EqualLogic FluidFS NAS Replication Technical Concepts and Best Practices

A Dell EqualLogic Reference Architecture

Dell Storage Engineering
July 2013

Revisions

Date	Description
July 2013	Initial release

Table of contents

Acknowledgements

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

Engineering: Srikanth Nandigam

Technical Marketing: Raj Hosamani

Editing: Camille Daily

Additional contributors: Mare Roll and Mike Kosacek

Feedback

We encourage readers of this publication to provide feedback on the quality and usefulness of this information by sending an email to SISfeedback@Dell.com.

SISfeedback@Dell.com

Executive summary

The defining characteristic of the information age is the creation and use of data. As businesses scale in size and complexity, implementing a continuously available IT infrastructure is a core capability no organization can live without. Maintaining continuous and uninterrupted access to the organizations data is an integral part of this equation. A variety of approaches have been designed to address the issue of maintaining uninterrupted access to data. Virtually all of them involve maintaining and managing multiple copies of data. A variety of approaches have been implemented that deliver different levels of data protection and timeliness of recovery. Replication is a time tested concept that delivers virtually seamless business continuity solution by enabling customers to easily synchronize data between two storage systems.

This technical reference architecture provides a detailed account of the replication functionality built into the Dell™ Fluid File System (FluidFS) EqualLogic™ network attached storage (NAS) appliances. The paper describes the business use cases for file based replication, an in-depth discussion of how replication is implemented within Dell EqualLogic FluidFS NAS, and a structured study of the performance characteristics of the feature.

1 Introduction

Dell™ EqualLogic™ FS7600 and FS7610 NAS appliances provide a highly scalable file system accessible via CIFS/NFS protocols using EqualLogic storage. These NAS appliances also provide features such as snapshots, replication, and NDMP backup support for protecting the mission critical data stored on FS76X0 systems. The built-in snapshot feature enables quick recovery of files or NAS volumes from data corruption, NDMP provides faster backup and restore times, and the replication feature allows the implementation of disaster recovery initiatives. The EqualLogic FS76X0 software includes support for NAS replication. This feature provides the ability to replicate NAS volumes to a peer NAS Cluster situated in remote locations without taking the source volumes offline. The NAS replication provides a disaster recovery option in case the original volume (or the entire storage array) is destroyed or otherwise becomes unavailable. FS76X0 NAS appliances have a point-in-time replication solution that offers extended distance replication over an IP network. It also provides asynchronous, incremental data synchronization between primary and secondary replicas. Scheduled replication events update the remote copy of the data with all the changes that occurred on the primary copy since the last replication event.

1.1 Benefits of replication

Replication was originally developed as a business continuity solution. By providing a mechanism by which data is automatically transferred to another storage system at a different location, organizations can maintain multiple copies of business critical data. This provides enhanced data protection and disaster tolerance in the event of an unforeseen disaster. The key benefits of replication include:

- **Business continuity:** Replication differs from backup mechanisms in that it is designed to be a business continuity solution. Separate storage systems that mirror one another are set up in different locations. In the event of a failure, the secondary site can be brought online to quickly restore access to business critical services. Unlike backup mechanisms, there is no need to restore data since the replication mechanism periodically mirrors the data from production systems to a secondary site.
- **Enhanced Data Protection:** Fundamentally, replication enhances the data protection profile of a data center. Although today's enterprise storage systems include robust data protection capabilities, they do not provide protection against site wide disasters. Replication solutions address this challenge by extending the data protection benefits to transcend site and geographical barriers.
- **Remote Data Access:** The FluidFS file system based replication allows the data to be accessed in a read only fashion at the remote site for functions such as application testing and backups.

1.2 Objective

This paper provides an overview of the replication features supported by FS7600 along with test results and derived guidelines for deploying an FS7600 based replication solution. The paper also presents test data based on the setup used for the testing conducted within Dell Labs. Other key points presented include:

- A high level understanding of FS replication
- The factors that affect effective replication deployment and an assessment of their impact
- Dell's guidelines for deploying an effective FS replication solution

1.3 Audience

The information in this paper provides storage administrators with a better understanding of the Dell FluidFS NAS replication feature and best practice guidelines for an effective deployment of a replication solution. Readers should be familiar with general concepts of EqualLogic iSCSI storage, Dell FluidFS NAS, and Ethernet local and wide area network (LAN and WAN) concepts.

1.4 Terminology

It is important to understand the following terms that are used throughout this paper.

Common internet file system (CIFS): The file sharing protocol used in Windows.

Dell Fluid File System (FluidFS): Dell's proprietary scale-out distributed file system that adds file services to Dell's EqualLogic, Compellent and PowerVault storage product lines. FluidFS running on EqualLogic FS Series NAS appliances is also referred to as FS Series Firmware.

NAS containers: Are created in a NAS cluster to provision NAS storage. Multiple CIFS and NFS shares can be created in these containers for user access. These NAS containers are also referred to as file systems or NAS Volumes.

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

NAS cluster: Consists of one or more Dell NAS appliances configured and networked running Dell FluidFS.

Network file system (NFS): The file sharing protocol widely used in Linux/Unix systems.

Recovery container: This file system is accessible when the replica container is promoted for client access on the destination cluster.

Recovery point objective (RPO): The maximum acceptable amount of data loss measured in time.

Recovery time objective (RTO): The maximum tolerable time between an unexpected failure or disaster and the resumption of normal operations and service levels.

Replica container: The point-in-time copy of the production file system on the destination cluster in a replication environment.

Replication: A service that produces a point-in-time copy of a production source file system and updates the remote file system to be consistent with the source file system.

Snapshots: A read-only, logical point-in-time view of a file system (container).

Source container: The production file system on the source NAS cluster that is accessible by clients via the CIFS/NFS protocol.

WAN emulator: A device used to simulate distance and impairments in a WAN.

2 EqualLogic FluidFS replication architectural overview

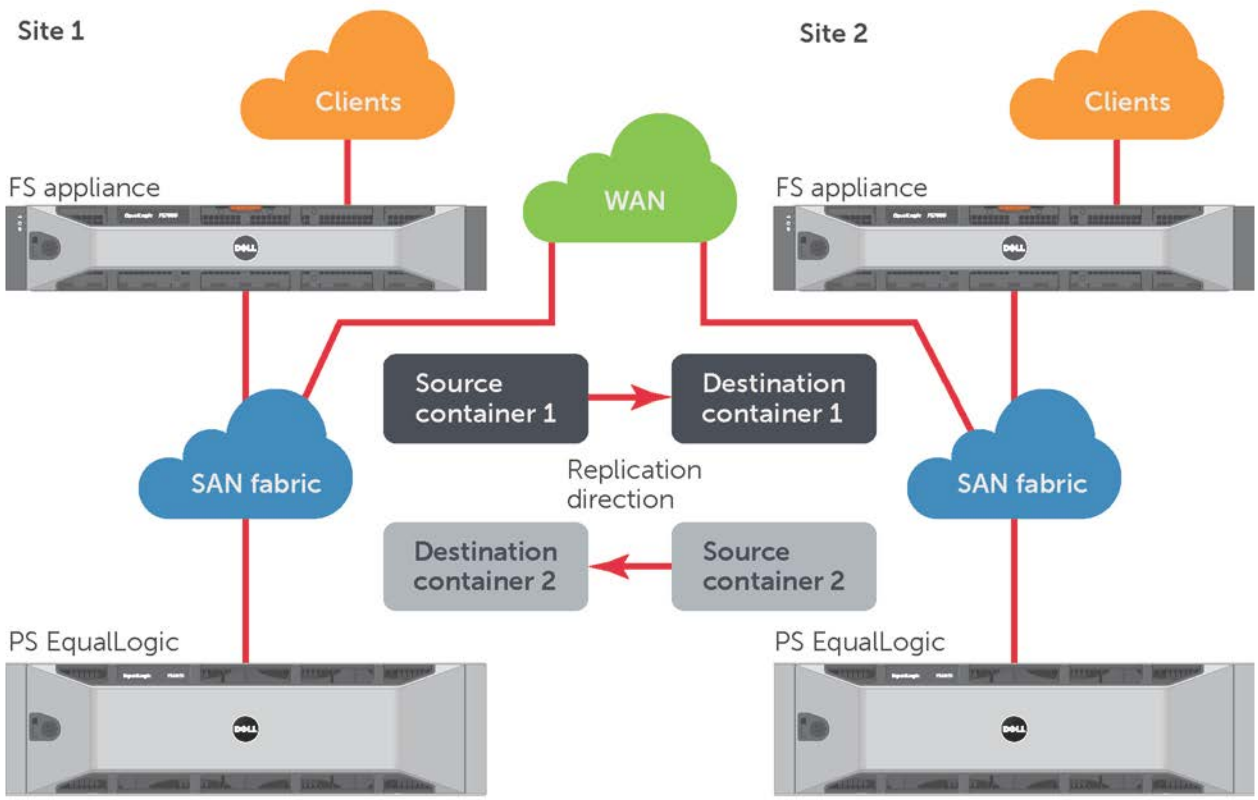

Figure 1 Replication Overview

FluidFS allows fast and reliable snapshot-based replication of any number of volumes to a partner over an IP network. To improve network bandwidth utilization, only incremental changes are replicated after the initial synchronization. This replication is native to FluidFS and does not require additional hardware. The data is always consistent on the partner site and available as read-only. Replication leverages the snapshot functionality to maintain crash consistency to the latest snapshot.

In addition to data, NAS configurations (volumes, exports, etc.) are replicated. This reduces administrative burden and enables continuous access to data in the event of a disaster or site failure to assure business continuity.

Replication is bidirectional, meaning that the same system can host both source and destination volumes. In addition, the direction can be reversed during the recovery without requiring a full resynchronization. FluidFS also supports one-to-many and many-to-one replication between NAS systems using unique volumes.

A FS76X0 FluidFS appliance utilizes a single Ethernet port per FluidFS controller (two per appliance) to replicate data to the remote system. Replication happens exclusively over the SAN network and does not use the client network. A Static Route is required on each NAS controller to access the SAN IP addresses of the partnered NAS Cluster. Although replication traffic flows through a single interface per NAS controller, data is read from the SAN storage using all available iSCSI ports. Replication bandwidth is mainly limited by the replicating interface speed of the controller. However, replication performance can be scaled linearly by adding a FluidFS appliance to the cluster (up to two FS76X0 appliances per cluster can be used).

2.1 Supported NAS replication topologies

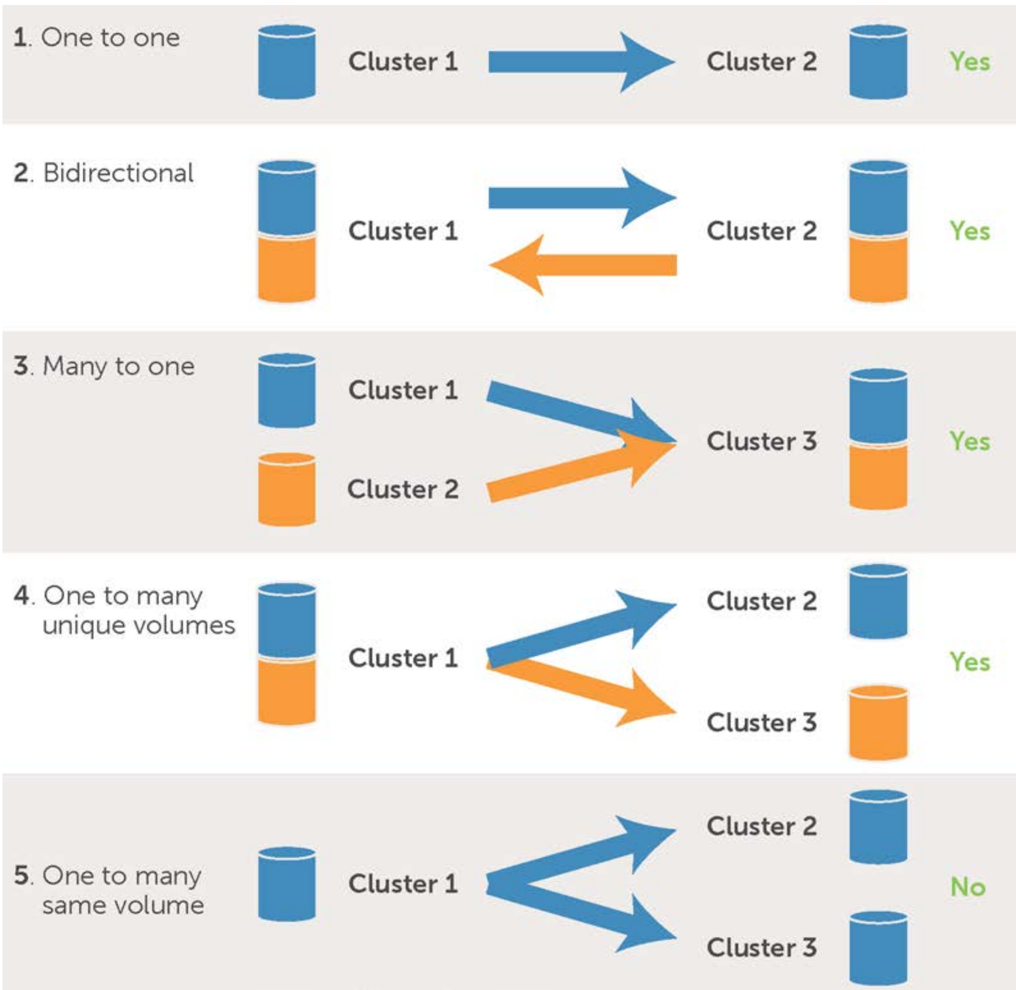

Figure 2 Replication topologies supported by a FS76X0

3 Setting up EqualLogic FluidFS NAS replication

Like SAN volume replication, NAS container replication requires that the EqualLogic (EQL) group containing the source NAS cluster be paired with a replication partner. After a replication partnership has been established, individual containers can be configured for replication. In addition to establishing EQL group level partnerships, a static route on each NAS cluster must be configured to communicate with its peer NAS cluster. Once the initial setup is complete, configure the containers for manual or scheduled replication to its remote cluster. In case of a disaster, manually promote the replica containers on the replication partner to access the data and then failback to the original source cluster that has been restored.

3.1 Network setup

FluidFS NAS replication involves at least two interconnected FluidFS clusters. It is important that the network between the replicating systems is configured correctly. Additionally, since replication is performed over standard IP connections, FluidFS replication requires configuring any firewall appropriately to enable smooth flow of data between the two sites. Network configuration for FluidFS replication across WAN link involves the following steps:

1. Configure the firewall to open the ports used by NAS replication.
2. If replication is being performed over a WAN link, add the appropriate static routes.
3. Enable jumbo frames (optional for WAN link, enabled by default for SAN).

FS7600 replication is performed using multiple and parallel TCP connections. When replication is started at the source, it establishes multiple socket connections to the remote cluster. In order to allow socket connections a specific firewall configuration is required when replicating across a WAN link.

3.2 Firewall setup

While the destination cluster listens on port 10560, the firewall in the network must allow requests to ports 10560-10576 for successful replication operations.

Table 1 Ports that must be opened on the firewall

Function	Protocol	Port Numbers
Replication set-up port	TCP	22, 26, 3260
Replication data port	TCP	2-Controller Cluster: 10560-10568 4-Controller Cluster: 10560-10576
Ephemeral port	TCP	40000 - 65535

3.2.1 Configuring static routes

When replicating over a WAN, a static route must be added. The WAN may be the public internet or a private network. A static route provides a network path for the NAS cluster and the NAS controllers to the SAN IP addresses for the partnered NAS cluster and NAS controllers. A static route must be added for the source and the destination systems.

The EqualLogic NAS manager CLI (**not** the EqualLogic Group manager CLI) can be used to add the static routes.

> **Note**: If the SAN network settings of either cluster are modified, the static routes may need to be updated to reflect the modification.

Use the following steps to add a static route when replicating over a WAN.

1. Log on to the EqualLogic NAS manager CLI for the source system.
   a. Use SSH to access the NAS Cluster SAN management VIP.
   b. Logon using **grpadmin** credentials.
2. At the EqualLogic NAS manager CLI prompt, enter the following command:

```
static-routes add <destination of the partner> gateway network  Management
–netmask <subnetmask of partner>
```

The example below demonstrates the process of configuring the static routes. For this example, the replication configuration specified in the following table is assumed.

Table 2 Static route configuration example

Source System		**Destination System**	
Name	Value	Name	Value
SAN Network	192.168.240.0/24	SAN Network	192.168.245.0/24
SAN Mgmt VIP	192.168.240.225	SAN Mgmt VIP	192.168.245.225
Gateway	192.168.240.1	Gateway	192.168.245.1

3. On the source NAS cluster enter:

```
> ssh grpadmin@192.168.240.225
[PrimaryNAS] CLI > static-routes add 192.168.245.0 192.168.240.1
Management –netmask 255.255.255.0
```

4. On the Remote NAS cluster enter:

```
 > ssh grpadmin@192.168.245.225
[RemoteNAS] CLI > static-routes add 192.168.240.0 192.168.245.1 Management
–netmask 255.255.0.0
```

3.3 Additional considerations

The following considerations apply to replica containers.

- There must be sufficient free space in the NAS container reserve on the replication partner to store the replica.
- The group will give the replica container the name of the source container, but with a randomly generated four digit string appended to the end of it. For example, if you replicate a container named *ContainerA*, then the replica may be named *ContainerA_1324*.
- If you replicate to an existing container, the destination container must either be an empty container containing no data or a container to which the source container has previously replicated.
- The source and replica containers do not need to be the same size.

4 FS7600 Replication test infrastructure

To properly design a replication scenario, administrators must understand how the network connection quality between two FS7600 NAS systems can affect replication. When data changes on the source volume, it is replicated over the WAN link. The amount of change directly affects how long it takes for each replica to complete. To help illustrate these points, two FS7600 NAS clusters for replication were setup for testing.

The test configuration (illustrated in Figure 3) consisted of a designated primary site and secondary site, although all of the hardware was physically located in the same datacenter. For storage on the primary site, three EqualLogic PS6100XV arrays were connected to a pair of Dell™ Force10™ S60 switches. On the secondary site, another pair of Force10 S60 switches were used with three EqualLogic PS6100XV arrays configured in the default pool in a single group. For redundancy, each pair of Force10 S60 switches was stacked. The primary and secondary sites were connected through an Apposite® Technologies Netropy® 10G WAN emulator. This allowed the bandwidth to be throttled in order to simulate the various speeds and conditions that might be encountered and to add impairments, such as latency, to the WAN connection.

On the primary site, a Dell™ PowerEdge™ R815 server was connected to the Dell™ PowerConnect™ 7048 switches. The PowerEdge R815 ran VMware® ESXi that hosted multiple Linux® and Windows® virtual machines. This allowed disk I/O to be generated and overwrite or change the existing data on the disk using CIFS/NFS. Several NAS containers of 100 GB, 500 GB, and 1 TB were created for the replication testing. Vdbench was used for creating different data sets from the Linux clients.

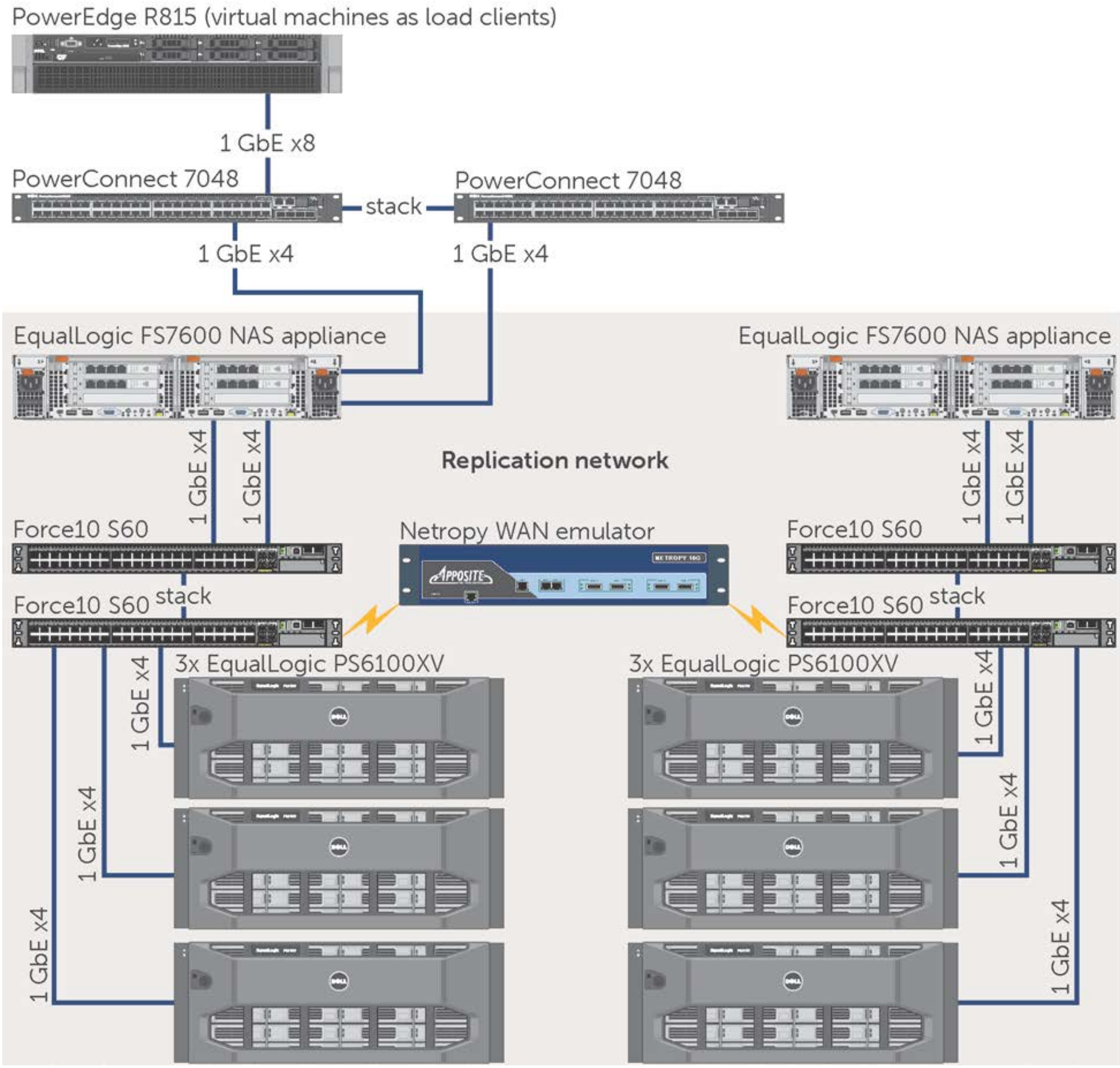

Figure 3 Replication test architecture

5 Replication performance characterization

This effort studied the replication performance of EqualLogic FS7600 NAS appliances systematically through carefully structured tests. Best practice recommendations are provided based on the insights gained from analyzing the test results. More specifically, the impact of the following factors on replication performance was studied.

- Available bandwidth for replication between sites
- FluidFS cluster size
- Average file size
- Presence of WAN maladies such as packet loss and latency

5.1 Dataset characteristics

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. Test data of varying file sizes (from 8K to 1G) were generated to simulate the data characteristics of a real NAS deployment. Vdbench is an open source tool and can be downloaded at http://vdbench.sourceforge.net/

The primary FS7600 was configured with three NAS containers. One volume was dedicated to small files, another with large files, and the third volume was loaded with a mix of small and large files.

NAS Container-1: Small files (from 4 KB to 500 KB) represented Microsoft® Office type files. The total size of this file system was around 250 GB.

NAS Container-2: Large files (> 512 KB) represented audio and video type files. The total size of this file system was around 500 GB.

NAS Container-3: A combination of small and large files. The total size of this file system was around 1 TB.

Using vdbench, a workload of 512 KB files with random I/Os of 70% reads and 30% writes was used for evaluating the performance impact on production clients during the replication.

6 FS7600 replication test results and analysis

This section describes the results and analysis of the replication tests run on the Equallogic FS7600 NAS appliances. The following tests were executed with the above setup and data sets, to measure the impact of various factors that impact the FS7600 replication.

6.1 Impact of intersite bandwidth

This test was run to demonstrate that FluidFS reasonably uses all the bandwidth available for replication. FluidFS replication is architected for high performance and is capable of consuming the available bandwidth for replication. In order to test this, a Netropy 10G WAN emulator from Apposite Technologies was used. The WAN emulator enabled simulating the commonly available WAN speed options by artificially throttling the network speed. This test measured the replication time to replicate 100GB of data from a single NAS appliance under different WAN link speeds.

Table 3 Tested WAN connection speeds

Connection Type	Max Available Bandwidth	Common Deployment Scenario's
10Gb Ethernet	10Gbps	Replicating across Campus Networks
1Gb Ethernet	1 Gbps	Replicating across Campus Networks
OC3	155Mbps	Replicating across city
T3	43 Mbps	Replicating across metro area
T1	1.5 Mbps	Replicating across a continent

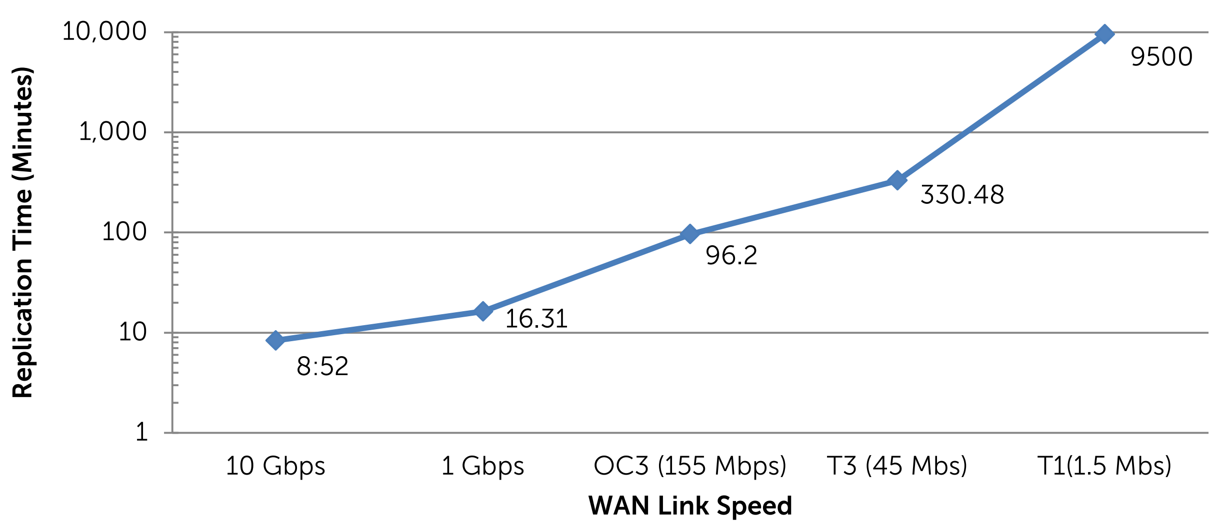

Figure 4 Time to complete 100 GB replication using a single FS7600 NAS appliance

Notice that in Figure 4, when a full 10 Gb bandwidth is available, replication did not consume the bandwidth entirely. This is because of the limitation of the test configuration. Achieving a sustained replication performance of 10 Gb requires more than the single NAS appliance configured in this test. Regardless, replicating over a 10 Gb link is less common when compared to the other connection options explored.

FluidFS replication is architected for high performance and is capable of performing at near line speeds for all of the commonly available connection options. The replication throughput is limited only by the available bandwidth.

6.2 FluidFS replication performance scalability

This test explores the scalability of FluidFS replication performance. A single NAS container of 500 GB (populated with 1MB files) was used to replicate from the source FS7600 cluster to the destination cluster. The replication performance was measured in MB/s using a range of one to four NAS controllers from the same NAS container. FluidFS scale out NAS architecture provides flexibility to scale performance and capacity independently. Additionally, the unique architecture enables a global namespace that can scale up to 509 TB in capacity. The flexibility that the architecture delivers for production I/O carries over to replication performance as well. As mentioned earlier, FluidFS replication utilizes two network interfaces per NAS appliance. By adding another NAS appliance to the cluster, replication performance can be enhanced. However, note that the source and destination NAS clusters should be identical in terms of cluster size.

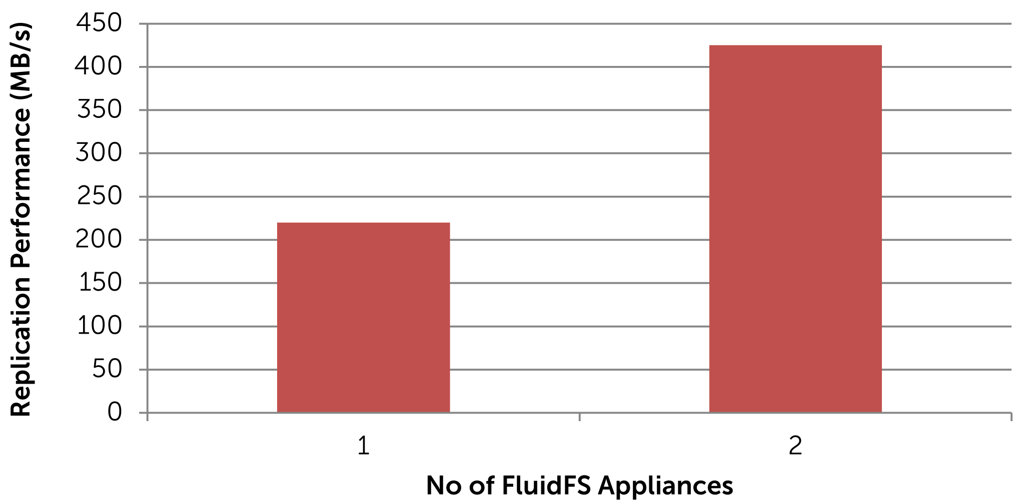

Figure 5 FluidFS scalability

FluidFS replication performance scales linearly with the number of appliances in the NAS cluster enabling a pay as you go methodology for increasing performance.

6.3 Average file size impact on replication performance

The average size of the files stored on a NAS system is one of the factors that influence the replication performance. This test explores the impact of file size on replication performance. For this test, data with average file sizes of 4k, 8k, 16k, 32k, 64k, 128k, 256k, 512k and 1MB on one single container was replicated and its impact on replication performance was analyzed. This test replicated over a WAN link that simulated a 1 Gb connection (max 125MBps) using a single FS7600 controller.

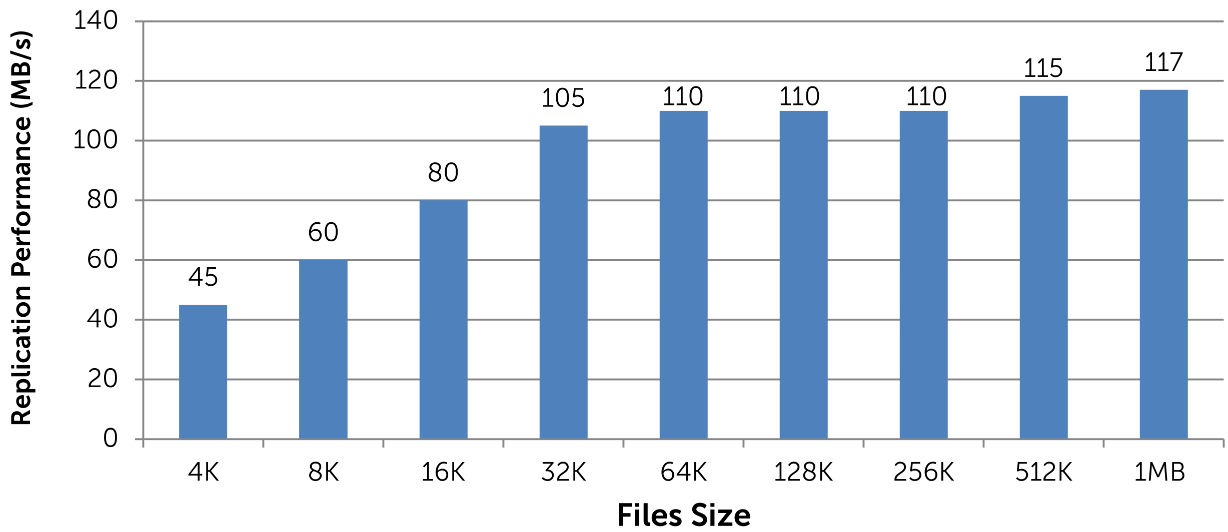

Figure 6 Effect of file size on replication performance

Notice that for 32K and larger sized files, near line rate replication performance was observed. This is because, unlike production I/O which is affected by the randomness of the access, replication processes access data in a more sequential manner. The number of NAS containers had no effect on replication performance, as the WAN link was saturated even with a single container as shown above. However, it is possible to create multiple NAS containers to store data of different file types/sizes. This allows appropriate replication schedules to be set on each individual container and meet different RPO business requirements.

FluidFS delivers near line rate replication throughput for average file sizes of 32K or larger.

6.4 Impact of WAN maladies on replication performance

In practice, most replication deployments utilize a WAN link between the primary and secondary sites. WAN links are different when compared to LAN links in certain attributes. It is important to understand the peculiar characteristics of a WAN link and how it affects replication in order to design effective disaster recovery architecture. It is important to understand the following attributes of a WAN link.

- **Network latency** is the time it takes data to go from the source to the destination and back. WAN latency is dependent on the distance travelled between the two network endpoints. Routers and other network devices also add latency. Typical LAN latency is approximately 5 ms or less and in WAN connections latency can range from 50 to 500 ms depending on the number of factors.
- **Packet loss** occurs at the network layer due to congestion in routers, link failures and other network device problems. An average of 0.1 to 1% packet loss on a WAN network is common in many networking environments, 5% packet loss can significantly increase the duration of the replication and ultimately affect the RPO and RTO.
- **Bandwidth congestion** occurs at several points in the WAN link. If multiple applications are using it, the network resources become over utilized and can restrict the replication throughput. It is essential to understand the current application load on the existing WAN link before implementing the FS7600 replication.

These attributes of a WAN link can significantly degrade the replication performance, and should be accounted for when defining the RPO and RTO of the enterprises. It is vital that system designers understand the impact of the common WAN maladies and the connection between the replicating sites in order to design an effective replication solution.

In this section, the impact of packet loss and latency characteristics of the inter-site connection being used for replication on replication performance is analyzed.

6.5 Impact of Packet Loss

Replication is most commonly deployed on TCP/IP connections due to the pervasiveness of the protocol and lower cost compared to other options such as dark fiber/ATM links. When TCP/IP packets are dropped or unacknowledged, the TCP/IP protocol invokes the slow start algorithm (see RFC 56812). Slow start is part of a normal congestion control strategy to learn the transfer capacity of the link and the receiving end point. In the presence of a lossy link, the slow start algorithm is invoked too often thereby affecting the network throughput. The packet loss characteristics of the most common connection options were studied as part of this effort. The vdbench I/O simulation tool was used to generate 1 MB files to a total 60 GB of data on a single container.

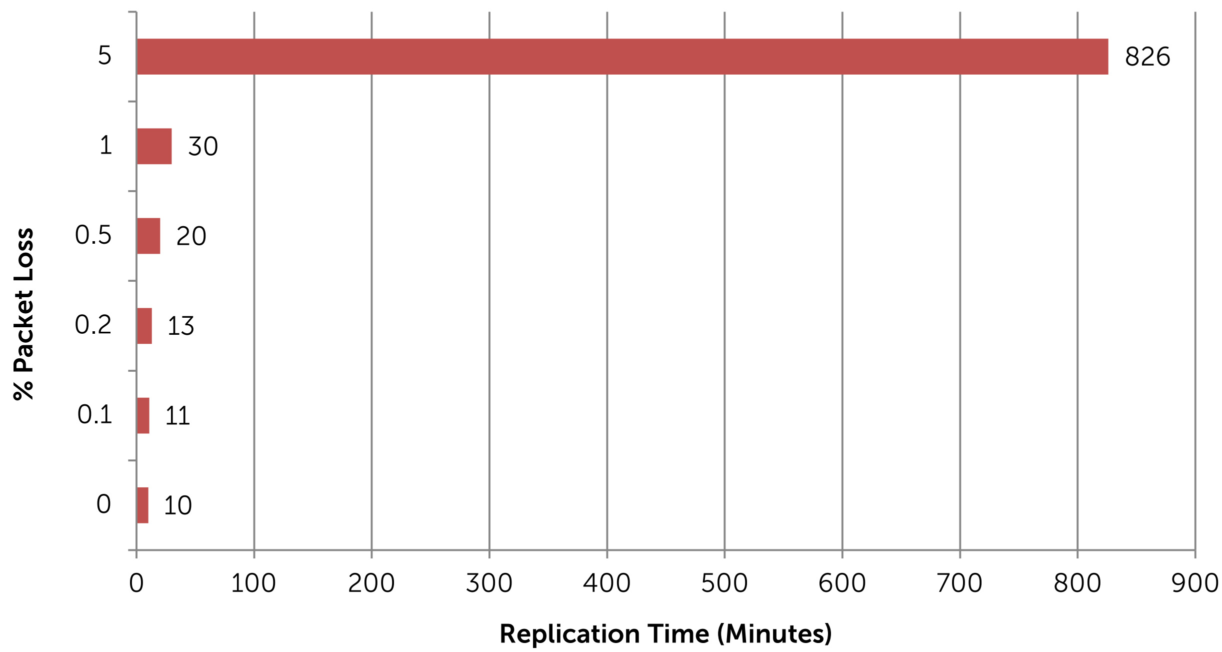

Figure 7 Packet loss effect on replication across a 1 Gbps link (60GB changed data)

As shown in the graph above, packet loss can significantly reduce the effective bandwidth of a connection and limit the effectiveness of the solution.

The testing concluded that the addition of a small amount of random packet loss (1/10th of a percentage) caused only a slight variation in the time it took to replicate 60 GB of modified data. However, when one percent of random packet loss was added, replication time tripled. When the packet loss was five percent (for example in a satellite connection), the replication time was significantly higher resulting in unacceptable RPO. Because packet loss can have a significant effect on replication time, it is important to monitor the quality of the link between sites.

It is very important to measure the effective throughput capability of the replication link in designing RPO and RTO expectations. As the replication link incurs significant packet loss, this has an adverse effect on the RPO and RTO.

6.6 Impact of Latency

Latency refers to delays that are incurred while processing network data. A low latency network connection is one that generally experiences small delay times, while a high latency connection suffers from long delays. Network tools such as ping tests and traceroute measure latency by determining the time it takes a given network packet to travel from source to destination and back. This is called the round trip time.

The round trip time for a data packet exchange will take longer in relation to the distance separating the devices. Another factor that affects round trip time is the number and type of network switching and routing devices that the packet encounters along the network path. It is more common for the device latency to cause an increase in the overall link latency rather than the distance alone.

It is not uncommon for replication to occur across continents as part of a comprehensive disaster tolerance strategy. Even though optics based networking technologies have made impressive strides, the fact that data has to traverse through several routers and switches within the internet backbone increases latency. Just as packet loss does, latency also has a limiting effect on the bandwidth a connection link can deliver.

A WAN emulator was used in testing to inject various amounts of latency into the replication network links. A NAS volume was then replicated across the 1 Gbps link and the time it took to complete the replication of 60 GB changed data was measured. The vdbench I/O simulation tool was used to generate 1 MB files to a total 60 GB of data on a single container

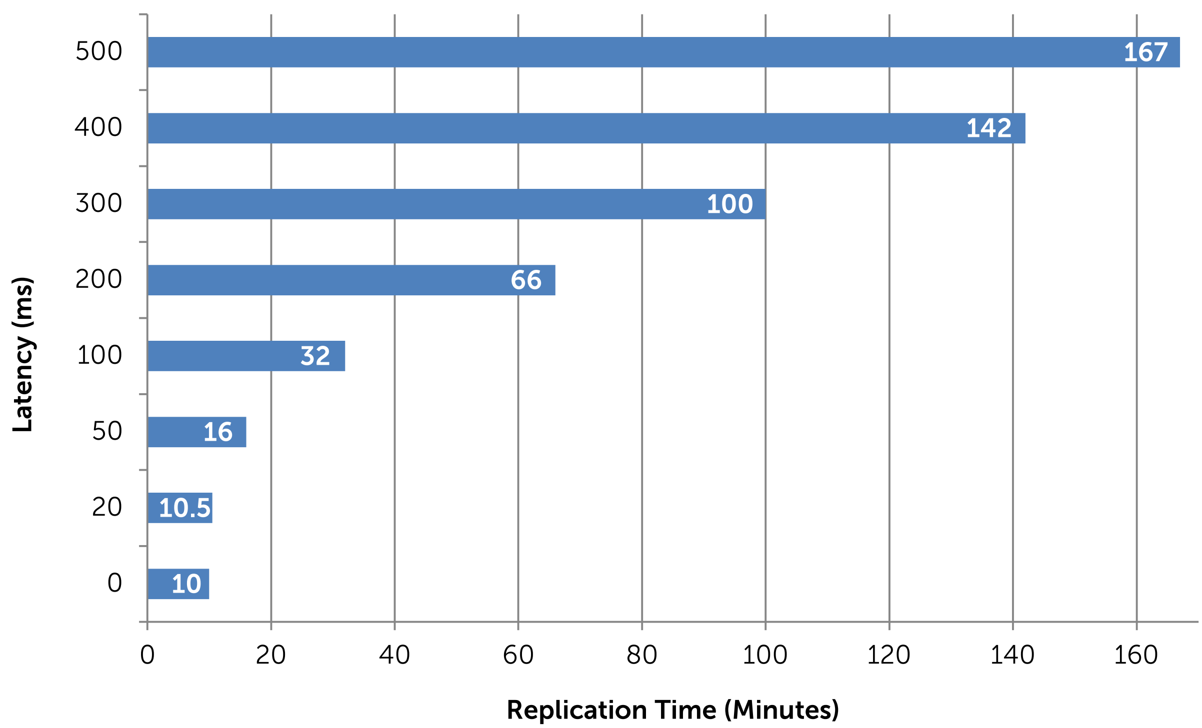

Figure 8 Effect of latency on replication across a 1 Gbps link (60GB of changed data)

As shown in Figure 8, network latency can significantly reduce the effective bandwidth of a connection thereby limiting the effectiveness of the solution. When 100ms latency was added to the replication link, the replication time was increased by a factor of 3 and when the latency was 500ms, the replication time was increased by a factor of 16.

It is very important to measure the effective throughput capability of the replication link in designing RPO and RTO expectations. The latency on the replication link has an adverse effect on the RPO and RTO and Dell recommends using a low latency link to achieve optimal replication performance.

7 FS7600 replication planning and design best practices

This section presents the variables and configurations that need to be taken into consideration when planning and implementing an efficient replication solution.

7.1 Designing achievable RPO/RTO

The primary goal of a replication solution is to meet the RPO and RTO requirements. These requirements dictate how often the replicas need to be synchronized. Understanding the data transfer capacity of the solution as it pertains to replication is important to determine if the solution meets the business RPO and RTO requirements.

7.1.1 Determining the replication capacity of the connection for the initial sync

The example below demonstrates a simple methodology to determine the capacity of the replication solution. As discussed in the previous sections of this paper, several factors affect the effective bandwidth available on a connection link. For instance, TCP/IP protocol carries protocol overhead which reduces the available bandwidth from the theoretical value. The methodology mentioned below can be used for a rough estimation of the replication capacity.

Maximum Throughput in MB/sec = [WAN link speed (Mb/sec)/ 8 bits per byte] x 93% protocol efficiency

Using the estimate for throughput, replication time for a given amount of data can be calculated as:

Replication Time = Data set Size (MB)/Throughput

Consider an OC3 link, which is rated at 155Mbps.

Maximum Throughput in MB/sec = 18.02 MB/sec

Time to transmit a NAS Container with 100GB data set = 102400/18.02 = 5683 sec (~95 minutes)

In practice, appropriate adjustments need to be made in determining the network throughput to account for other factors such as network latency and packet loss.

Monitor the changed data at the source for calculating the actual bandwidth required to meet the selected RPO.

7.1.2 Determining the bandwidth required for replication

Assume there is a **2 TB** NAS container with a **5% change** (or new data), i.e. 100 GB per day or **~4 GB/hour**. If the RPO is one hour, then the replication of 4GB data needs to be completed in an hour which translates to a theoretical bandwidth of **11 Mbps**. With a latency of 10%, the actual bandwidth required for replication in this example is **12.1 Mbps**. Also note that the change rate is not uniform throughout the day, use the peak data change rate for the calculations.

7.2 Tuning the replication link

Data flowing through a replication WAN link will most likely travel through the internet backbone comprised of several routers. Congestion in the WAN link can cause dropped packets which in turn affects the bandwidth available for replication. System administrators can implement several measures to mitigate the effects of these uncontrollable factors.

- When possible, consider implementing a WAN link dedicated to replication.
- Consider implementing QoS (Quality of Service) which provides a pre-negotiated level of network bandwidth.
- Latency adds up time to complete the replication, so a low latency network is recommended to achieve optimal replication performance.
- Consider implementing WAN optimization products. WAN optimization solutions mitigate the effects of WAN maladies through network based compression, de-duplication and other packet optimizations that improve the efficiency of a slower WAN link.

7.3 Planning storage needs and replication schedules

NAS replication makes use of the snapshot capability within FluidFS. As such, during the replication, the snapshot reserve space should be closely monitored and adjusted to ensure that there is enough reserve space capacity.

Snapshots occupy space on the NAS volume. To use snapshot space efficiently, administrators can configure alerts that respond when a predefined percentage of volume capacity has been consumed by snapshots. The total capacity consumed by snapshots (snap reserve) for a volume at a given point in time is the sum of the snapshot delta size for all the snapshots on the volume. The snapshot delta size increases dynamically, based on the update rate of the active file system. Only existing blocks that are updated in the active file system are accounted for in the snapshot delta size or the overall snapshot reserve. New data blocks added to the active file system do not impact the space consumption of existing snapshots.

Actual user data, estimated growth, change rate and retention requirements are needed to plan the size of a NAS container. Each time the replication process is run, a new snapshot is created on the source system that is used as a baseline to copy the changed data to the destination cluster.

Use the FluidFS CLI to monitor snapshot size and delta size.

```
#ssh admin@<FluidFS-Management-IP> "data-protection snapshots actions view-delta
<VolumeName> <SnapshotName>".
```

A 1 TB NAS container was set up with 500 GB of data for these tests. A 2 % growth of new file data was simulated every day and 1% of the data changed daily. The following calculations were used to measure the required snap reserve space on the NAS for the replication. This does not include the additional space required for snapshots created outside of the replication process. It is possible for these snapshots to consume the snap reserve space based on the number and retention of the snapshots.

Current space utilization = FS7600 with 1 TB NAS container containing 500 GB of file data.

Daily growth in user data (assume 2%) = Growth rate x current space utilization = 10 GB.

Capacity Consumed by Snapshots for replication = Change rate x (current space utilization + daily growth in the user data) x the number of snapshots retained = 15.3 GB. With a 1% change rate between successive replications, this equates to 5.1 GB per snapshot (1% of [500GB + 10 GB]). Replication retains three snapshots. Total capacity used by snapshots is 15.3 GB (5.1 GB per snapshot x 3 snapshots).

In the above example, the estimated total NAS reserve space required for replication is 15.3 GB.

As there is some overhead associated with each replication pulse, it is strongly recommended to schedule replication pulses so that replication backlog is not accumulated for any NAS volume. If there's an I/O impact to the production clients, consider replicating during non-peak hours.

Ensure RPOs are aligned with the business needs. Not all data is equally important, so the frequency of replication can vary among different NAS containers. Dell recommends scheduling the replication of NAS containers according to the business needs to meet certain RPOs.

7.4 Planning for initial synchronization

Although it would be ideal to implement a replication solution from day one, in practice that is not always possible. Replication solutions can be implemented as business requirements grow. Effectiveness of a replication solution is highly dependent on the bandwidth available on the replication link. Typically, the replication bandwidth is procured based on the expected change rate of the data that needs to be protected. As the amount of unstructured data that organizations manage grows, completing the initial synch can be a slow process on slow WAN links.

One option to consider is to connect the secondary storage system in the local (primary) data center, establish replication partnership, and perform a full initial synchronization over a LAN link. After the initial replication is complete, shutdown the secondary storage system and move it to the remote site. After the secondary storage system is setup in the DR site, further replication sessions will synchronize only the changes made since the initial replication.

Another option is deploying WAN accelerators or optimizers to improve the efficiency of replicating large amounts of data over a WAN link.

7.5 Monitor Replication Status

It is important to monitor the replication status and its performance to achieve the required RPO. The storage administrators can monitor the success and failure of replication sessions using the EQL Group Manager GUI by monitoring the replication history. Set an alert (SNMP) for replication failures and look at the event log for root cause analysis. For details see the Dell EqualLogic Group Manager Administrator's Manual on www.support.equallogic.com.

8 Best practices and recommendations

The following list summarizes the best practices and recommendations mentioned in this paper that will enable optimal performance of the EqualLogic FS7600 NAS cluster replication.

- The source and destination FS7600 NAS clusters should be identical in terms of cluster size (number of appliances and controllers). However, the underlying EqualLogic storage at both sites can be different.
- When possible, consider running the same FS series firmware version at both the source and destination NAS clusters. Replication from one FluidFS cluster running a newer version to another FluidFS cluster running an older version is not supported.
- The FS7600 replication engine fully utilizes the bandwidth available on the WAN link. To achieve optimal throughput and scale the replication performance, data should be load balanced across the file system domains (FSDs). This enables load balancing the requests and allows the replication engine to read from all the controllers available in parallel to maximize the replication throughput.
- NAS replication makes use of the snapshot capability within FluidFS. As such, during the replication, the snapshot reserve space should be closely monitored to ensure the availability of enough reserve space.
- The number of containers has no effect on the overall throughput of the replication.
- The FS7600 replication performance scales linearly as more controllers (NAS appliances) are added to the NAS cluster. Each controller can start independent replication sessions.
- Evaluate the impact of current workloads due to the replication background process and schedule the replication process appropriately.
- For replication throughput greater than 120 MB/s, use a network link greater than 1 Gbps between the source and remote cluster.
- When possible, consider implementing a WAN link dedicated to replication.
- Monitor the replication performance and history using the EQL GUI. Set SNMP or SMTP alerts for replication failures.
- Ensure RPOs are aligned with the business needs. Not all data is equally important, so the frequency of replication can vary among different NAS containers. Schedule the replication of NAS containers according to the business needs to meet certain RPOs.
- Loss of packets in a WAN link will add it to the time it takes to complete replication and may jeopardize the required RPO/RTO. Monitor the WAN link for any packet losses greater than 0.5%.
- Latency adds to the time it takes to complete replication. Use a low latency network for FS7600 replication.

9 Conclusion

The Dell EqualLogic FS76X0 NAS appliance is a compelling choice that meets the challenges posed by the growth of unstructured data. The underlying Fluid File System delivers advanced capabilities such as a high performance virtualized scale out file system and a rich feature set. Performance and efficiency were key principles in designing this file system in which the replication feature also benefits. Furthermore, the EqualLogic licensing model carries over to the NAS offering, which means that functionality such as snapshots and replication are included at no additional cost.

FluidFS snapshot based replication solution offers information protection and disaster recovery. Disaster Recovery can be vastly enhanced by the use of FluidFS replication. Data on the replicated system can be kept up to date as desired depending upon the amount of data being modified on an ongoing basis and the speed of the network link between the two FluidFS systems.

A Test Infrastructure

Table 4 Component design details

Solution Components		Purpose
Servers	Dell PowerEdge R815 Server running ESXi 5.0.0 hosting Linux and Windows VMs	To generate client I/O.
Network	4 x Force10 S60 : Firmware 8.3.3.7 2 x PowerConnect 7048R: 5.0.1.3	Connectivity to the SAN and replication Connectivity to the client network.
Storage	3 x EQL 6100XV storage arrays at each site with) firmware V6.0.2 Dell EqualLogic FS7600 (1 Source +1 Destination Fluid FS Firmware 2.0.6220	Storage with enough space for creation of file systems. One FS7600 NAS appliance at each site for testing remote replication.
Test Tools	test tools Vdbench to create file system data sets	To simulate CIFS and NFS file traffic
Performance Monitoring	Windows Perfmon , Linux sar, Ibmonitor tools.	Performance monitoring in Microsoft Window, Linux, SANHQ
SAN Headquarters Server	EQL SAN HQ 2.2	Monitor PS array metrics.

A.1 EqualLogic storage array configuration

Three EqualLogic PS6100XV arrays were used as a back end for the FS7600 appliance at each site. Two dedicated Dell Force10 S60 switches were used for SAN connectivity. The front end client connections were made to a separate a pair of PowerConnect 7048 switches. The two sites were connected by a 10 G link and used the Netropy 10G WAN emulator to emulate different link speeds.

A.2 Network configuration

In this architecture, there were three different networks. Each FS7600 controller had two quad port 1 GbE NICs. These ports were dedicated for specific networks:

Client Network: Used for accessing NFS exports and CIFS shares hosted by the NAS server. There were four dedicated 1 GbE ports on each controller (eth0, eth1, eth2, eth3 – bonded as one logical interface)

Storage network: Used for communicating between FS7600 controllers and the EqualLogic array members (eth30, eth31, eth32, and eth33).

Replication network: eth30 on FS7600 was also used for establishing the replication session with its partner cluster and used for transferring data during the replication.

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 EqualLogic Configuration Guide:
http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19852516/download.aspx

Documents pertaining to FS7600 replication details are located at https://eqlsupport.dell.com (this site requires an account to log in):

- Dell EqualLogic Group Manager Administrator's Manual
- Dell EqualLogic FS Series Disaster Recovery by Utilizing FluidFS NAS replication.
- Dell EqualLogic FS Series Installation and Setup Guide

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

- http://dell.to/sM4hJT