Increasing Throughput for Data-Intensive, High-Performance Computing (HPC) Environments for Microsoft® Windows® HPC Clusters

HPC Solutions Engineering

By

Sanjay Kumar

Dell Product Group
August 2009

Table of Contents

Executive Summary

This document describes high-performance solution architecture in a data intensive HPC environment with Windows HPC Server® 2008, Dell™ Power Edge™ servers, Quantum StorNext® file system, and Dell EqualLogic™ iSCSI storage. This paper provides a reference configuration to be used to evaluate and select the Dell solution that best fits individual requirements, based on the performance needed with the appropriate storage solution that delivers high-aggregate throughput. However, this solution can also be based on other Dell servers and Dell high-performance storage to meet needed requirements.

Additionally, this paper will describe the reference configuration, performance results, and best practices for the Dell solution including data management capabilities. It will not contain information regarding the installation and deployment of the different components such as Windows HPC Server 2008, and the Dell server and storage. This paper is not intended for performance evaluation for either the storage or file system, but instead as guideline to be used to select a similar configuration.

Introduction

Microsoft Windows HPC Server 2008 provides enterprise-class tools, performance, and scalability for a highly productive, compute intensive environment. It provides an integrated cluster environment that includes the operating system, a job scheduler, message passing interface v2 (MPI2) support, and cluster management and monitoring components. Windows HPC Server 2008 is composed of a cluster of servers that includes a head node, and one or more compute nodes for the required computational power. The head node generally controls and mediates all access to the cluster resources, and is the single point of management, deployment, and job scheduling for the compute cluster. A simple Windows 2008 HPC cluster is represented as in Figure 1.

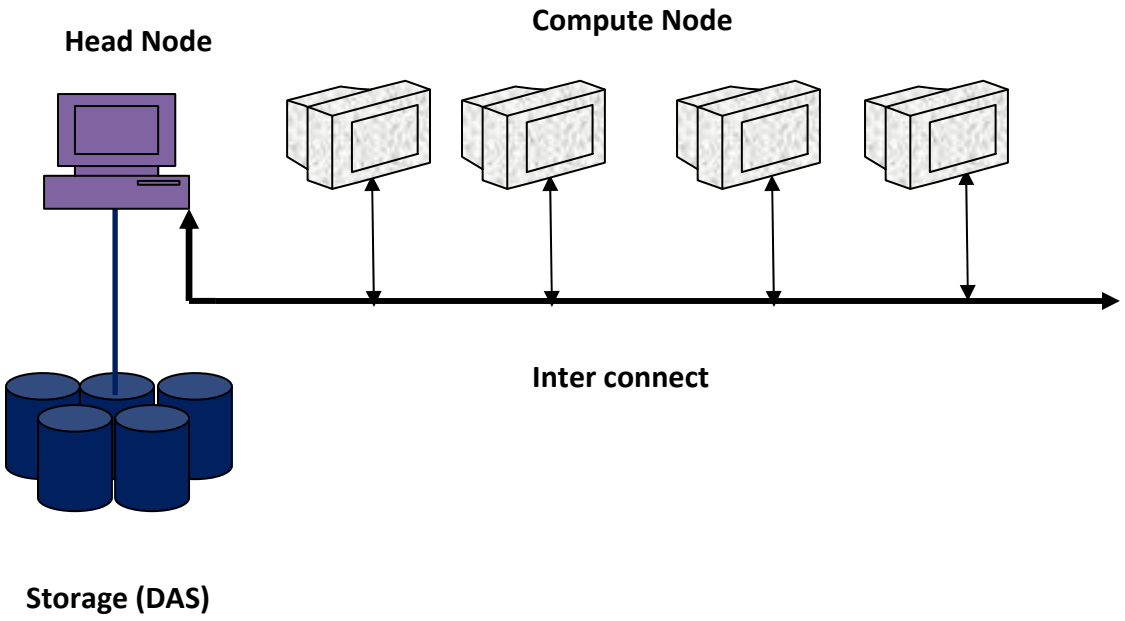

Figure 1: Simple Windows HPC Cluster

More power can be added easily, based on computational requirements, with multi-core and faster -available servers. Apart from being compute intensive, today's HPC environments are becoming data intensive as more data is generated from models and analysis. With data-intensive HPC environments, solutions should deliver a high-aggregate throughput with the help of a high performance file system and attached storage. As the amount of generated data increases, additional data management software will play a vital role for the HPC solution.

Data-Intensive HPC Environments

With data-intensive HPC environments, high-performance I/O solutions and data management are key factors along with the compute power. Depending on compute power and the amount of data generated, Windows HPC clusters may vary in size and throughput requirements; for large cluster sizes, storage, file systems, manageability, and data management requirements will vary. The most important decision to be made when implementing a data-intensive HPC environment, is the selection of the "best fit" storage with the "best-fit" file systems based on *performance, capacity, accessibility, cost and high availability needs*.

- ***Performance*:** The size of the cluster and hence the number of concurrent applications running will vary based on the throughput, accessing the storage directly or using storage nodes. The file system and storage should deliver the required aggregate throughput with the ability to scale with the storage hardware, storage nodes and compute nodes directly accessing the storage. The required aggregate throughput can be estimated based on the required throughput per node.

- ***Capacity:*** Online storage capacity planning is partially based upon the performance requirements. Near-line storage can be planned considering cost and where offline data that has not been accessed recently can be pushed with the help of hierarchical storage management (HSM) features.

- ***Accessibility*:** Define the storage accessibility based upon the needs of the applications. Some applications may not require that all compute nodes access the storage directly, but instead with the help of the storage nodes. At the same time, some applications may require that the compute nodes directly access the storage. File system software should be able to handle both situations.

- ***Cost and High Availability needs*:** Select the storage and file system solution considering high availability needs. I/O transactions should not fail because of a storage, storage node, or metadata server failure in case it is centralized. Generally, storage hardware has this feature built-in and this is taken care of by storage vendor itself. High availability should be present at storage node and metadata node, if there is a centralized metadata server, at the file system level. In the case of distributed metadata, this issue is taken care by storage nodes itself.

Dell solution for data intensive HPC

In order to meet the above criterion, Dell offers scalable storage, and high-performance cluster file system solutions, for Windows HPC Server 2008. With these solutions, mixed compute node configurations are possible, where the compute nodes can access the storage directly or with

the storage nodes based upon accessibility need. The compute nodes can access the data in parallel with multiple storage nodes. Storage accessibility from the compute nodes can be set for either direct access to storage or using storage nodes. High availability features are available at the storage node and metadata node level. Dell high performance SAN storage and Quantum StorNext file system together provide high performance and scalable data storage solution.

Overview of Quantum StorNext File system

StorNext is a heterogeneous shared file system that enables multiple servers to access a common disk repository, regardless of operating system (OS) type supported by StorNext. Access to the file system is controlled by metadata controllers (MDC). The MDC is a server that sits outside the file system data path, and is responsible for negotiating access and indicating the client buffering mode. Nodes running StorNext clients communicate with the MDC using an IP connection to obtain information about file location, block allocation, and other information that provides direct, block-level access to the disk.

Clients can either be LAN based, a distributed LAN client (DLC), or SAN based. With SAN based access, clients communicate directly with the storage. With LAN based access, DLCs communicate to the storage using SAN clients designated as gateway servers for data access. StorNext file system has the following components:

- Metadata controller (MDC): a separate machine that handles all client metadata transactions. For high availability, MDC supports failover.
- Distributed LAN client (DLC): these clients access the data over IP using either single or multiple gateway servers.
- SAN client: directly attached to the SAN, and can access data directly at storage speed.
- Gateway server: these send the I/O request from the LAN connected clients to the SAN. Gateway server acts as a clustered gateway that provides high-performance failover and load balancing.

Overview of Dell HPC Windows 2008 Solution with Quantum's StorNext File System

With faster multi-core servers and high-capacity storage systems currently in use, today's HPC applications are generating greater amounts of data. With data intensive HPC environments, applications require high performance and scalable I/O subsystems. I/O subsystems, such as storage and the file system, should be able to meet the application's requirements in terms of throughput, as well as in capacity and scalability. In addition, today's HPC applications require the following:

- Performance: a high performance storage and file system that can deliver high-aggregate throughput
- Accessibility
- High availability
- Data management

Following sections in this document propose a solution that addresses the requirements defined above. This solution also offers the flexibility to access the storage either using DLC's, or directly using the SAN. The proposed solution consists of following:

- Head node
- Compute node with either a SAN client, or distributed LAN client capability to access storage
- Metadata controller that consists of a single server without HA mode, and two servers utilizing HA capability
- DLC gateway server that consists of multiple servers, based upon application needs

Each of the compute nodes will be configured as StorNext DLCs. In case compute nodes need to access the storage directly, it can be configured as SAN client with proper storage connectivity. Figure 2 exhibits the Dell Solution for Windows HPC 2008.

In the figure, **A** represents a compute node configuration that can access the storage using the gateway server. **B** represents the DLC gateway server that provides a way for each compute node to access the storage. **C** displays the compute node configured so that it can access the storage directly. **D** represents the MDC. With high-availability, two MDC servers are required to run the StorNext failover (FO) option.

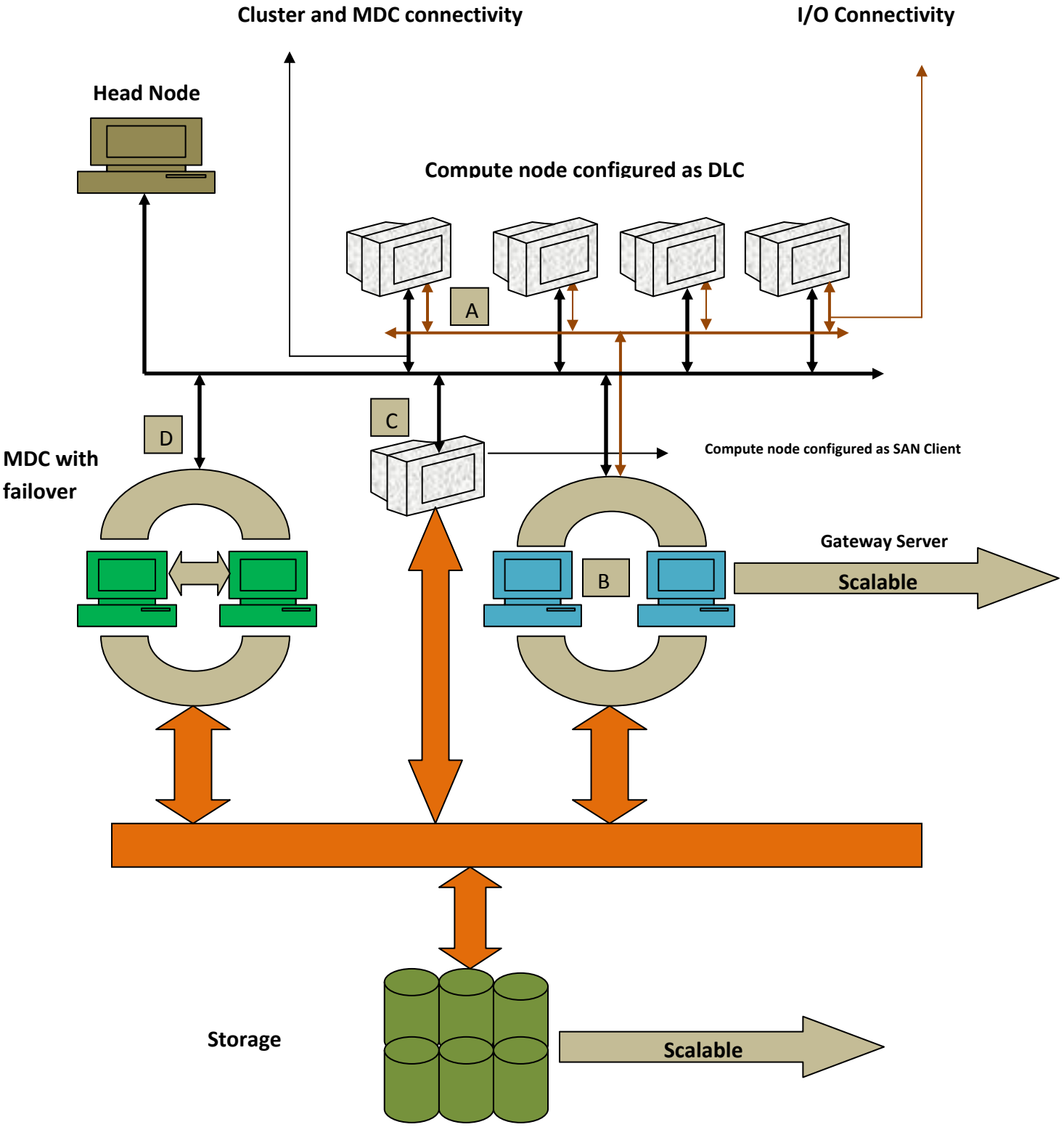

Figure 2: Dell Windows HPC Cluster Solution with Quantum's StorNext File System

Reference Configuration Overview

A reference configuration has been evaluated using Dell EqualLogic iSCSI storage. Figure 3 shows the reference configuration setup that was used for testing.

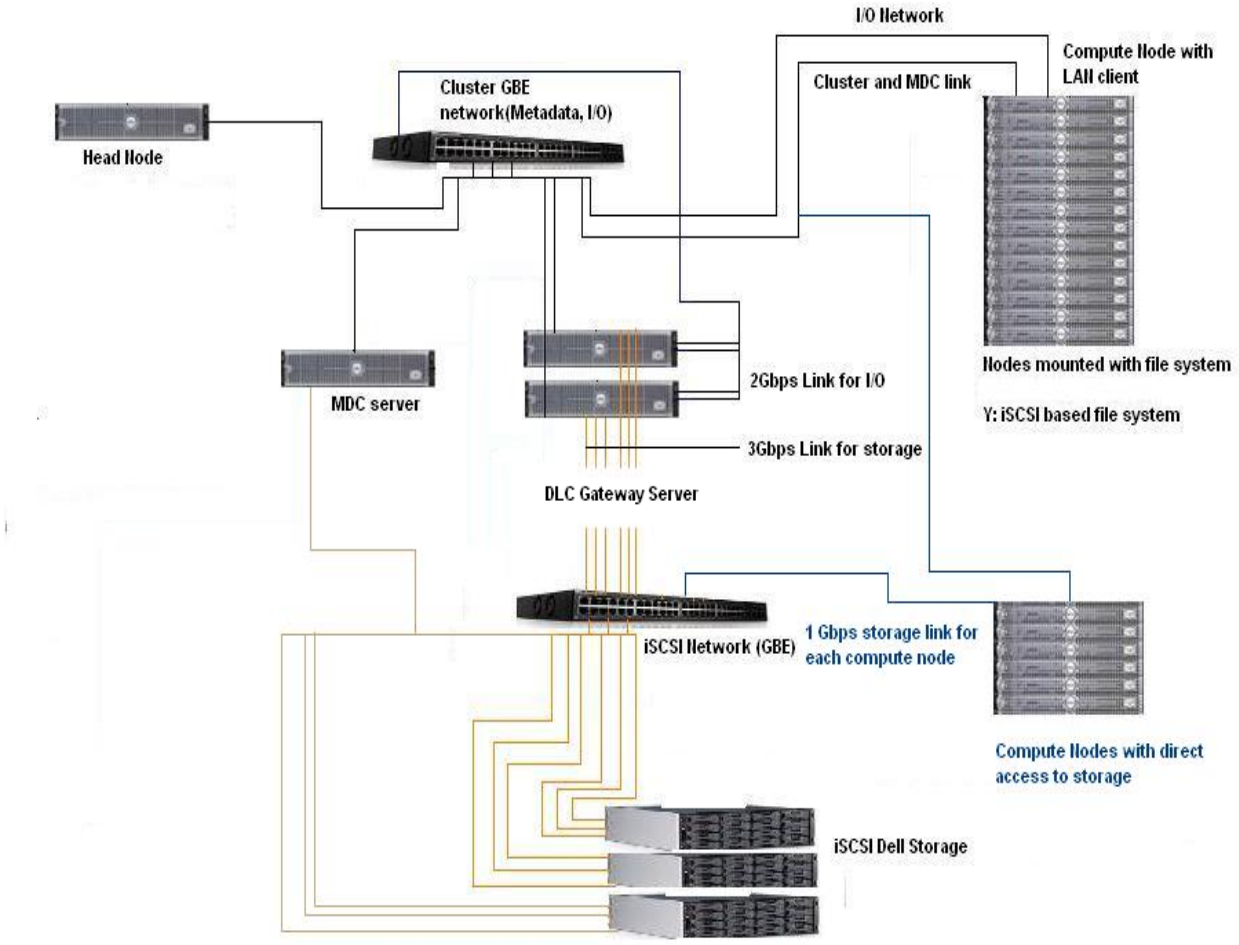

Figure 3: Reference Configuration Setup for Windows HPC Cluster using Dell Servers, EqualLogic Storage, and Quantum's StorNext File System

High availability was not used for the MDC server in this configuration; however two DLC gateway servers provide high availability with load balancing for data access. Three Dell EqualLogic PS Series arrays were configured in a non-HA mode, with three GBE iSCSI links for the respective storage array. EqualLogic iSCSI software on DLC gateway server automatically

load balance across three iSCSI links. DLC gateway servers were configured with 2Gbps network for I/O. Similarly, each compute node was configured with 2Gbps for DLC network I/O distribution to match with number of network on DLC gateway servers. StorNext automatically load balance across NICs and gateway server. The reference diagram as shown in figure 3 also shows that the compute node was configured for direct access to storage with a 1Gbps network link that was accessing the same file system. A file system was created and mounted on each of the compute nodes.

Reference Configuration Hardware and Software

The following hardware and software was used in the reference configuration.

Hardware	Component	Description
	Metadata controller (MDC)	1x Dell PE 1950
	DLC gateway server	2 x Dell PE 1950
	Head node	1 x Dell PE 1950
	Compute nodes	4 Dell PE servers
	Storage	3 Dell EqualLogic PS series array
	Network switch	2 x Dell PowerConnect™ 6248
Software		
	OS	Windows Server 2008 HPC edition
	File System	StorNext
	Host software	Dell EqualLogic bundled software (HIT)
Benchmark		
	IOZONE	IOZONE build with Cygwin

Dell EqualLogic PS Series Array Configuration

Three Dell EqualLogic PS Series arrays were used in the reference configuration. Each PS Series array has two control modules. Three network interface ports from one of the control modules for each array were connected to the DLC gateway server using a network switch. iSCSI software initiator was enabled for each iSCSI network interface on a DLC gateway server. Multiple iSCSI connections were leveraging the failover and load balancing features, with the help of the Dell EqualLogic software installed on the DLC gateway servers. The section below describes the storage configuration used.

- 3 Dell EqualLogic PS Series arrays
- 8 RAID5 volumes
- Storage cache size of 3 GB.
- SAS disk with 15K rpm.

- For each array, the MDC and DLC gateway servers were connected to a Dell PowerConnect 6248 switch forming an iSCSI network exclusively for I/O between the DLC gateway server and the array.

Network Configuration

Network configuration plays a vital role as each of the compute nodes configured as DLC's need to communicate to the Gateway Server as well as to Metadata Controller. Also, the head node needs to communicate to the compute nodes for management. Performance is limited by network bandwidth available on these nodes in case compute nodes are accessing via DLC Gateway Server. Three different networks were configured for the cluster: metadata and I/O by using IP subnet. A single Dell PowerConnect 6248 was used for the cluster, metadata, and I/O traffic. The following list items describe the three different networks using IP subnet mechanisms:

- Cluster network: One of the network interfaces, a 1Gpbs link, for each server (head node, MDC, gateway server, compute nodes) was configured as subnet (172.20.0.x) for cluster communication.

- Metadata network: Metadata and cluster traffic were shared. If cluster network is heavily utilized, there should be a dedicated metadata network.

- Distributed LAN network for I/O: Each DLC gateway server was configured with a different subnet (eth1:172.20.1.x and eth2:172.20.2.x) for I/O traffic. Each compute node was configured with two subnets (eth0:172.20.0.x and eth1:172.20.2.x). In order to leverage available bandwidth, one virtual interface was configured for eth0:172.20.1.x; one network was shared between metadata and I/O traffic.

- ISCSI network for storage: Each DLC gateway server was configured with a 3 Gbps link (3 x 1Gbps) for iSCSI traffic, and was connected to an Ethernet switch that was dedicated to ISCSI traffic. Each storage array was configured with a 3 Gbps (3 x 1Gbps) link on one control module, and connected to the network switch.

    - The ISCSI software initiator was enabled on one MDC server network interface that was participating in the iSCSI SAN. Multiple NICs with iSCSI software initiators should be enabled to leverage the failover and load balancing features.

    - Compute node with direct storage access was configured with a 1 Gbps iSCSI link. To use the failover and load balancing features, multiple NIC should be used.

File System Layout and Configuration

One 3 TB file system was created. This file system was mapped to each compute node as ”Y” drive. It was created and configured on the MDC with a “name server configuration”. Two DLC gateway servers were configured to handle I/O from the compute nodes. Each of the compute nodes accessed the file system using DLCs, except for some compute nodes that were accessing the storage directly using SAN Clients. MDC, DLCs for the compute nodes, and DLC gateway servers were setup and configured. The following list items describe the file system component configuration:

- MDC configuration: The StorNext file system was installed and configured on the metadata server. For simplicity, the name server service was also configured on the MDC but can be configured on any other server. Configuring the name server is mandatory as without, the DLC gateways, DLCs, and SAN clients cannot communicate with the MDC. The following basic steps were done for file system setup and creation.
    - All 8 volumes were labeled using disk labeler; disk labeler is the part of Quantum StorNext software.
    - FS name server configured on metadata server. IP address of MDC server was added in the name server configuration.
    - By using the file system configuration tool, three different stripe groups, namely metagroup, datagroup, and journalgroup, were created. Metagroup and journalgroup contained one RAID5 volumes exclusively for the metadata and journal. Datagroup contained six RAID5 volumes for data. The file system configuration was saved for the above configuration.
    - A file system of 3 TB size was created using 8 volumes, namely iSCSI-SNFS with the help of the advanced file system creation tool. The configurable parameters, such as block size and journal size, were set to the default values. For the 8 volumes, six of them were used for I/O while the other two were used for journal and metadata separately.
    - The volumes selected for I/O were tuned with a file system stripeBreadth of 256K. For metadata and journal, the stripeBreadth was the default of 64K.
    - The file system was created using the `cvmkfs` command. The `cvmkfs` command allows the file system and configuration file names as parameters.

- DLC gateway server configuration: The DLC gateway server was configured after the file system is up and running on the MDC. In order to configure the DLC gateway server, the following steps were performed:
    - Configured each DLC gateway server with the help of the client configuration tool. The server capability was enabled in the distributed LAN section.
    - Two network interface cards were selected for I/O, in order to handle the traffic from the compute node.
    - The file system was mapped to Y drive. The file system can be mapped to the some directory, however in this case it was mapped as drive.
- Compute node configuration as a distributed LAN client: Each of the compute nodes was configured as a distributed LAN client. These nodes were accessing the data using the gateway server. The following configuration was performed on each compute node.
    - The file system was mapped as "Y" drive locally on each of the nodes using the client configuration tool.
    - All nodes were enabled with client capability in distributed LAN tab.

Reference Configuration Performance Test

The performance test was performed to measure the I/O throughput on DLC gateways, as well as on the compute nodes configured as distributed LAN clients. Compute nodes were accessing the files using the DLC gateway. For each test, a sequential work load and large file sets were used. Multiple IOZONE threads were executed in parallel on both on DLC gateway servers and compute nodes at different times. The intent of the performance test was to stress the file system with a large file set, and observe the throughput on the gateways and compute nodes. I/O throughput on the gateway servers will suggest the maximum available throughput, and provide a guideline for a similar configuration. Tests performed on the compute nodes will suggest the throughput available on those nodes with a similar configuration. Throughput on the compute nodes was limited by network bandwidth and the gateway servers. Throughput on the DLC gateway server was limited by the 3 Gbps Ethernet link on each of them. For performance testing, the file system benchmark tool IOZONE was used.

Results and analysis

For iSCSI storage, the theoretical I/O bandwidth was limited to 375 MB/s (3 x 1Gbps), on each of the DLC gateway servers. The total storage bandwidth available with two gateway servers

was 750 MB/s, but the total available bandwidth will be less if TCP and other overhead were considered.

The aggregate throughput for the compute nodes was limited by the 2 Gbps network on each of the gateway servers. Each DLC gateway server can handle a theoretical maximum of 250 MB/s with the reference configuration; throughput was limited to 500 MB/s (250MB/s * 2 gateway servers). Each compute node had one 1 Gbps link shared for metadata and I/O traffic. This shared link could produce some overhead for aggregate throughput on compute nodes when metadata transactions are very high.

DLC Gateway Servers

The IOZONE test was performed on one gateway server with 16 threads running concurrently and a peak write throughput of 315 MB/s and read throughput of 297 MB/s was recorded. At this throughput, network utilization was very high for all three gateway server iSCSI interfaces. When another DLC gateway server was added to run 32 IOZONE threads in parallel, the read and write throughput increased to 535 MB/s and 463 MB/s respectively. Chart 1 shows the aggregate throughput for one and two DLC gateway servers.

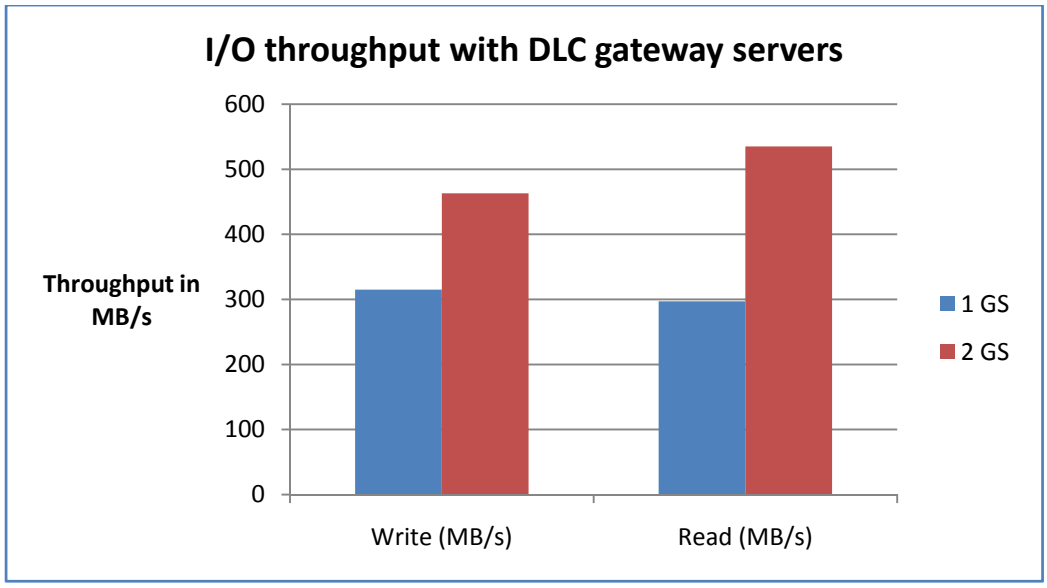

Chart 1: I/O throughput for Gateway Servers

With enough storage bandwidth, the gateway server can become a bottleneck for the compute nodes; therefore multiple gateway servers should be implemented based on available storage bandwidth and the required aggregate throughput.

Chart 2 below shows the performance trend on a DLC gateway server when there are more and more IOZONE threads running concurrently.

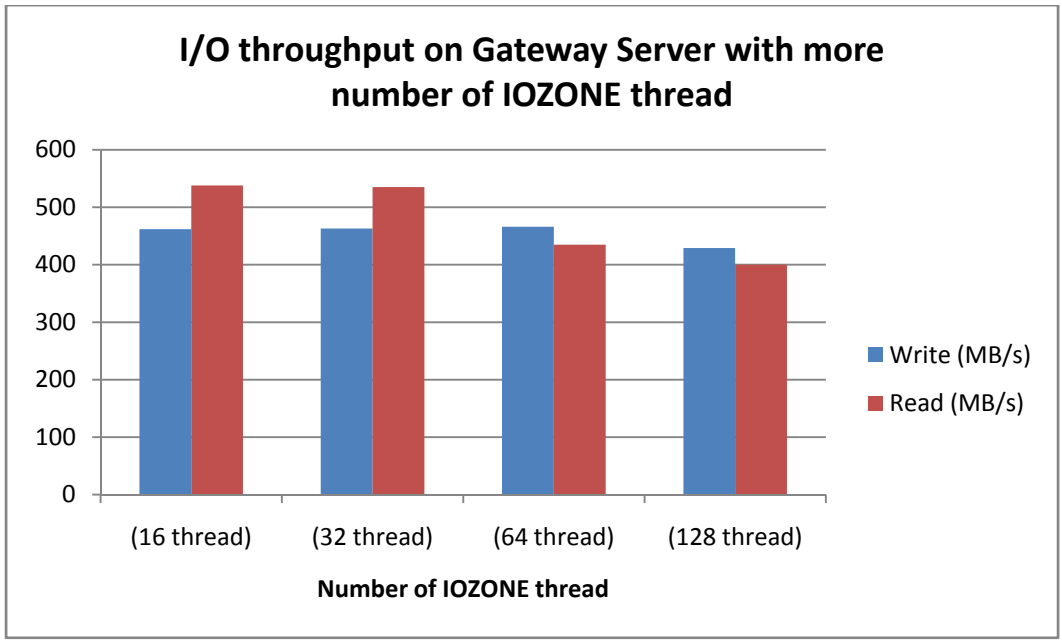

Chart 2: Throughput on Gateway Servers with Multiple IOZONE Threads Running

Throughput was limited by the 3 Gbps network on each DLC gateway server. We observed a network utilization of 90-95%, which is very high for 32 threads. The performance pattern shows a diminishing throughput due to the network limitation; the network was close to saturation with 32 IOZONE threads. This finding provides a guideline for available storage throughput with a similar load factor; with a similar number of threads and files. The result for 128 IOZONE threads suggests how the aggregate throughput will be limited on compute nodes. DLC gateway server throughput can be increased by adding more storage, or by adding more DLC gateway servers until the storage is saturated.

Compute Nodes

Each compute node was configured as distributed LAN clients (DLC) and access storage using the DLC gateway server only. With IOZONE on one of the compute nodes, the write throughput was 230 MB/s and read throughput was 212 MB/s. The total theoretical bandwidth available was 250 MB/s (2 x 1 Gbps). If we consider protocol overhead running TCP, the theoretical limit of GbE will be 117 MB/s, and therefore the total I/O bandwidth available was 234MB/s. One of the 1 Gbps links was shared between the metadata and I/O traffic. The file system efficiency achieved at the compute nodes was 91-98% of available network bandwidth. The read throughput can be increased further by tuning the file system tunable performance parameters. Chart 3 below shows the read and write throughput on single compute node with 2 Gbps of I/O network.

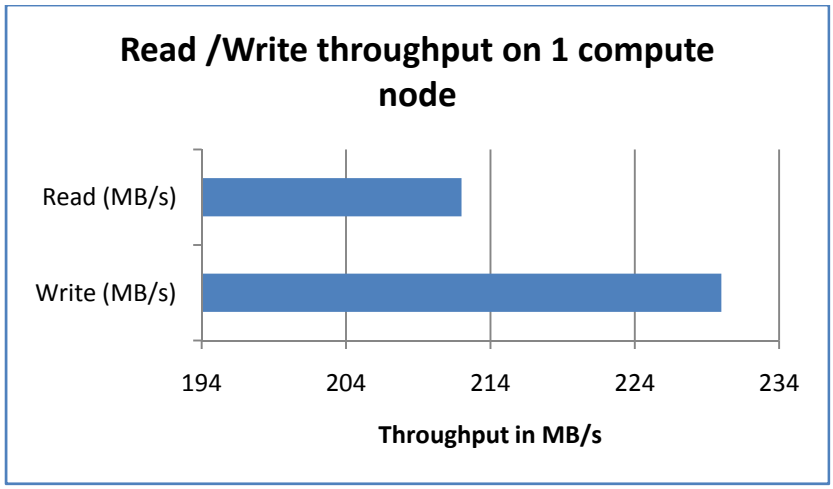

Chart 3: Read and Write Throughput on 1 Compute Node with 2Gbps of I/O Network.

Chart 4 below shows the I/O performance trend with multiple IOZONE threads running in parallel on four of the compute nodes.

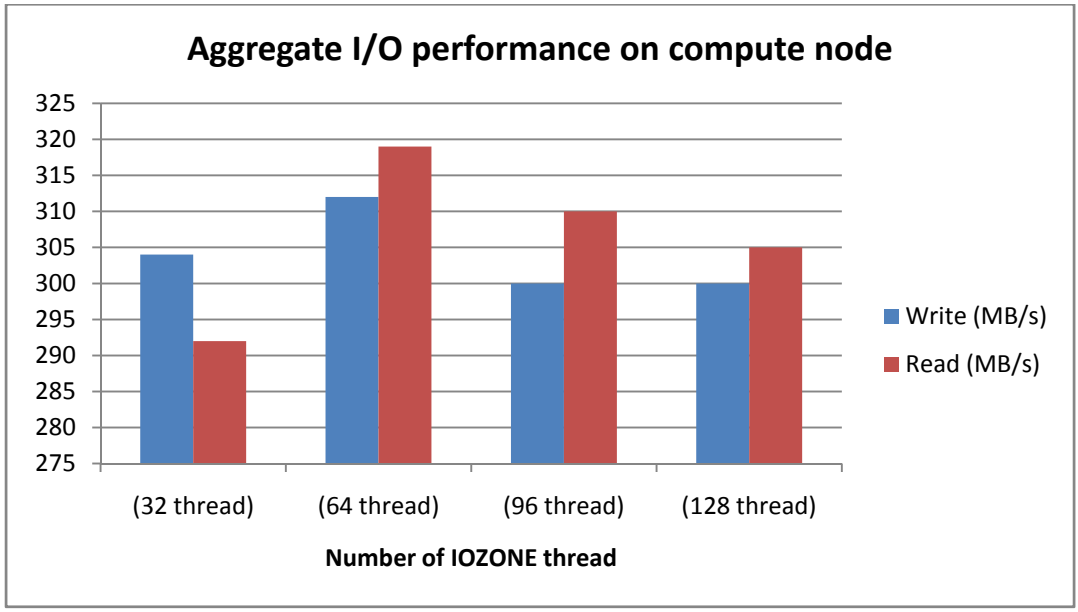

Chart 4: I/O Performance on Compute Nodes

The chart above shows that the peak throughput is achieved with 64 threads running in total on four compute nodes. With more than 64 threads, there is diminishing throughput. Also, with 64 threads running the network utilization on both of the DLC gateway servers was 90-95%. When there is a larger file system load, the throughput is expected to degrade. The aggregate throughput can be increased by adding additional DLC gateway servers to leverage the storage

bandwidth, and simultaneously increase the overall available I/O bandwidth for compute nodes.

The multiple performance tests show the various throughput trends on DLC gateway servers as well as on compute nodes. It also demonstrates how throughput can be limited by the DLC gateway server itself, the available I/O network on that server, and the iSCSI network. Based on the application requirements, the number of DLC gateway servers, I/O networks, and storage arrays should be planned in order to achieve required aggregate throughput based on capacity and performance requirements.

Since iSCSI uses the network interface for storage traffic, some applications may need to access the storage directly and bypass the DLC gateway server. However, these applications will use the same file system as the other compute nodes that access the storage using the DLC gateway server. With this architecture, it is possible to configure some of the compute nodes as SAN clients and access the data directly from storage. The reference diagram (figure 3) shows how you can connect compute nodes, and access the storage using same file system that is configured for the other compute nodes. In this configuration, multiple compute nodes configured as SAN clients, as well as distributed LAN clients, will access the same file system using different mechanisms.

If the compute node is configured as a SAN client, it should not be used as the DLC gateway server for the performance reasons. The DLC Gateway Server should be a dedicated system. Also, the metadata server should be stand alone, and have no other applications running on it.

SNFS Configuration Best Practices

This section describes some of the SNFS best practices involving the network, ISCSI storage, and file system.

Network Configuration

The following list contains the best practices for Quantum StorNext file system configuration.

- Use a dedicated metadata network if you have a large cluster size.
- Chose the distributed LAN network based on the throughput requirement. Multiple distributed LAN clients can fill several GBE networks easily, and therefore it can be a throughput bottleneck for the DLC gateway server. If you plan on more than one interface for I/O traffic on the DLC gateway server and distributed LAN client, make sure that they are on different IP sub networks.
- Use same number of NICs on DLC gateway server as well as on distributed LAN client for distributed network I/O using IP sub networks.

iSCSI Storage Configuration

- A RAID level should be selected based on the needs of the deployment since it can't be changed after file system creation.
- The network ports on both the controller module of the storage enclosure should be connected to the switch. This will provide a redundant path to the array should an array controller fail. At the same time, it will ensure there is no performance degradation if there is a control module failure.
- In order to provide redundancy at each component level, a dual controller with a dual switch should be used.

File Server Selection and File System Configuration

- The metadata controller should be a dedicated system with no other applications running on it. If the server contains other applications, it can impact performance. If MDC is going to host multiple StorNext file systems, and then plan for more memory than what is needed for a single file system. Refer to the StorNext technical documentation for more detail, or contact a Dell representative at: http://content.dell.com/us/en/enterprise/hpcc.aspx?c=us&cs=555&l=en&s=biz.
- The DLC gateway server should be a dedicated system for handling client I/O requests. Memory size will vary depending on the number of file systems, distributed clients, NICs/interfaces per server, and the transfer buffer count and size. Refer to the StorNext installation documentation for more details, or contact Dell representative at: http://content.dell.com/us/en/enterprise/hpcc.aspx?c=us&cs=555&l=en&s=biz.
- The StorNext file system has the ability to logically group the different storage resources using a stripe group. By using the stripe group feature, mixed and differently configured storage, such as JBOD, RAID1, and RAID5, can be part of same file system. Metadata and journal can use RAID1, and data can use RAID5 and other RAID configurations for better performance. The RAID level should be selected based on the deployment requirements.
- For better performance, the RAID stripe size should match the file system's stripeBreadth parameter.
- For sequential I/O throughput, the proper direct memory access (DMA) and I/O transfer size can boost performance.

Additional StorNext Features

The following list details the additional features available in the Dell solution using StorNext file system.

- Dynamic resource allocation: Dynamic resource allocation increases uptime during a service operation, by allowing online expansion of the file system and a transparent swap out of a disk during a hardware upgrade. Storage capacity can be added and scaled while system is active. Also, using the stripe group feature, data can be moved from one disk volume to another. For example, if the metadata and journal need to be moved on to another volume, this feature allows that migration to take place easily.

- Storage virtualization: StorNext offers the following storage virtualization that can be beneficial in many ways.

    - The file system can be spanned across multiple heterogeneous SAN storage systems.

    - Data can be moved dynamically from one storage system to another while data is being accessed.

    - Data can be moved from one disk volume to another using the stripe group feature.

    - In order to scale and meet the additional storage capacity requirements, file system can be dynamically expanded.

    - Can move the data automatically and transparently between online and near line storage, including disk and tape.

    - All SAN and LAN clients will have access to same data, regardless of operating system and the storage connectivity type. Refer to StorNext release notes for supported OS types.

Conclusion

This study provides an example of how a scalable I/O susbsystem can be built for Windows HPC clusters using Dell EqualLogic storage and Quantum StorNext file system, After analyzing the performance tests, it has been observed that Quantum's StorNext file system performs well with the Dell EqualLogic storage and delivers high efficiency up to greater than 90% for sequential I/O. This solution can be scaled up to meet the requirements for most HPC applications for even larger cluster sizes by adding gateway modules and storage enclosures as needed. You can also utilize storage virtualization such as mixed storage, stripe group, and other features in variety of ways based on application needs.

Application characteristics and business needs are the two most important factors in designing an I/O subsystem for HPC clusters. The proposed architecture in this paper provides a verified base starting point and allows flexibility by enabling systems to be built for your unique needs.

References:

Dell HPC Solution

http://content.dell.com/us/en/enterprise/hpcc.aspx?c=us&cs=555&l=en&s=biz

Dell Enterprise Servers

http://www.dell.com/us/en/enterprise/enterprise/ct.aspx?refid=enterprise&s=biz&cs=555&~ck=mn

Dell EqualLogic storage

http://www.dell.com/us/en/enterprise/enterprise/equallogic/cp.aspx?refid=equallogic&s=biz&cs=555

Quantum StorNext File System

http://www.quantum.com/Products/Software/StorNext/Index.aspx

Microsoft Windows 2008 HPC

http://www.microsoft.com/hpc/en/us/default.aspx

Appendix A: Install and Configuration

This section describes the installation and configuration of a file system in a Windows HPC environment that will aid in setting up the StorNext file server.

Server and Storage Preparation

1. Prepare the head node and compute nodes by following the Windows Server 2008 HPC installation procedure.
2. Install Windows Server 2008 on the MDC and DLC gateway server.
3. Configure the storage using the storage management software.
4. Implement the required storage connectivity between storage, MDC, and each of the DLC gateway servers. For a large number of DLC gateway servers, a switch may be required.
5. Install storage software, such as EqualLogic software (HIT) for PS series, for failover and load balancing features with multiple connections and other storage management software.
6. Make sure the MDC and each DLC gateway server reports the same number of volumes (as per storage configuration).

Metadata Server (MDC) Configuration

1. Install the StorNext software on the MDC.
2. Using the disk labeler tool, label all the volumes that will be part of the file system.
3. Using advanced file system configuration tool, create the disk type, disk definitions, and stripe group. For disk types, you can define the type of disk such as data, journal, or metadata with the proper number of sectors. Using the disk definitions tab, add the labeled disk to the proper disk type. Finally, using the stripe groups tab, create a different stripe group and add the available disk with the proper attributes, such as metadata, journal, or exclusive. Save this configuration.
4. Create the file system using the `cvmkfs` command.
5. Configure the name server on the MDC using the name server configuration tool. Add the IP address of the MDC. The DLC gateway server and compute node should be able to ping the same IP.
6. Using the StorNext file system administrator, start and activate the newly created file system; the file system should be up and running.

DLC Gateway Server Configuration

1. Install the StorNext software on each DLC gateway server. After installation, configure each server using the client configuration tool; configure the name server first. Add the IP of the MDC where the name server is configured. Stop and start the SNFS services on

the DLC gateway server. The software should scan the configured file system, and will appear in the client configuration window.

2. Using the edit drive mapping in Tools, map the drive. While mapping the drive using the distributed LAN tab, select **Enable server** and select **Distributed LAN network for I/O** that the client will use for I/O traffic.

Compute Nodes for Distributed LAN Client Configuration

1. Install the StorNext software on each compute node. After installation, configure each node using the client configuration tool; configure the name server first on each compute node. Add the IP of the MDC where the name server is configured. Stop and start the SNFS services on each node. The software should scan the configured file system, and will appear in the client configuration window.
2. Using the edit drive mapping in Tools, map the drive. While mapping the drive using the distributed LAN tab, select **Enable client**. After this file system is ready for use.