the supported dump levels for different types of backup.

Table 1 NDMP backup types – Dump levels

Backup type	Dump level	Description
Full backup	Dump level 0	Dump level 0 indicates full backup and the entire file system content is backed up.
Differential backup	Dump level 1	The dump level for differential backups is always 1 which indicates that all changes since the last full backup (dump 0) are copied.
Incremental backup	Dump level 1-9	Controlled by dump level parameter (Range: 1 to 9). Copies only the changes since the last incremental backup.
Token based incremental backup	Time token based	DMA maintains the timestamp database and controlled by time token used during each incremental backup. This method does not rely on level based incremental backups.

An example of a backup schedule is taking a full backup at the beginning of the week followed by incremental or differential backups during the week.

For a detailed discussion on various backup strategies using NDMP backup types to meet the required RTOs and RPOs, refer to *Understanding Snapshots in Dell Fluid File System NAS*.

3.4 NDMP direct access recovery

Data protection is a continuous process of ensuring that data can be quickly recovered if it is lost. The RPO requirements include tolerance for data loss and RTO which specifies the tolerance for down time while a recovery is in progress. There are various methods available to achieve data protection, each with their own advantages and challenges.

For example, file system snapshots can be created every day and used to recover corrupt or deleted files. Typically, snapshots are not retained for extended periods of time since consumed storage space may be high based on the data change rate. When snapshots reside on the same NAS appliance, failure with NAS components might result in data loss.

NDMP DAR functionality can extend the granular recovery advantage of snapshots by dramatically reducing the time it takes to restore single files or directories. In a normal restore operation, the DMA must sequentially search a backup target for files or directories. This can be a time consuming process with large backups.

Under NDMP DAR, the DMA directly accesses backup data anywhere in a target backup set without having to read the backup set sequentially. Only the portion of the backup set that contains the data to be restored is read. DAR capability makes it feasible to quickly restore single files or directories that are no longer available on snapshots.

3.5 Fluid File System support for NDMP

FluidFS NAS solutions support standard backup software using NDMP version 4. Dell is working with industry leaders to provide comprehensive backup solutions that integrate with FluidFS. The supported backup software at the time of publication for this paper includes:

- Symantec Backup Exec™ 2012 & 2010 R3
- Symantec NetBackup™ 7.x
- CommVault Simpana 9.x
- IBM Tivoli® Storage Manager 6.3 or later
- Quest NetVault Backup 9
- EMC Networker 8.0

Other backup applications supporting NDMP version 4 may work, but were not tested by Dell prior to the publication of this paper.

NDMP support is included on each FluidFS appliance as part of the NDMP service that handles requests to backup and restore data.

3.6 Backup and restoring data

FluidFS NAS solutions support full, incremental, and differential NDMP backups (dump levels 0-9), as well as DAR, in the three-way (or remote) configuration shown in Figure 4. In this configuration, the DMA server mediates the data transfer between NAS appliance and storage device. The current release of FluidFS does not support backup to locally attached tape or disk devices.

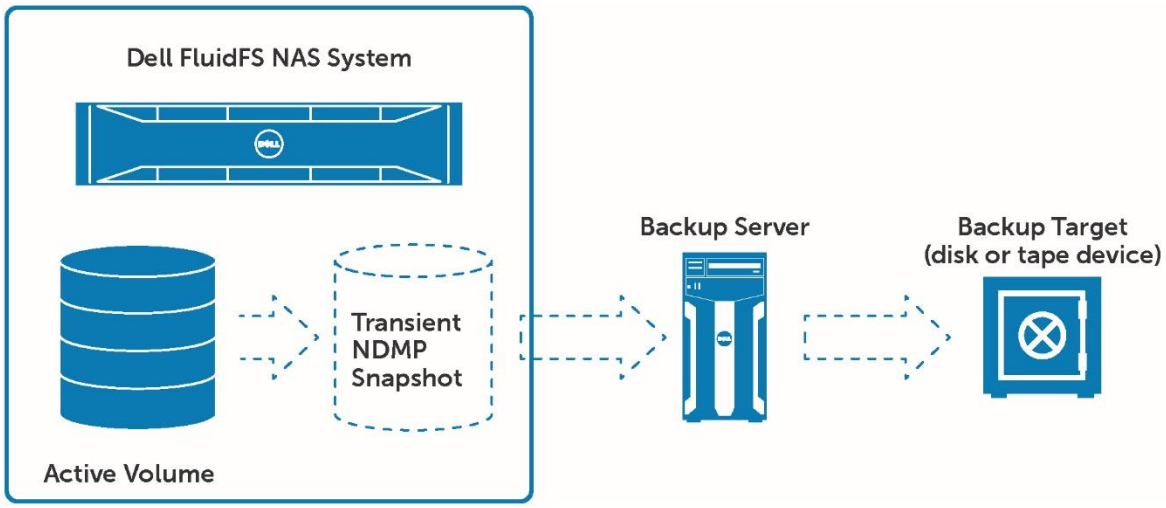


Figure 3 Typical three-way NDMP backup configuration

In the three-way (or remote) deployment, an NDMP based backup application, such as CommVault Simpana, manages a backup to the backup target. When the backup is initiated, the NDMP component on the FluidFS NAS appliance takes a snapshot of the target NAS volume or NAS container. This snapshot provides a consistent image of the file system to the NDMP component during the backup process.

To perform backup and restore operations, the DMA must be configured to be able to access the NAS appliance over the client network. The FluidFS NAS cluster solution does not use a dedicated address for backup operations, so any configured client network address (NAS Virtual IPs) can be used for backup and restore operations.

4 NDMP backup and recovery test methodology

NDMP provides backup software vendors with the flexibility to offer backup and restore capabilities without installing any software agents on the NAS servers. There are many data protection products available for performing NDMP backup. In this solution, a Dell DL disk based backup and recovery appliance with CommVault Simpana software was used as the backup server. The backup server or DMA is responsible for managing the control data (such as scheduling, backup, restoring, etc.) on the NAS server

4.1 Test infrastructure: Component design details

The high level architecture for three-way (or remote) NDMP backup configuration used in the testing is shown in Figure 5.

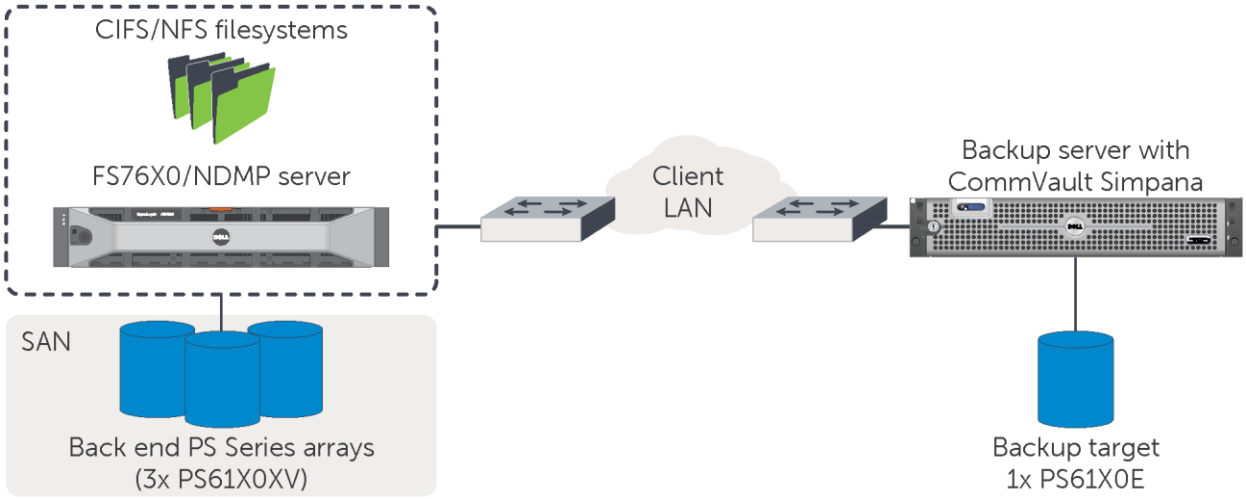


Figure 4 FS Series NAS: Three-way (or remote) NDMP backup

In this architecture, a single FS7600 was used as the NDMP host/server and has three PS series 6100XV arrays connected in the back end. A DL server with CommVault Simpana software was used as the backup server (NDMP client). This backup server was connected to a PS series PS6100E array as the backup target. Additional tests were executed using 10 GbE based FS7610 appliances with 10 Gb PS series arrays using the same architecture. A more detailed network topology of client, NAS, and SAN components used in this test are presented in Appendix A.

4.2 Test objectives

The primary objectives of the tests were to characterize the NDMP backup and recovery scenarios using FS76X0 for use cases listed below.

- Unstructured data comprised of Microsoft® Office®, Adobe® pdf, and media files. These files are usually smaller to medium in size and range from 4 KB to 1 GB.
- File shares storing streaming video and media files. These files are usually large in size and range from 1 GB to 10 GB.
- Backup of multiple containers consisting of large and small sized files.
- Single high capacity file share hosting mix of large and small sized files under multiple directories.

The primary goals of the tests were:

- Optimize NDMP backup/restore performance of file systems or containers on an FS76X0 by tuning network settings and other configuration parameters.
- Characterize NDMP backup/restore throughput for small and large sized files.
- Determine the benefits of utilizing multiple data streams for backup and recovery operations.
- Evaluate the benefits of DAR feature while performing single file recovery.
- Document key observations and provide configuration best practices based on the test results.

The tests were designed, executed and tuned using various configuration parameters to achieve optimal RTO and RPO requirements.

4.2.1 Three-way (or remote) NDMP backup – I/O flow

Characteristics for the I/O flow of the three-way (or remote) NDMP backup are:

- A backup server or DMA sending the backup request to FS76X0
- The backup data is sent to the backup server over the client LAN
- The backup server adds the index information to each file to enable faster single file recovery; this operation is done at the backup server so SAN or production I/O performance is not affected
- The backup server writes the backup data and additional index information to the backup target (PS61X0E array).

See Figure 5 for the test configuration.

More details about the solution infrastructure components, solution architecture, storage array configuration, backup server setup, network configuration, and backup server configuration can be found in Appendix A.

4.3 Test approach

The test approach can be summarized as:

- The backup operations were performed with simulated real-world file transactions.
- The NDMP backup performance of a single NAS container was measured with default network settings to obtain a baseline.
- The tests were executed to characterize the backup throughput using data sets consisting of small and large files.
- Additional tests were executed to characterize multi-container, multi-directory backups, and various file recovery scenarios.
- The network configuration was tuned to achieve optimal backup throughput.
- These tests were repeated using a 10 Gb based FS7610 NAS appliance.
- The benefits of DAR were also evaluated for single file and entire file system recovery operations.

4.4 Dataset characteristics

A dataset consisting of small to medium files, large files, and a single file share containing multiple directories was used for testing.

To characterize backup of small files, the NAS container was populated with 8 K and 16 K files. The total size of the file system consisting of 8 K files was 26 GB (approximately 1.6 million files). Similarly, the total size of the file system comprising of 16 K files was around 48 GB (about 2 million files).

With the NAS file shares storing streaming video and media files, file systems with fifteen 5 GB files (capacity: 75 GB) were used. A 1.2 TB NAS container (also referred to as a file system) was created and populated with files of varying sizes to represent the single file share with multiple directories use case. The NAS container consisted of seventeen directories and each of them were populated with files of varying size. The distribution of file size in the data set is illustrated in Figure 6.

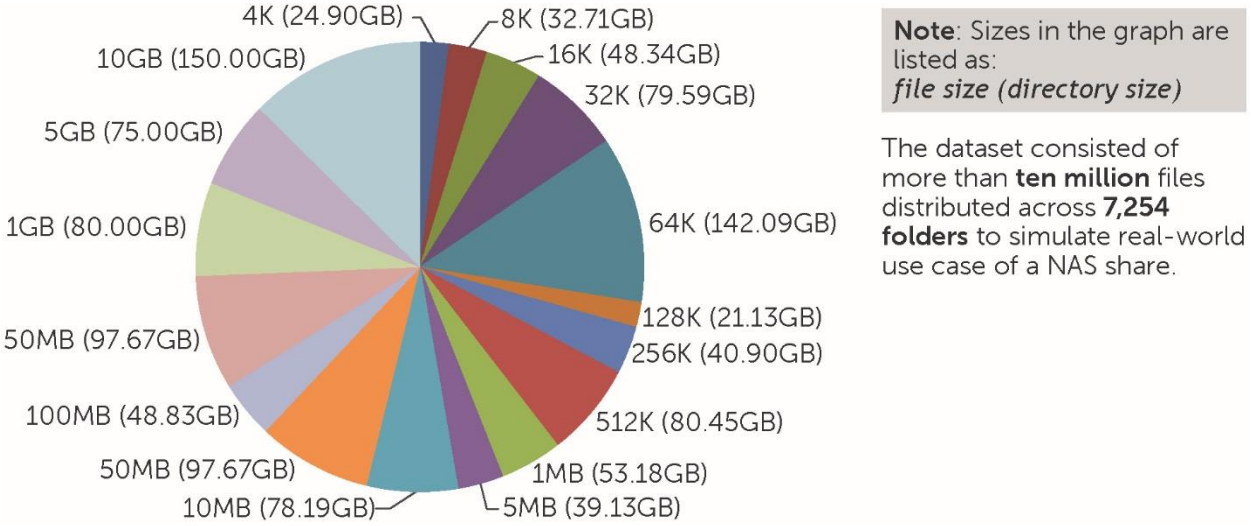


Figure 5 Dataset characteristics and distribution of files (total size: 1.2 TB)

4.5 Test tools

Load generation and monitoring tools were used to complete the tests and provide the best practices in this paper.

4.5.1 Load generation

The vdbench file system workload generator was used to populate the NAS container with different sized files. These files simulated a real world NAS data distribution. A workload of 8 K random I/Os with 70% read and 30% writes was used for evaluating the performance impact on production I/O during NDMP backup. This was achieved by executing I/O and backup operations simultaneously.

Vdbench is an open source tool and can be downloaded at http://www.oracle.com/technetwork/server-storage/vdbench-downloads-1901681.html.

4.5.2 Monitoring tools

Network utilization, CPU and memory utilization on both controllers of FS76X0 were monitored while performing NDMP backup. The PS series SAN HeadQuarters (SAN HQ) tool was used for monitoring back end storage array performance. Detailed performance metrics were also captured on the backup server using the Perfmon utility. In order to better understand the results, several engineering mode only performance monitoring utilities were used in our tests.

5 NDMP backup and recovery test results and analysis

This section describes the different NDMP backup and restore tests performed as well as the key findings from each test. For all 1 GbE configuration tests, a single FS7600 NAS appliance consisting of two active controllers and CommVault software installed on the backup server was used to perform NDMP backup. Three PS6100XV arrays were connected as NAS backend and a PS6100E array was used as the backup target. The NDMP backup and restore operation was performed on file systems consisting of files of varying size ranging from 4 K to 10 GB.

Additional tests were executed on an FS7610 appliance to characterize NDMP backup performance on the 10GbE NAS configuration. A single FS7610 appliance with three PS6110XV arrays as NAS backend was used for this purpose. Different configuration parameters were explored and tuned to achieve optimal backup/restore throughput. The detailed test results and observations are discussed in the following sections.

More details about the solution infrastructure and test configuration are available in Appendix A.

5.1 NDMP backup test scenarios

It is recommended to perform backup when there is minimal I/O on the production file system. However, it's not always possible to do this; especially in environments with 24/7 operations. Tests were executed specifically to measure the performance impact on production NAS system during an NDMP backup. Detailed test results and key observations are discussed in Section 6.1.1.

The NDMP backup throughput characterization was established first by analyzing the baseline results. In order to obtain these results, the NAS and the backup software were used in an out-of-the-box fashion without any additional optimizations. The intent was to establish a baseline that could be used to understand the impact of several optimizations evaluated through the rest of the paper. Tests executed on a 1GbE configuration using an FS7600 NAS appliance to evaluate the baseline NDMP backup performance without any optimization and the key findings are discussed in detail in Section 6.1.2.

Additional tests were executed on a 10GbE configuration using an FS7610 NAS appliance and the key results and observations from these tests are discussed in Section 6.1.3.

Various configuration parameters were tuned and optimization techniques explored to improve backup and recovery performance. These are discussed in detail on Section 6.1.4.

5.1.1 NDMP backup performance impact

These tests were executed to measure the performance impact on a production NAS system during an NDMP backup. Vdbench was used to simulate a real-world NAS client working environment. A workload of 8K random I/Os with 70% reads and 30% writes was simulated to represent a NAS client end-user collaboration environment.

Note: The performance numbers displayed in the base line graphs are not representative of the maximum performance capacity of the NAS.

The vdbench I/O workload was executed on six containers consisting of 5 GB files. Each NAS container consisted of fifteen 5 GB files (with a total capacity of 75 GB). NDMP backup of all six NAS containers was performed simultaneously while the vdbench workload was running on the production file system.

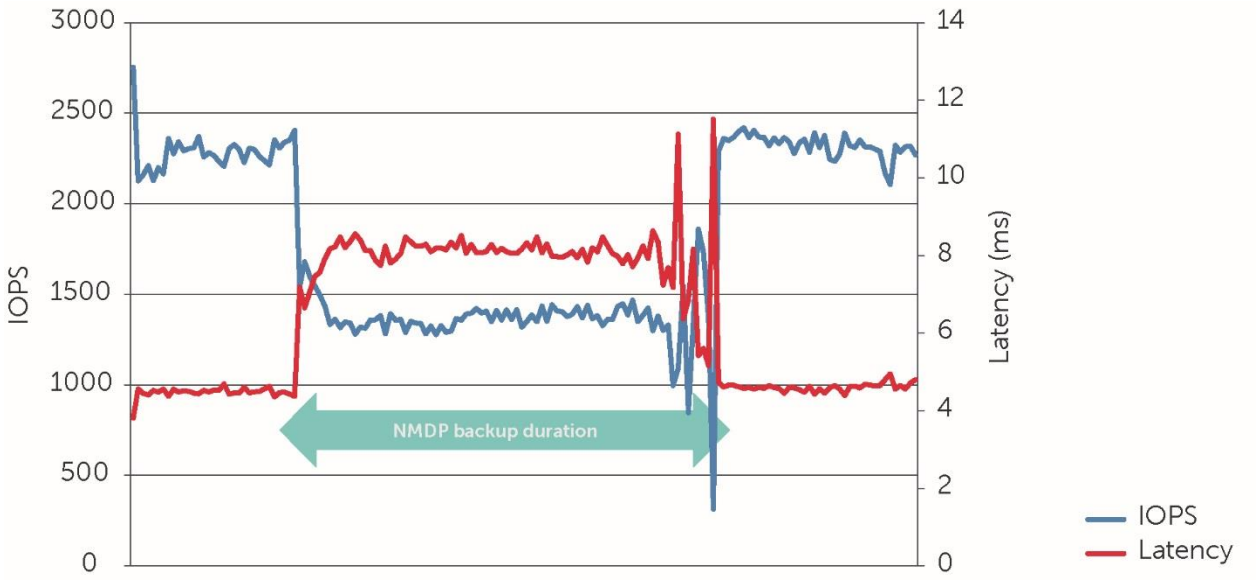


Figure 6 Performance impact due to NDMP backup (IOPS and latency)

As shown in Figure 6, there was an impact to performance during NDMP backup. The IOPS reduced from 2350 to 1350 and the response time increased significantly from 4.4 to 8.4 milliseconds. The extent of I/O degradation is largely dependent on the I/O workload issued to the NAS by NAS clients and backup processes. This test was executed using the FS7600 test configuration, but the same kind of I/O degradation is expected even on a 10GbE based FS7610 configuration during backup. The data in Figure 6 clearly demonstrates the need to find a time of low production I/O activity to schedule backups.

Note: Dell recommends performing backup operations in planned windows with low production I/O activity as there will be significant impact to the performance of production system during NDMP backup operation.

5.1.2 Unoptimized NDMP backup performance for large and small sized files

These tests were executed using the default NDMP configuration without any optimization to establish a performance baseline to evaluate the following objectives.

- Unoptimized baseline performance of a data set consisting of large size files
- Unoptimized baseline performance of a data set consisting of small size files

5.1.2.1 NDMP backup throughput for large sized files

This scenario was designed to represent the use case of streaming/video editing type of a deployment. NDMP backup of a single container consisting of 5 GB sized files was performed without running any I/O workload on the production file system. The total size of the NAS container was 75 GB and it consisted of fifteen 5 GB sized files.

The FS7600 configuration utilized one 1 Gb NIC on the backup server in the default NDMP settings, as a result, the observed backup throughput of 90 MBps was close to the capability of a single NIC. Later in this paper, easy to implement optimizations are presented that take advantage of the higher throughput delivered by FluidFS. A detailed discussion of FluidFS load balancing can be found in the white paper titled, *Dell Fluid File System Overview* at http://en.community.dell.com/techcenter/extras/m/white_papers/20441492.

Without any optimization, CommVault Simpana uses only one data stream per backup path to transfer backup data.

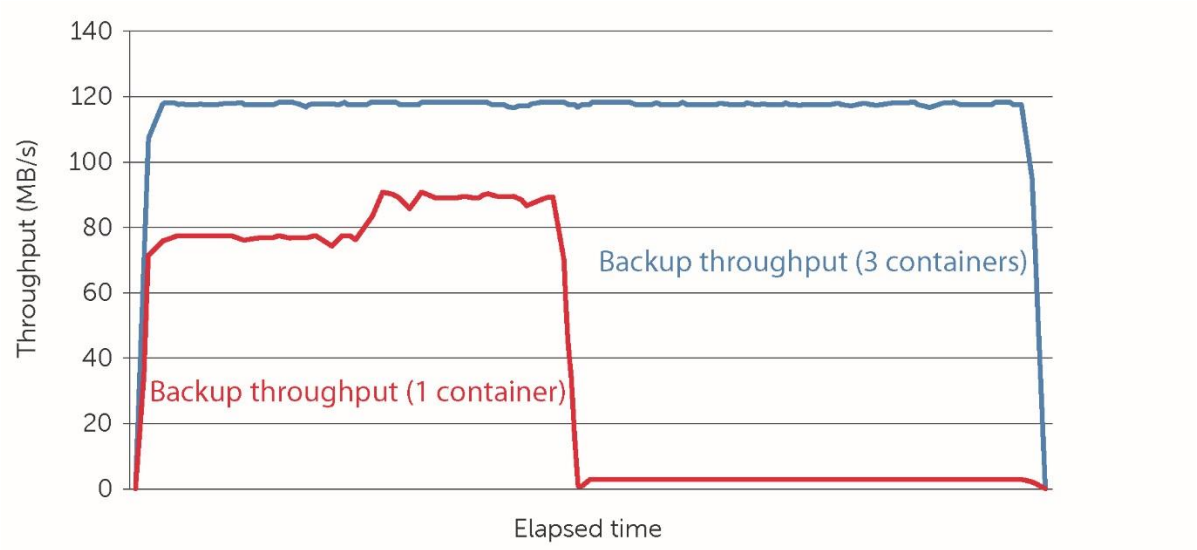


Figure 7 FS7600 Configuration: NDMP backup throughput with one and three containers

To understand the impact of file system sprawl on backup operations, a backup of three NAS containers was performed simultaneously so that multiple data streams could be utilized by CommVault. Each NAS container consisted of fifteen 5 GB files (with a total capacity of 75 GB). This time, three data streams were used while backing up three containers. A total throughput of 117 MB/sec was observed which saturated the throughput capacity of a single 1 Gb NIC. Utilizing more data streams helped in increasing the throughput from 90MB to 117MB/sec resulting in a reduction of backup time. But without any optimization on the backup network configuration, the maximum network throughput observed was around 117 MB/sec because one 1 Gb NIC

was effectively used in the backup from the FS7600. Performing backup of more than three NAS containers did not result in improved throughput as the 1 Gb NIC was saturated to its maximum capacity.

The CPU utilization on the FS7600 controllers did not exceed 40%.

Table 2 Backup of 5 GB sized files

	Avg Throughput (MB/sec)	MAX Throughput (MB/sec)	Avg CPU Utilization (%)	MAX CPU Utilization (%)
1 Container	76.84	90.73	29.1	34.24
2 Containers	112.04	116.76	25.71	36.54
3 Containers	116.9	117.57	34.74	36.2

Note: FluidFS handles backup operations of large size files (such as image and video files) effectively delivering near line rate throughput while maintaining minimal to moderate CPU utilization.

5.1.2.2 NDMP backup throughput for small sized files

This scenario is representative of typical user home shares and departmental home shares. This represents use case of file shares consisting of smaller to medium sized files. This is by far the most common use case where NAS systems are deployed. Such implementations are characterized by large number of containers (file systems) which support a large number of users creating large number of files of small to medium size. It is not typical to have a single NAS container consisting of only small sized files on a NAS appliance such as an FS76X0. For example, there might be a NAS container consisting of small program files for the engineering department, a similar one for quality assurance department, and separate file shares for the marketing and accounting departments. System administrators may setup multiple file systems (containers) for administrative and ease of management purposes. In order to characterize the NDMP backup performance in such deployments, we varied the number of NAS containers from one to 12.

In this test, NAS containers consisting of small sized files (8 K and 16 K) were used. The total size of the NAS container consisting of 8 K sized files was around 26 GB (approximately 1.6 million files). Similarly, the total size of the NAS container comprising of 16 K sized files was around 48 GB (about 2 million files).

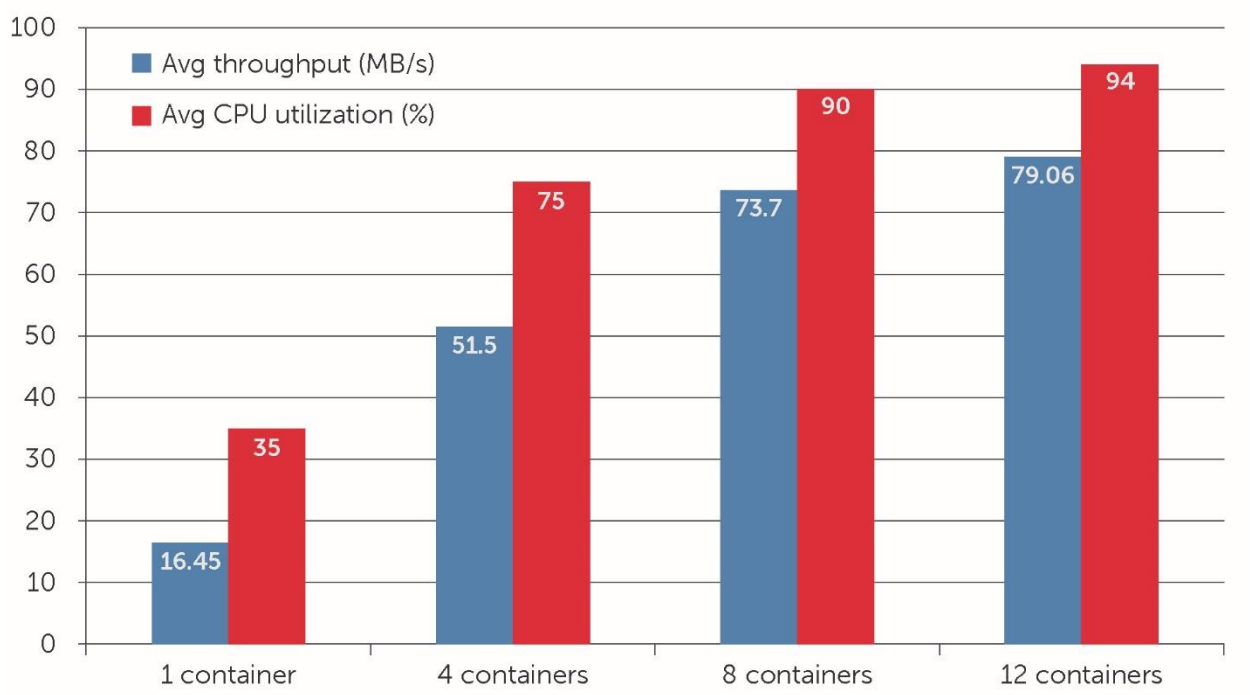


Figure 8 Backup throughput and CPU utilization for small sized files

As can be seen from Figure 9, the backup throughput scales linearly as the number of file systems increased. This resulted in significant reduction in backup time. This is a very desirable behavior as real world deployments are likely to have many file systems configured on the NAS.

It is also important to note that the average CPU utilization of NAS controller is higher compared to backup of large files as explained in Table 2. CPU utilization increases with the number of files because each file contains an independent set of metadata that must be accessed.

Backup throughput scales linearly as the number of containers are increased. This is due to the fact that CommVault will use one data stream per file system. As more data streams are utilized simultaneously, it results in higher network throughput. However, the backup server is still limited to using one NIC as the process is not aware of the FluidFS scale-out architecture.

5.1.3 FS7610 10GbE NDMP backup performance

Media files such as video and images present unique challenges to system administrators. Media files are typically larger in size ranging from several megabytes to gigabytes in size. As such, access to these files is characterized by a need for high throughput performance making the FS7610 NAS an ideal choice. The FS7610 provides 10GB Ethernet connectivity to clients and the SAN backend in addition to providing the raw bandwidth necessary to meet the needs of large file access. The backup processes leverage the high throughput performance provided by the 10GB Ethernet connectivity.

The 10GbE test configuration consisted of three PS6110XV arrays as NAS backend.

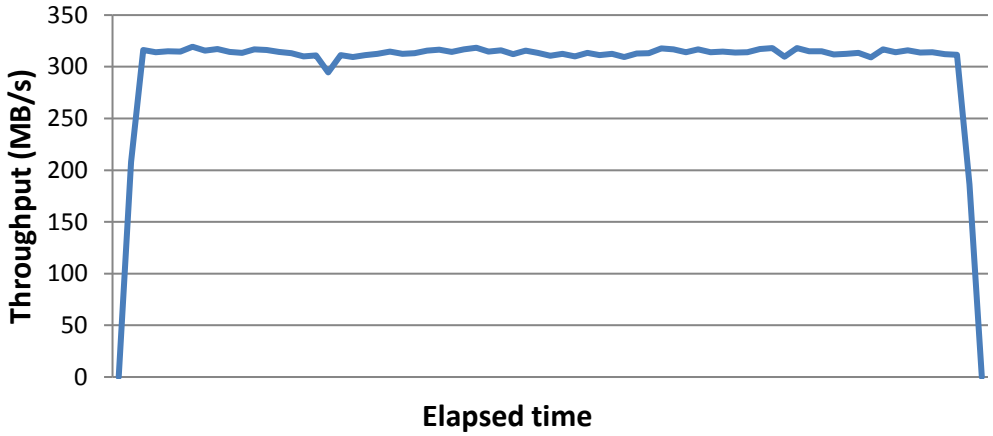


Figure 9 FS7610 Configuration: NDMP backup throughput with twelve containers

This configuration utilized a single 10 GbE NIC on the backup server in the default NDMP configuration. More than 300MB/sec backup throughput was achieved using the default FS7610 NDMP configuration compared to 117 MB/sec throughput on 1 GbE FS7600 NAS configuration as discussed in Section 6.1.2.1.

These results clearly show the significant benefits of FS7610 (10 GbE based NAS appliance) specifically for customer deployments which include large sized files and high bandwidth requirements.

A 10 GbE version of the FluidFS NAS delivers 2.5 times higher backup performance when compared to a similar 1 GbE deployment.

5.1.4 NDMP backup optimization

The following sections explore several approaches to improve backup performance. The optimizations fall in two broad categories:

- More effective utilization of NAS and backup server resources. This is enabled by the unique features built into FluidFS and CommVault Simpana
- Intelligent backup strategies that enable a divide-and-conquer approach to backup large amounts of data

5.1.4.1 Optimizing backup rate: Utilize multiple streams and multiple NICs for efficient backup

Without optimizations, the backup processes utilize one network interface on the backup server and the FS7600 appliance limits the maximum backup throughput to 120 MB/sec (theoretical max of a 1 Gb NIC). This is clearly demonstrated in the test results in Section 6.1.2. This limitation is not acceptable when the amount of data stored on the NAS grows because storage administrators will not be able to meet aggressive RTO and RPO requirements. The load balancing features built into FluidFS, combined with the Data Interface Pairs feature of CommVault, enables a significant improvement in backup and restore rates.

The following procedure utilizes multiple client facing NICs on FS7600 and the NICs dedicated for backup operations on the backup server effectively. The FluidFS architecture allows up to eight Virtual IP addresses to be defined per NAS appliance. These virtual IP addresses can be used to address the NAS independently. In a backup context, this allows creation of independent backup paths to the NAS thereby allowing the FluidFS firmware to dedicate exclusive resources to independent backup paths. As a result, a significant backup throughout performance improvement is expected.

In order to take advantage of the scale-out performance enabled by the network architecture of FluidFS, the Data Interface Pairs feature of the CommVault Simpana backup software was utilized. This feature allows creating pairs of network interfaces. Each pair consists of a single interface on the backup server and a single Virtual IP (VIP) on the NAS. The unique source-destination addresses of data interface pair enable the backup process to take advantage of all the networking resources on the NAS. Theoretically, the data interface pairs can be created using all eight client network interfaces per FS appliance, but in our test configuration we used four NICs for configuring data interface pairs as shown in Figure 11. The principles explained in this section can be applied to scale back-up rates by utilizing more NICs in deployments consisting of more arrays.

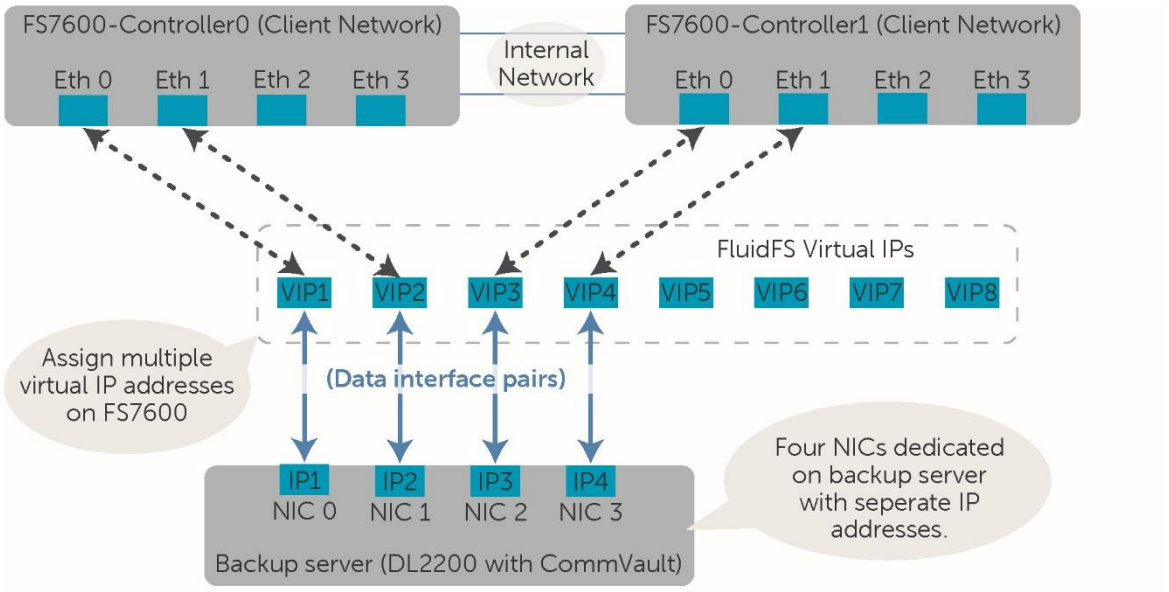


Figure 10 FluidFS virtual IP and data interface pairs configuration

The following steps utilize multiple front end NICs on the FS7600 and map the multiple NICs dedicated for the client network on the backup server.

1. Four virtual IP addresses were defined on the FS7600 as shown in Figure 12.

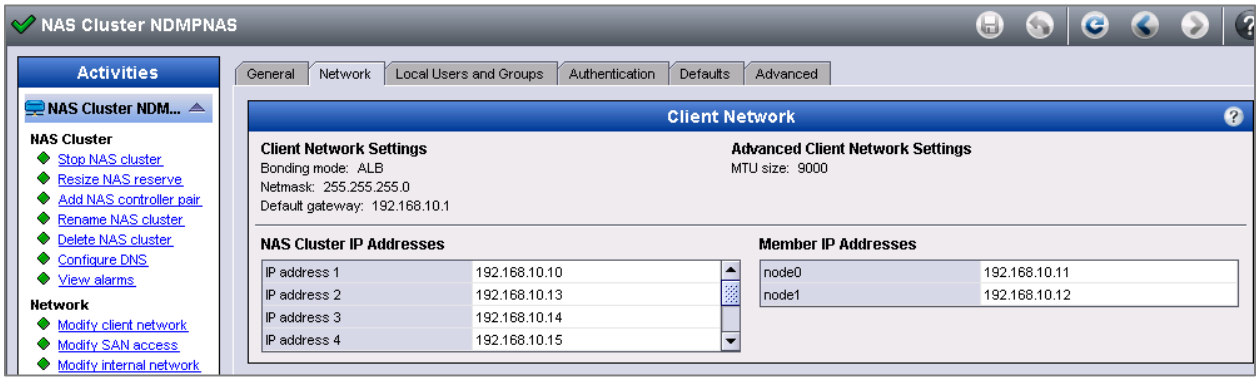


Figure 11 FS Series NAS (FS7600) virtual IP configuration

2. Four NICs were dedicated on the backup server which were configured with separate IP addresses
3. The data interface pairs feature from CommVault was used to distribute the network throughput across all four NICs dedicated for client connectivity. This feature helped in utilizing the front end NICs and thereby eliminating the single NIC network bottleneck discussed in Section 6.1.2.1.
4. As shown in Figure 13, the four virtual IP addresses on the FS7600 were added as 1st Machine Interfaces. Similarly, four IP addresses on the backup server were added as second machine interfaces. This pairing was created using the CommVault CommCell console.

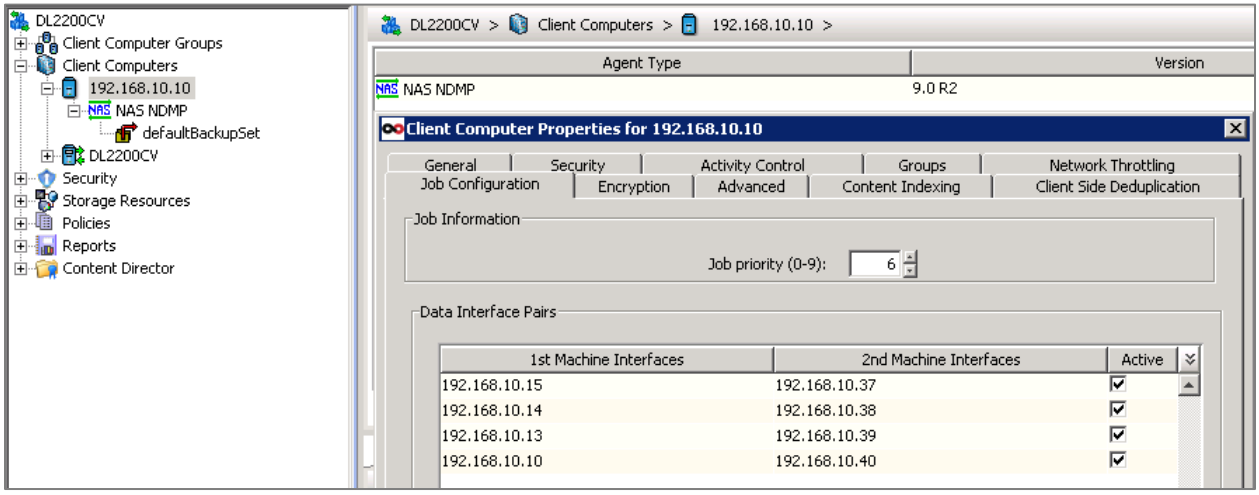


Figure 12 CommVault data interface pairs configuration

This configuration used the FluidFS load balancing feature across multiple NICs and helped increase the network throughput to achieve faster backups.

This optimization achieved an average backup throughput of approximately 278 MB/sec as shown in Figure 14 compared to about 117 MB/sec as in the case of the unoptimized configuration. The optimization created four 1 G lanes for backup and improved the backup performance by approximately 235%.

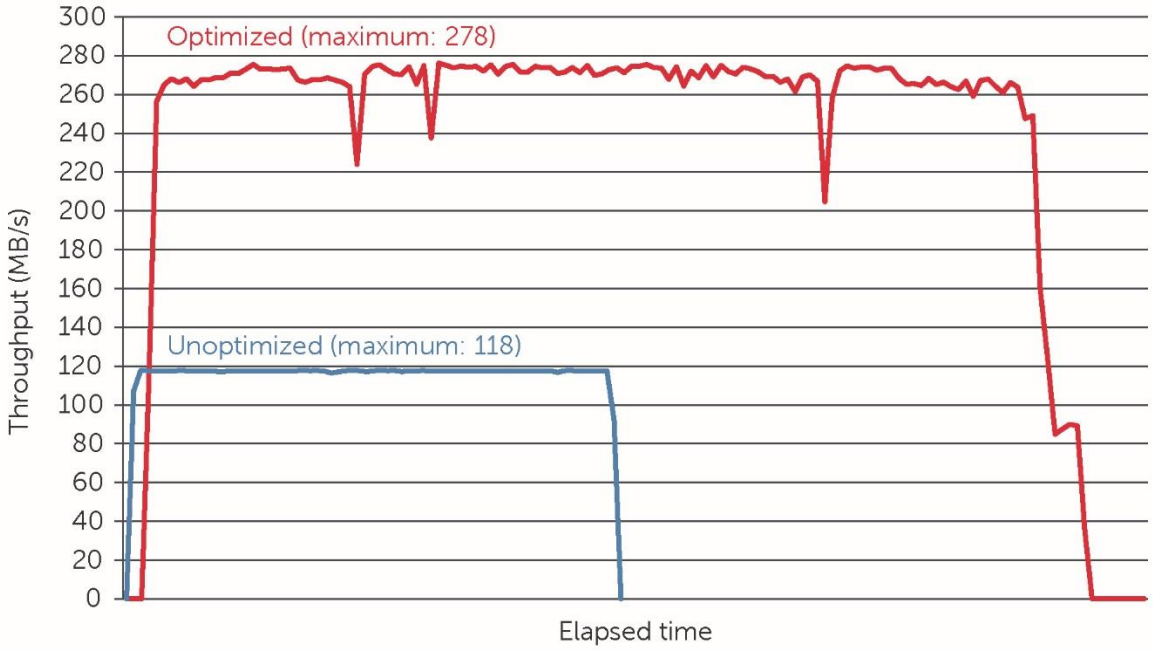


Figure 13 NDMP backup optimized throughput using FluidFS load balancing and data interface pairs

Utilizing the data interface pairs feature and FluidFS VIP together improved backup performance by more than 200%.

5.1.4.2 Optimizing backup utilizing multiple controllers

The building block of FS Series architecture is a NAS appliance which consists of two active-active NAS controllers serving a single file namespace. Each NAS controller is a file serving platform with I/O ports that connect to the client side network and the PS series SAN side network.

Built-in and automated load balancing distributes client traffic over all NAS controllers. Additionally, all disks that are part of the NAS storage pool are used to store user data. The NAS controllers also make intelligent use of the system memory to cache read and write data. These features help improve performance and maximize resource utilization.

Section 6.1.4.1 described ways to improve backup performance by utilizing multiple NICs, however, this procedure does not ensure that resources on both active controllers are effectively utilized. The following procedure should be followed to effectively utilize NAS resources on both active controllers to achieve increased backup throughput.

As explained in Section 6.1.4.1, FluidFS architecture allows defining multiple virtual IP addresses per NAS appliance and these can be used to address the NAS independently.

1. Ensure that at least two virtual IP addresses are defined per NAS appliance.

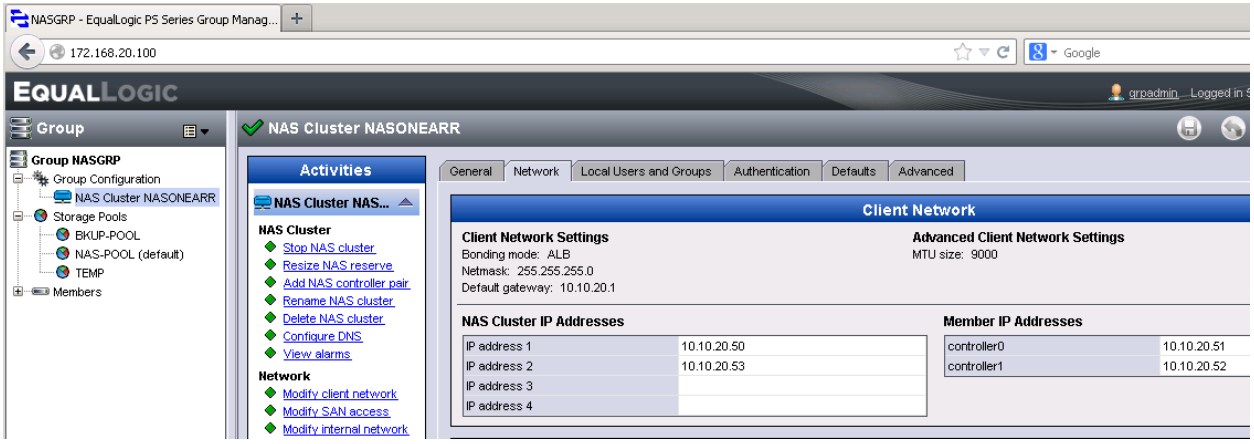


Figure 14 FS Series NAS (FS7610) virtual IP configuration

2. Add a NAS client for CommVault (Using the Commvault CommCell GUI) corresponding to each virtual IP address.

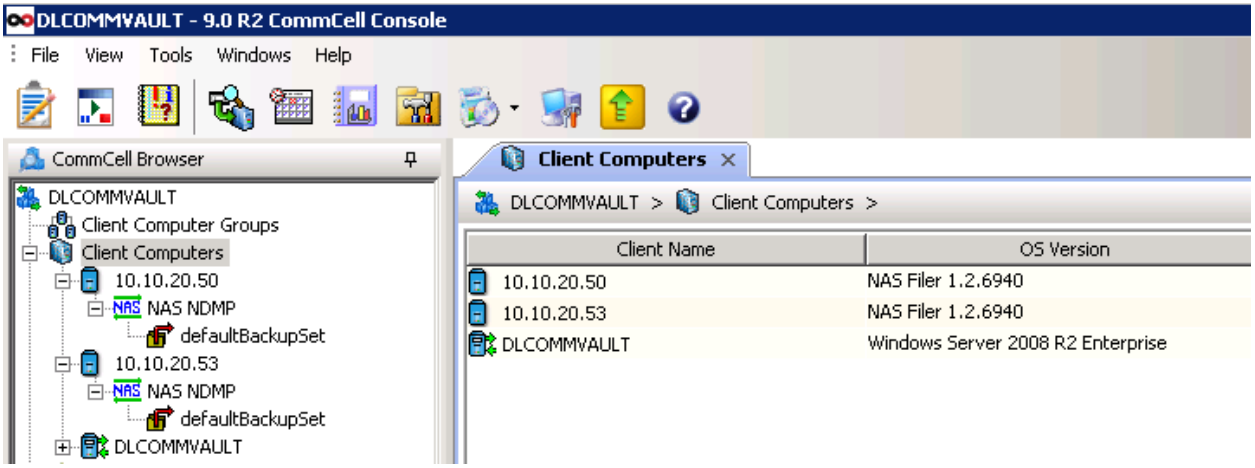


Figure 15 CommVault data interface pairs configuration

As you can see in the above image, two virtual IPs are added as different NAS clients on a CommVault server.

3. Schedule the backup jobs from both NAS clients simultaneously.
FluidFS will automatically load balance the client network traffic from each virtual IP to different controllers. This will ensure that client network traffic from the first virtual IP is handled by a specific controller and another controller is used to handle the traffic from the second virtual IP.

Implementing this technique along with the data interface pairs described in Section 6.1.4.1, will ensure that NAS resources are utilized effectively on both controllers in addition to the effective utilization of network resources.

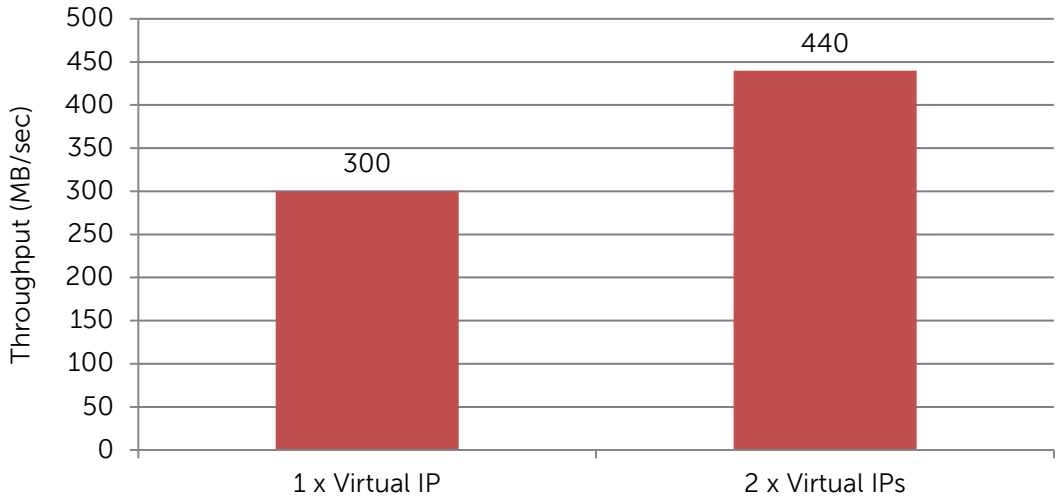


Figure 16 Backup throughput scaling with VIPs

The optimization enabled the utilization of resources on all active controllers and this effectively improved the backup performance by approximately 47% versus using a single virtual IP.

5.1.4.3 Optimizing backup of large containers

One of the key differentiating features of FluidFS is the ability to transcend the limitations of traditional file systems. The scale out architecture enables creation of a single namespace and file shares that can scale up to 509 TB per Dell PS Series group. Though a single large namespace helps in data administration, it also poses challenges in backing up large amounts of data even at theoretical line rates. System administrators may setup fewer large file shares to store data that scale up to 10s of terabytes in capacity. This approach may be used to ease management and fulfill specific business needs of the customer environment. However, it presents a challenge since the backup of 10s of terabytes within a reasonable amount of time requires a very high network throughput. Furthermore, without further optimization, the backup software may only use one stream to back up large file systems which severely limits the backup throughput.

Hence, in addition to the throughput optimizations mentioned above (Sections 6.1.4.1 and 6.1.4.2), system administrators need to fine tune their backup strategies to effectively backup containers of large size. In this series of tests, the backup performance using two approaches is explored.

- Brute force approach: The backup targets the entire container.
- Divide-and-conquer approach: Multiple jobs are setup with select top level directories.

To test this scenario, we populated a single file share with 1.2 TB of data that consisted of more than ten million files distributed across 17 top level directories as described in Section 6.1. NDMP backup was performed on this file share using both the brute force and divide-and-conquer approaches. The backup jobs were setup as follows:

- Brute force approach: A single NDMP backup job was created that executed the backup on the entire container.
- Divide-and-conquer approach: One backup job was created for each top level directory (17 in total). The directory backups (which used a sub-client for each directory) were scheduled to start at one minute increments.

The results in Figure 16 demonstrate the effectiveness of the divide-and-conquer approach. Using this approach, the backup completed in five hours and 20 minutes compared to the brute force approach which took roughly 17 hours to complete. A 300% improvement in backup performance was observed using the divide-and-conquer approach.

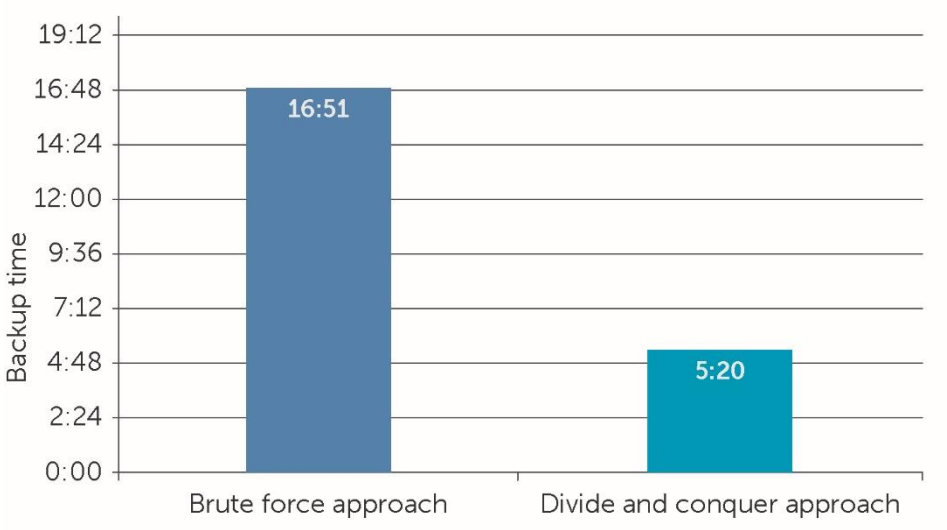


Figure 17 Backup duration – Brute force versus divide-and-conquer approach

In order to understand this behavior, we analyzed the network traffic during the backup operations. As seen in Figure 17 and Figure 18, the brute force method executed the backup in a sequential manner. As a result, the effective backup rate was bottlenecked by the throughput observed on small sized files. This was primarily because the CommVault backup software utilized one stream, and in doing so prohibited parallelism.

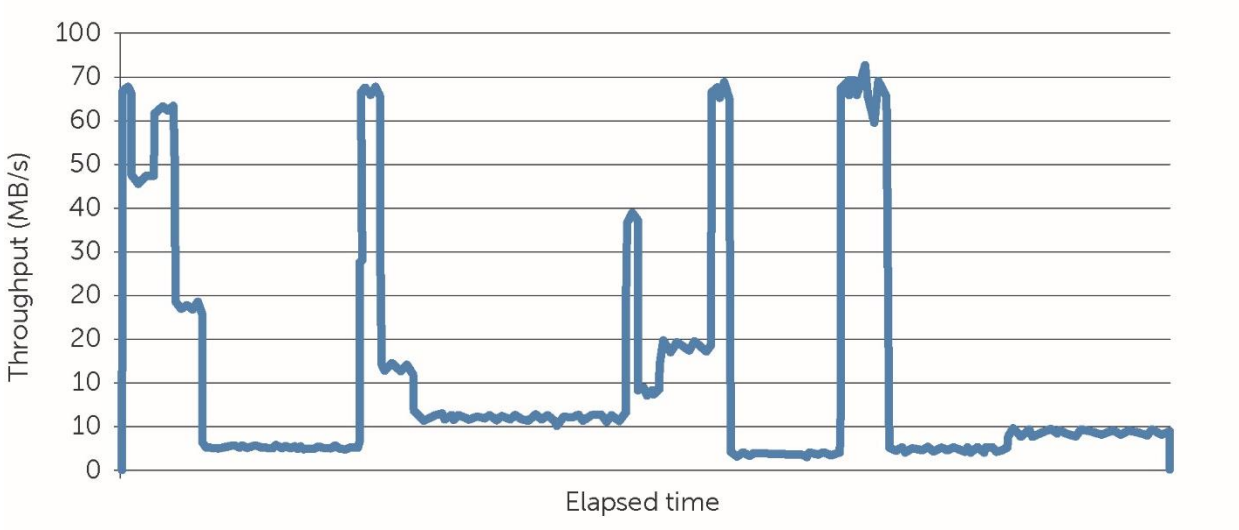


Figure 18 NDMP backup throughput using the brute force approach

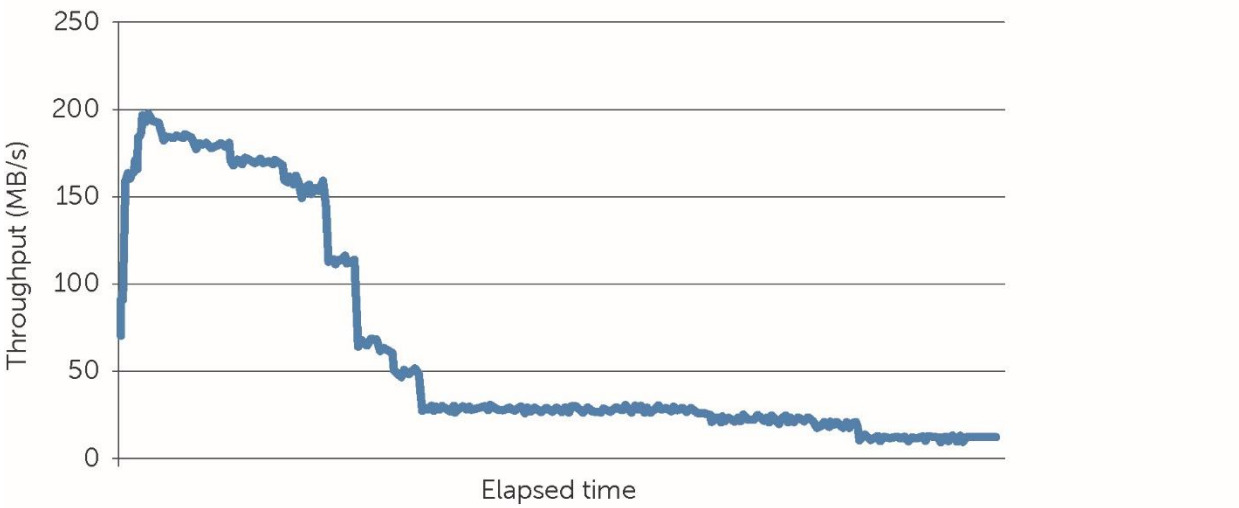


Figure 19 NDMP backup throughput using the divide-and-conquer approach

In contrast, the divide-and-conquer approach benefited from the fact that multiple streams were used to perform the backup. This resulted in a high degree of parallelism during the initial phase of the backup; effectively eliminating multiple peaks and valleys in network utilization. During the initial phase of backup, the cumulative backup throughput consistently exceeded 150 MBps. The long tail at the end of the backup was caused when the network utilization decreased after backing up directories containing large files and only small files remained.

It is important to note that effective backup throughput is largely dependent on the size and nature of data on the file share. When setting up large file shares, system administrators should understand the file size distribution, planning the directory layout, and how to use the divide-and-conquer approach to satisfying SLAs.

Implementing a divide-and-conquer approach (simultaneous backup of directories) as opposed to the brute force approach (entire file system) for large containers enables the use of multiple streams and a 300% reduction in the backup time.

5.2 NDMP recovery test scenarios

In this section, the performance characteristics of restore operations from an NDMP backup target are explored. The tests in this section utilized the backup sets from the previous sections. Recovery is an important component of a data protection strategy. The exponential growth of data presents interesting challenges to system administrators when the need for restore arises. Some of the key challenges of restore operations are:

- Improve RTO: Effective restore throughput
- Improve RPO: Granularity of a restore
- Flexibility of a restore process in regard to in-place and out-of-place restores

5.2.1 NDMP restore performance

The effectiveness of a backup strategy is often dependent on the ability to perform accurate and timely restores when the need arises. The RTO and RPO objectives completely depend on the flexibility and speed of the restore process.

This series of tests:

- Determines the restore rate of single and multiple containers which helps in achieving better RTOs
- Determines the impact of enabling DAR
- Characterizes the granular restore operations from a DAR enabled backup which helps in achieving better RPOs

5.2.1.1 Optimizing restore rate of small to medium file systems

Restore operations have unique characteristics when compared to backup. A backup attempts to make a copy of all the relevant data either in full or incrementally. A recovery is used when a catastrophic event causes data loss or data corruption. As a result, restore operations have to optimize full restores and selective restores. The first scenario we will explore is optimizing full file system restores which is needed when widespread data loss or corruption has unexpectedly occurred.

To characterize the scalability of the restore rate, tests that measured restore performance of one and four containers were run. Each container consists of 75 GB of data comprising mostly of large files. Figure 19 shows the throughput of restoring a single NAS container and simultaneous restore of four NAS containers.

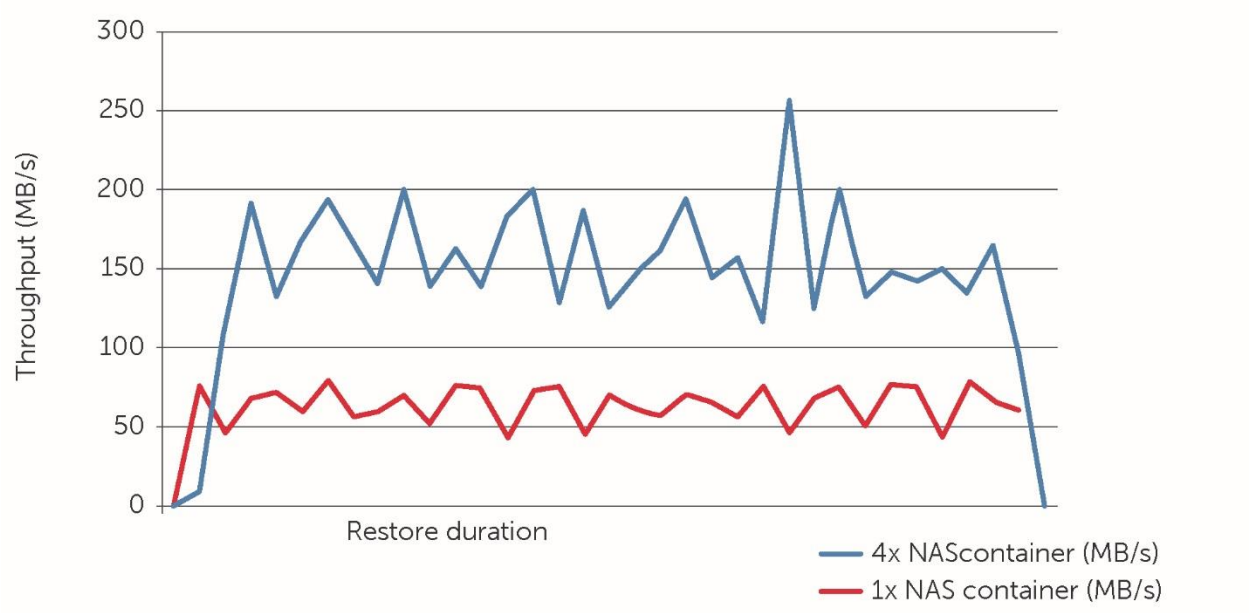


Figure 20 Restore network throughput with one and four NAS containers

As seen in Figure 19, a single container restore delivered a throughput of 70 to 90 MBps. This is primarily because the restore operation was designed to utilized just one stream. Performing restore of four containers simultaneously delivered three times the performance with restore throughput touching close to 200 MBps consistently.

The restore throughput also scales linearly as the number of containers are increased. This is because the optimizations enable the backup process to leverage multiple data streams.

5.2.1.2 Optimizing restore rate of large containers

In backup, restoring very large containers that span multiple terabytes of data requires more planning. System administrators should carefully analyze the data layout to implement a divide-and-conquer approach because restore rates, as in backup rates, are faster for large files than they are for small files.

The restore operations in this section returned 1.2 TB of data from 17 directories consisting of more than 10 Million files. The best practice for improving the network throughput with FluidFS VIPs and CommVault Data Interface Pairs was utilized in this test. Additionally, 17 subclients were setup, one per directory, to achieve parallelism. This setup was similar to the setup used in the divide-and-conquer backup approach.

Figure 20 illustrates the network throughput during the restore process where the initial part saw high degrees of parallelism. This resulted in consistent restore rates of approximately 200 MBps followed by slower rates for small files.

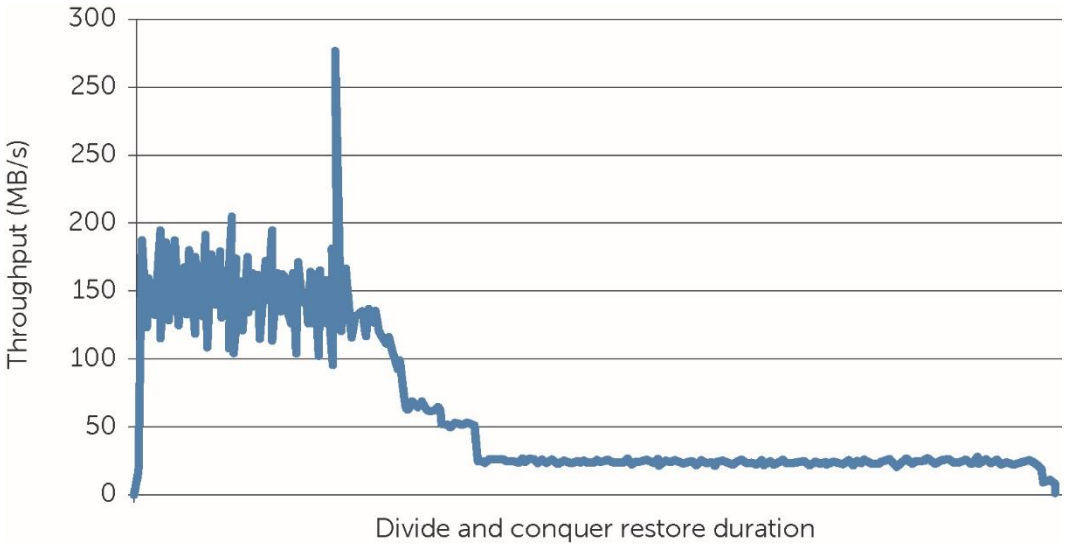


Figure 21 Multi-directory restore throughput, divide-and-conquer approach

Performing simultaneous restore of directories utilized multiple data streams and indicated significantly higher restore throughput.

5.2.2 Flexible restore

Backup strategies serve three primary purposes. First, restore all the data defined by the RPO objectives. Second, restore part of the data that has been compromised due to an unforeseen event; and finally, can redirect restore data to another location for business processes or to validate the backup data. Backing up data to a remote location is called an out-of-place restore.

The following sections analyze two scenarios which require flexible backup processes.

- Compare in place and out of place restores
- Characterize the DAR feature of NDMP

5.2.2.1 In place and out of place recovery

CommVault provides an advanced restore option called overwrite that is not enabled by default. This option helps prevent concurrent execution of multiple restore operations on the same file system.

When the overwrite option is selected, existing items will be overwritten by restore items with the same name. When deselected:

- In place: Backed up files or directories whose names are identical to those in the restore path are not restored unless they are newer than the existing file or directory
- Out of place: Backed up files or directories whose names are different from those in the restore path are restored

Test results for both in place and out of place restore scenarios are summarized in Table 3.

Table 3 In place and out of place recovery

Recovery scenario	Duration (Hr:Min:Sec)	Remarks
Recovery to the same path	01:16:27	All the files in original location were manually removed before performing a restore
Recovery to the same path using the overwrite option	01:23:27	The NAS overwrite option was used on CommVault. The files on the original location were not removed before the restore
Recovery to a different path	01:17:29	Restored files to a new file system that did not contain any of the original files

These test results show a slight increase in the restore time when the overwrite option was selected. This is because during an overwrite operation, the backup process has to check whether the file exists and perform the appropriate changes. This is potentially a metadata intensive operation. The slight increase in restore time demonstrates the metadata efficiency of Fluid File System architecture.

There was no significant difference in restore time for in-place and out-of-place recovery scenarios.

5.2.3 Direct Access Recovery

FluidFS supports the NDMP DAR feature which enables granular file recovery. This is achieved by the backup software that adds metadata to the backup file. This enables quick recovery of single files and directories without requiring a parsing of the entire data set.

CommVault Simpana enables DAR by default. When a DAR enabled backup is performed, the backup software adds 8 K bytes worth of metadata for each file that is being backed up. This metadata is added at the backup server and increases the storage requirements at the backup target. Given that the metadata is only 8 K per file, the additional storage requirements are modest for large files but may be significant if the primary data is comprised of a large number of small files. System administrators should carefully weigh the benefits of a quick recovery prior to deciding whether enabling DAR. Dell recommends enabling DAR as the benefits far outweigh the incremental cost incurred for backup target storage needs (which typically use lower cost drives).

In this test, a single 5 GB file was restored with DAR both enabled and disabled. The NAS container had fifteen 5 GB files and a single file was restored.

A single 5 GB file was restored in 3:30 minutes with DAR enabled, compared to 22 minutes when the DAR feature was not enabled.

Restoring small files is a much bigger challenge. To test this, a file system containing 48 GB of data comprising two million small size files. The data was part of a complex directory structure as shown in Figure 21.

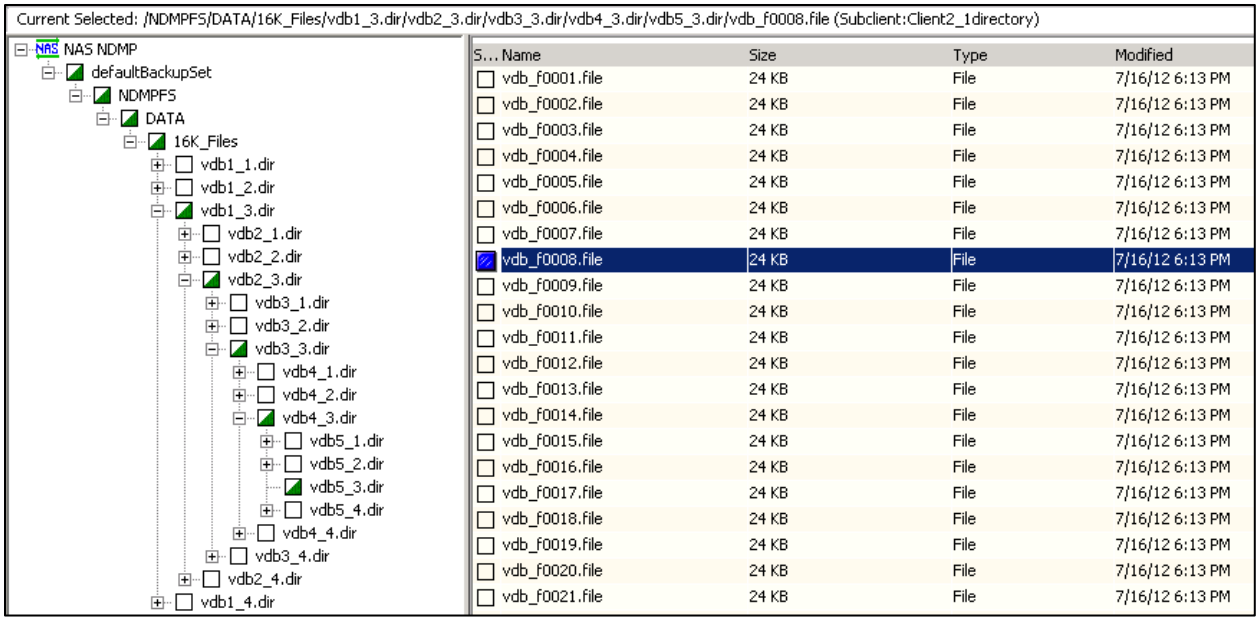


Figure 22 Directory structure for the single file restored using DAR

"vdb_f0008.file" (located under six levels of directories) was restored within one and half minutes. The same operation would have taken more than one hour and forty minutes if DAR had not been enabled.

The above results clearly show the significance and benefits of the NDMP DAR feature in a NAS environment with a huge amount of unstructured data.

Because DAR enables very fine grained, exceedingly fast recovery times for a very small increase due to additional metadata tracking, Dell recommends using the NDMP DAR feature by default.

6 Best practices: Putting it all together

As demonstrated in the previous sections, the NDMP functionality of a FS Series NAS Appliances combined with the capabilities of CommVault Simpana, allows system administrators to implement effective data protection strategies. This section summarizes the key best practices identified through analyzing the test results. The best practices fall in three categories: those unique to the FS Series platform, those unique to CommVault Simpana, and procedural best practices.

6.1 FS series Best practices

Dell recommends implementing the following best practices on the FS7600 to enable fast backup and recovery operations.

- Setup multiple Virtual IPs (VIPs) to enable better load balancing on the FS7600. The number of VIPs configured should be equal to the number of network interfaces on the backup server.
- Retain the default Adaptive load balancing (ALB) mode.
- Jumbo frames (MTU = 9216) should be set on switch ports used for SAN and internal network.
- FluidFS limits the number of snapshots per container (file system) to 512. Each backup session to a file system results in a snapshot. Therefore, if a divide-and-conquer approach is planned for a file share, limit the total number of top level directories on the file share to less than 512.
- NDMP backup makes use of the snapshot capability within FluidFS. As such, during a NDMP backup, the snapshot reserve space should be closely monitored to ensure enough reserve space capacity. The default configuration should suffice for most use cases. However, Dell recommends monitoring and adjusting the snapshot reserve space according to the specific needs of the environment.

6.2 CommVault Simpana best practices

CommVault Simpana is a robust tool used to perform backup and restore operations. Dell recommends the following actions to achieve optimal backup and restore performance.

- Where possible, configure the backup server with multiple NICs. The number of NICs on the backup server should be dictated by the backup window and RTO objectives. Dell recommends at least two NICS per DMA server. Remember to setup as many VIPs as there are NICs on the backup server.
  - Enable data interface pairs functionality. The CommVault data interface pairs feature can be used to create a relationship between the FS7600 front end NICs and the NICs dedicated for client connectivity on the backup server.
  - Each NIC on the backup or DMA server should be paired with a unique VIP on the FS7600 side.
  - The load balancing feature built into FluidFS combined with the Data Interface Pairs feature of CommVault, enables a significant improvement in backup and restore rates.
  - Refer to Section 6.1.4.1 for details about configuring data interface pairs on CommVault and how this feature, along with the FluidFS load balancing functionality, helped in performing efficient backup.
  - Refer to Section 6.1.4.2 for details about configuring multiple VIPs and multiple NAS clients to effectively utilize resources of both active controllers which helps in increasing backup throughput.

- Configure backup jobs to enable multiple streams. CommVault implementation generates one backup stream per NAS container, therefore, the backup jobs have to be structured to enable creation of multiple streams. See Appendix B for an example of how multiple streams can be configured while performing backups. The recommendations are:
  - Set the number of Data Readers equal to the number of containers being backed up within a single subclient
  - Limit the number of containers being backed up per subclient to less than 10
- Retain the Simpana default enabled DAR functionality for granular restore functionality.

6.3 Procedural and miscellaneous best practices

This section details procedural best practices that provide optimal backup and restore performance. Procedural best practices are as important as the configuration best practices.

- Because backup operations impact primary data access performance, Dell recommends choosing a window of low I/O activity to run backup processes.
- Plan a backup strategy based on the nature and layout of data. Dell recommends the following rules of thumb:
  - It is important to understand the nature and layout of the data on the NAS
  - With a large number of small containers, setup backup jobs with each job simultaneously backing up multiple containers up to 10 per job.
  - With a small number of large containers, prioritize backup requirements of directories based on business needs.
- Design a divide-and-conquer approach that groups directories as required per business need.
- Budget additional storage on the backup target for DAR metadata as DAR adds 8 Kb of additional metadata per file. Typically, this should have minimal impact, if any, on budgetary considerations. Dell recommends analyzing the requirements while configuring the backup target.
- While configuring multiple jobs, schedule them at least a minute apart. Following this best practice ensures appropriate quiescing of active I/O during background snapshot creation. Refer to Appendix B for an example of implementing a scheduling scheme.

7 Conclusions

The growth of unstructured data presents significant challenges to system administrators in designing and implementing effective data protection strategies. NDMP protocol has emerged as a popular choice for backup and restore of unstructured data. This paper evaluated NDMP V4 based backups using a FS76x0 NAS appliance.

The FS76x0 NAS appliances provide near line rates for large file backups. The scale-out architecture of FluidFS increases backup performance with the number of file systems being handled. When the scale-out architecture of FluidFS is combined with the unique features of the CommVault Simpana backup solution, it enables creation of a flexible and high performance backup strategy. The 10 GbE based FS7610 NAS appliances help with customer deployments that require high data transfer rates or backup throughput.

When the recommendations and strategies presented in this paper are combined with a FluidFS based NAS and CommVault Simpana backup solution, system administrators are enabled to design and deploy efficient data protection strategies. Furthermore, other features included in the FluidFS (such as snapshots and replication) present system administrators with a wealth of tools that can be adapted to meet aggressive RPO and RTO objectives.

As demonstrated by the testing performed in this paper, a Dell DL disk based backup solution when combined with the FluidFS based NAS, provides a robust solution for meeting today's backup and recovery requirements. Furthermore, it is important to note that the principles and best practices detailed in this paper could easily be applied to backup deployments involving any other DMA choices.

The scale out architecture of FluidFS allows pay as you grow for data protection. The results presented in this paper clearly demonstrate that system administrators do not need to worry about fork lift upgrades when RTO and RPO requirements change. By simply adding more data interface pairs and arrays to the storage backend, system administrators can scale the file system capacity and backup/restore rates simultaneously without disruption to production I/O.

A Solution configuration

This appendix provides details of the test configurations used to support this paper.

Table 4 Solution infrastructure components

Solution Configuration - Hardware Components:		Description
Servers	2 x Dell PowerEdge R810 Servers: • ESXi 5.0 • 6 internal disk drives • Broadcom 5709 Quad port NIC	8 VMs were hosted on one of the ESXi servers for hosting the NFS export. Vdbench I/O workload was executed on these VMs during backup operation. Another eight VMs were hosted on the other ESXi server for hosting the CIFS share. Vdbench I/O workload was executed on these VMs during backup operation.
Management server	1 Dell PowerEdge R710 server: • Windows Server 2008	Management server for the entire infrastructure. EQL Manager and SAN HQ EQL Monitor.
Network	Dell FS7600 1 GbE Configuration: Four Dell Networking N3000 switches Dell FS7610 10GbE Configuration: Four Dell Networking S4048-ON switches	Two switches stacked together for NAS front end. Similarly, two switches stacked together for NAS backend
NAS	One Dell FS7600 • Firmware 2.0.6940 • NDMP ver 4.0 • NDMP Remote Backup One Dell FS7610 • Firmware 2.0.6940 • NDMP ver 4.0	NAS appliance on which NDMP backup/restore operations were performed.
Storage	Three Dell PS6100XV: • 24 x 146GB 15K SAS • Quad port dual controllers • Firmware: 6.0.2 One Dell PS6100E: • 24 x 2TB 7.2K NL- SAS • Quad port dual controllers • Firmware: 6.0.2 One Dell DL Disk based backup and recovery appliance: • CommVault Simpana • Backup target: PS6100E • 10Gig PS Series variants for • FS7610 testing.	Back end storage for FS7600. Backup target attached to Dell DL disk based backup and recovery appliance. NDMP Backup appliance. Back end storage for FS7610.
Performance Monitoring	• SAN HeadQuarters – 2.2.0 • IOMeter output • Perfmon	• Performance monitoring on PS Series arrays. • Workload monitoring on NFS/CIFS servers. • Performance monitoring of each server (NFS and CIFS).

A.1 Solution architecture

The high-level architecture block diagram of the 1GbE based FS7600 test configuration is shown in Figure 22. The network connectivity from FS7600 to both LAN and SAN are illustrated, as well as the connection paths from the PS Series arrays to the SAN.

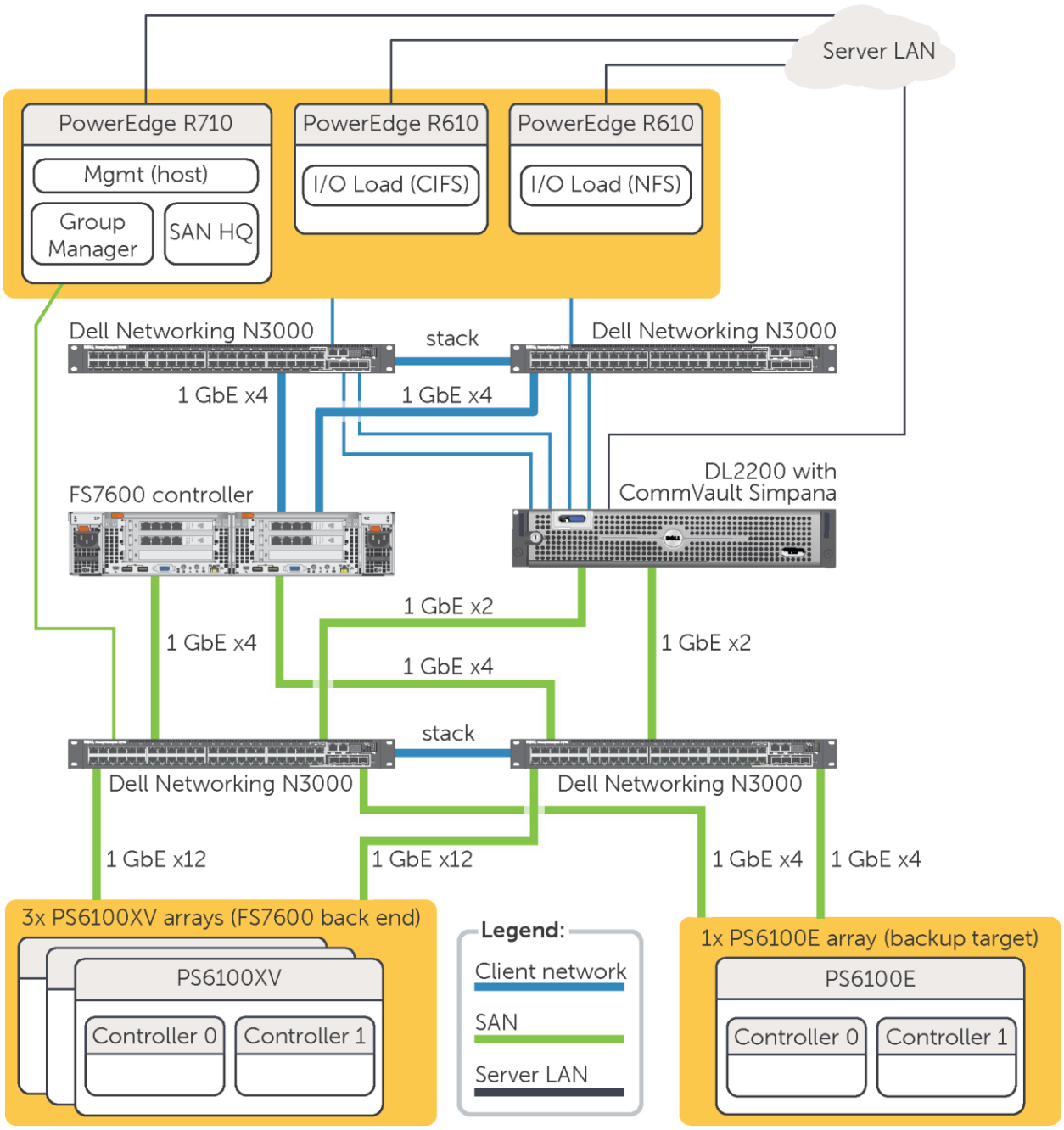


Figure 23 1GbE based FS7600 solution architecture

The high-level architecture block diagram of the 10GbE based FS7610 test configuration is shown in Figure 23. The network connectivity from FS7600 to both LAN and SAN are illustrated, as well as the connection paths from the PS Series arrays to the SAN.

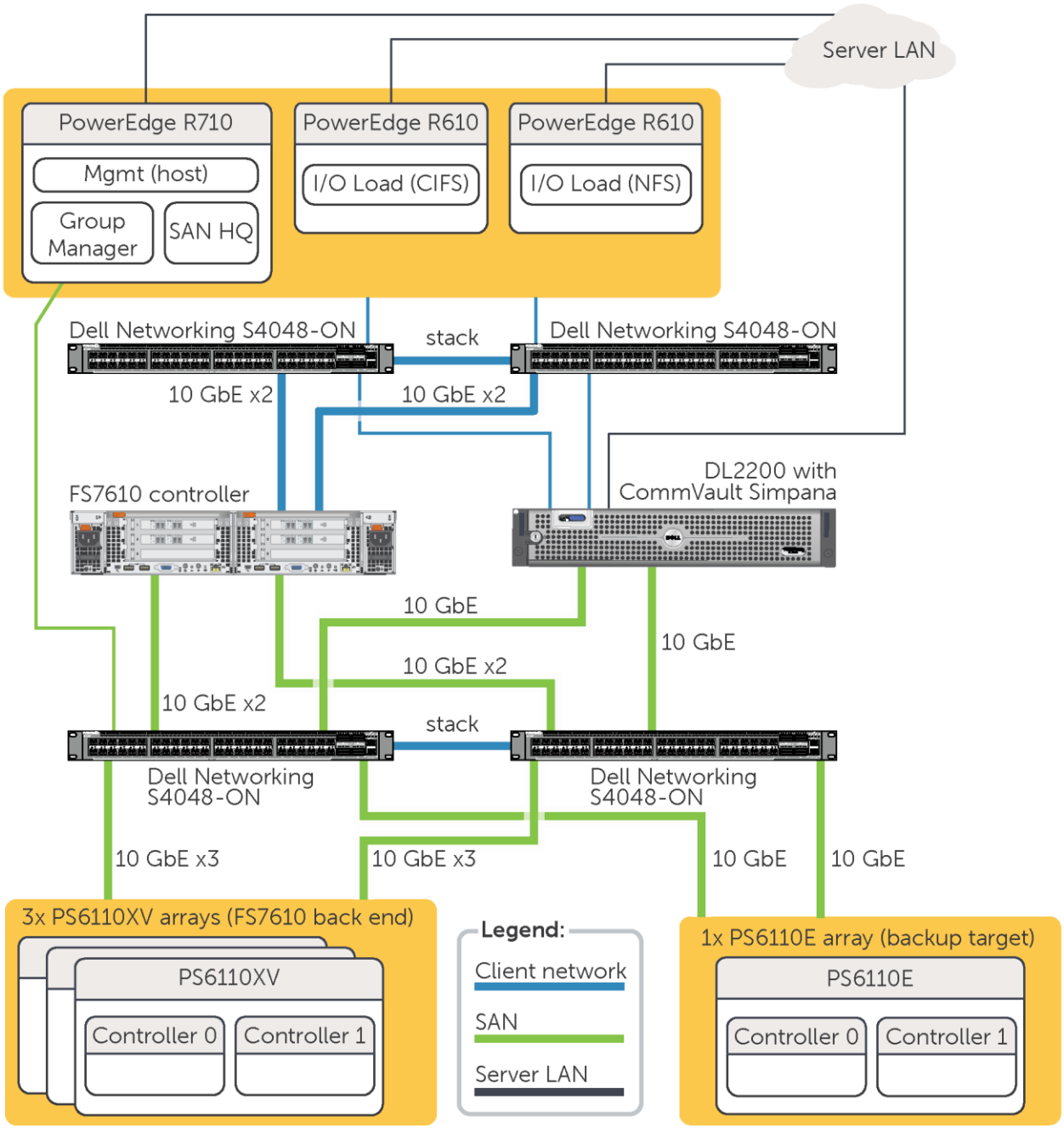


Figure 24 10GbE based FS7610 solution architecture

A.1.1 PS Series array configuration

Three PS6100XV arrays were used as back end for the FS7600 appliance. Four virtual IPs were configured on FS7600 and used for NDMP backup/recovery operations. Two dedicated Dell Networking N3000 switches were used for SAN connectivity. The front end client connections were made to a separate pair of Dell Networking N3000 switches.

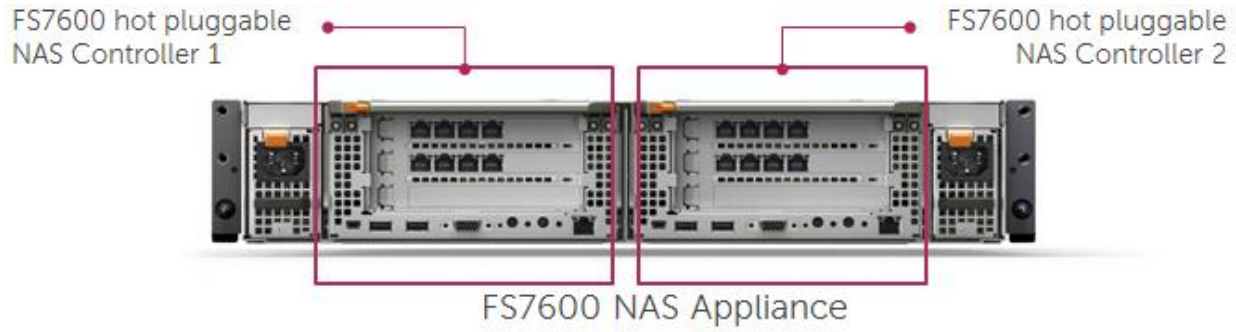
Three PS6110XV arrays were used as back end for the FS7610 appliance. Two virtual IPs were configured on FS7610 and both were used for NDMP backup/recovery operations. Two dedicated Dell Networking S4048-ON switches were used for SAN connectivity. The front end client connections were made to a separate pair of Dell Networking S4048-ON switches.

A.1.2 Backup server configuration

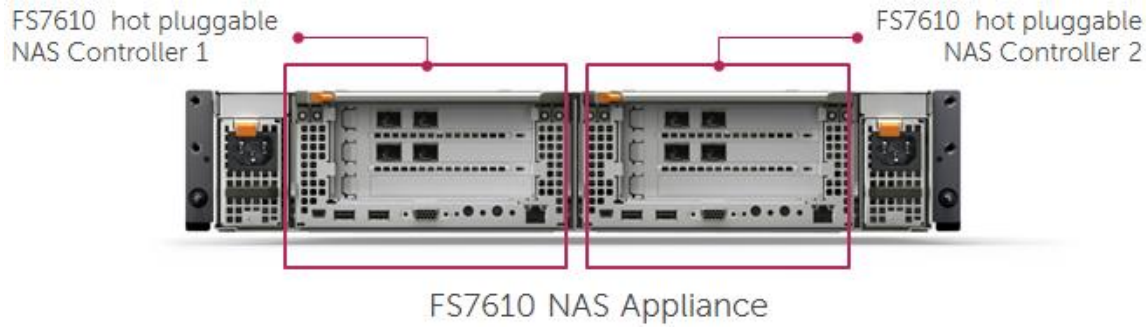
The PowerVault DLserver loaded with CommVault Simpana software used a dedicated PS series PS6100E array as the backup target. Four 1 GbE NICs were dedicated for client connectivity, and another four 1 GbE NICs for accessing the back end PS Series storage.

A.2 Network configuration

The FS7600 has a total of 16 1Gb Ethernet ports, eight ports on each controller. A total of eight ports are used to connect to the client network and eight ports to the SAN.



The FS7610 has a total of 8 10Gb SFP+ ethernet ports, 4 ports on each controller. A total of 4 ports are used to connect to client network and 4 ports to SAN.



B Backup optimization techniques

B.1 Using multiple data streams for backup

CommVault does not support multiple streams within a single container for NDMP backups. Using multiple data streams is highly recommended for better backup and recovery performance. It is also important to set the **Number of Data Readers** variable equal to the number of NAS containers being backed up simultaneously. This allows a dedicated data stream to be allocated for each NAS container.

If multiple containers require backup, then a single subclient can be created with all different NAS containers included in the same job as shown below.

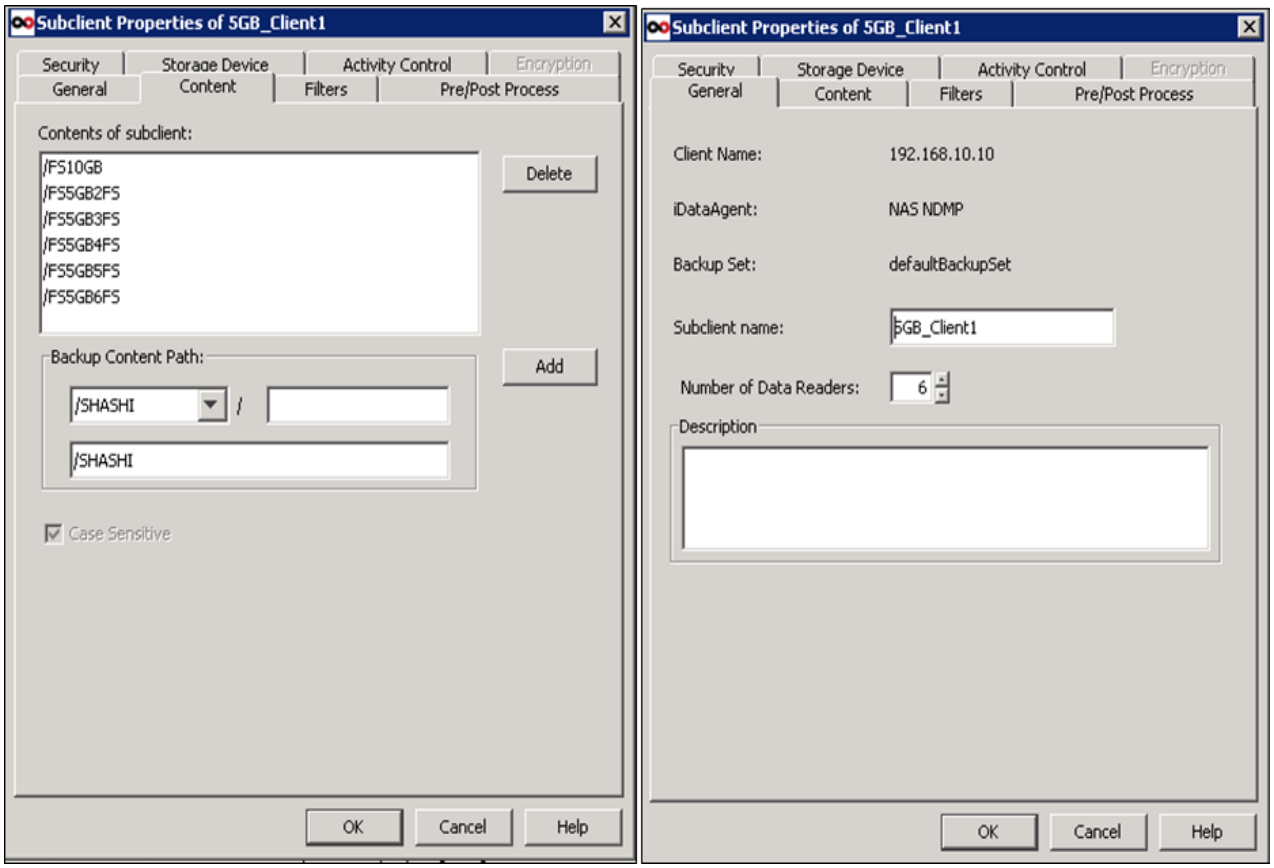


Figure 25 CommVault multi-stream backup configuration

In this example, six NAS containers were included for backup in the same subclient. Because six NAS containers were included for simultaneous backup, we need to ensure that the Number of Data Readers variable is set to six as demonstrated in Figure 23. This ensures that a dedicated data stream is allocated for the backup of each NAS container which increases the overall backup throughput.

B.2 Scheduling multi-directory backups

If there are multiple directories within a NAS container to be backed up, then a single subclient should be created for each directory as shown in Figure 24.

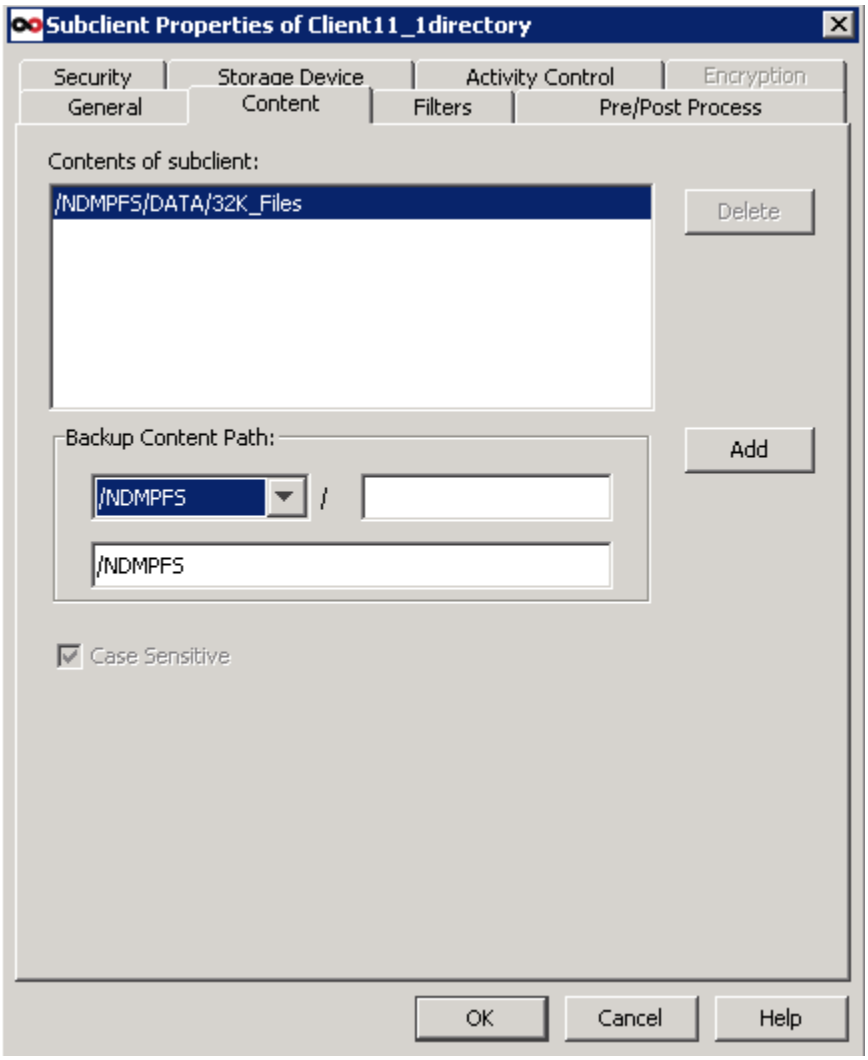


Figure 26 CommVault multi-directory backup configuration

Additional resources

Support.dell.com is focused on meeting your needs with proven services and support.

DellTechCenter.com is an IT Community where you can connect with Dell Customers and Dell employees for the purpose of sharing knowledge, best practices, and information about Dell products and your installations.

Referenced or recommended Dell publications:

- Dell Fluid File System Overview:

  http://en.community.dell.com/techcenter/extras/m/white_papers/20441492

- FS7500 NAS Scalability and Deployment

  http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19923201

- Integrating the Dell FS7500 into an existing SAN

  http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19924996

For PS Series best practices white papers, reference architectures, and sizing guidelines for enterprise applications and SANs, refer to Storage Infrastructure and Solutions Team Publications at:

dell.com/storageresources