Microsoft Server 2012 for Dell Converged Blade Data Center: Reference Architecture

Release 1.0 for Dell PowerEdge Blade Servers, Force10 Switches, and EqualLogic Storage

Dell Global Solutions Engineering

Revision: A00

October 2012

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

October 2012| Rev A00

Revision History

Revision	**Description**	Date
A00	Initial Version	October 2012

Contents

Figures

Tables

1 Introduction

Microsoft® Windows® 2012 for Dell Converged Blade Data Center is an enterprise infrastructure solution that has been designed and validated by Dell™ Engineering. Dell Services will deploy and configure the solution tailored for business needs and ready to be integrated into your datacenter. This solution utilizes Microsoft® Windows Server® 2012 with Hyper-V® role enabled Hypervisors. This document will define the Reference Architecture for the Dell™ PowerEdge™ M420 Converged network with Dell EqualLogic™ PS-M4110X storage solution.

The architecture defined in this document includes Dell™ PowerEdge™ M420 blade servers, Dell EqualLogic™ Storage, Dell Force10™ network switches, Microsoft® Windows Server® 2012 Datacenter Edition with Hyper-V® Role enabled, and two Dell PowerEdge R620 servers that manage the solution by hosting Dell management tools.

The configurations also include EqualLogic SAN HQ and Dell OpenManage™ Essentials (OME). The configurations vary in the number of PowerEdge M420 blade servers and EqualLogic storage enclosures to meet virtualization resource needs.

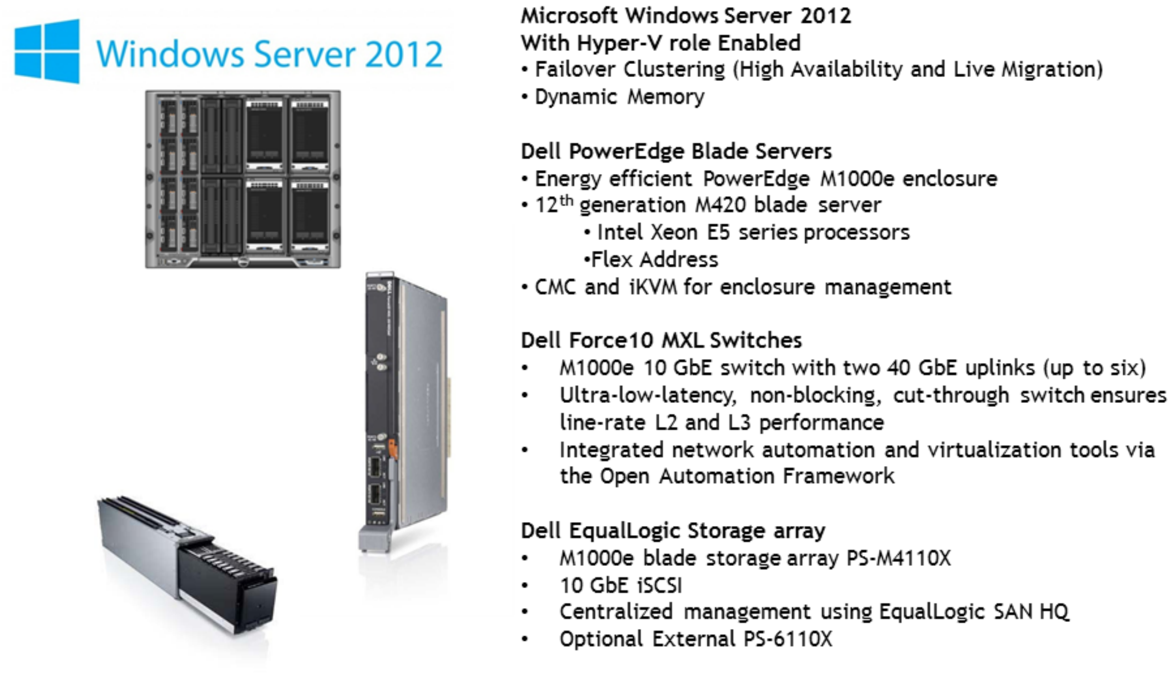
2 Audience

This document provides an overview of the solutions. Readers, including CTOs and IT managers, can use this document to understand the overview and scope of the solution. IT administrators and managers can use this document to understand the solution architecture.

3 Solution Overview

This section provides a high-level product overview of Microsoft Hyper-V, Dell PowerEdge blade servers, Dell Force10 Network Switches, and Dell EqualLogic Storage, as illustrated in Figure 1 and Figure 2 below. Readers can skip the sections of products with which they are familiar.

Figure 1: Overview of Dell Blade Servers, Dell Force10 network switches, Dell EqualLogic Storage, and Microsoft Hyper-V



Microsoft Windows Server 2012
With Hyper-V role Enabled

- Failover Clustering (High Availability and Live Migration)
- Dynamic Memory

Dell PowerEdge Blade Servers

- Energy efficient PowerEdge M1000e enclosure
- 12th generation M420 blade server
    - Intel Xeon E5 series processors
    - Flex Address
- CMC and iKVM for enclosure management

Dell Force10 MXL Switches

- M1000e 10 GbE switch with two 40 GbE uplinks (up to six)
- Ultra-low-latency, non-blocking, cut-through switch ensures line-rate L2 and L3 performance
- Integrated network automation and virtualization tools via the Open Automation Framework

Dell EqualLogic Storage array

- M1000e blade storage array PS-M4110X
- 10 GbE iSCSI
- Centralized management using EqualLogic SAN HQ
- Optional External PS-6110X

Figure 2: Dell Blade Servers, Dell Force10 network switches, and Dell EqualLogic Storage

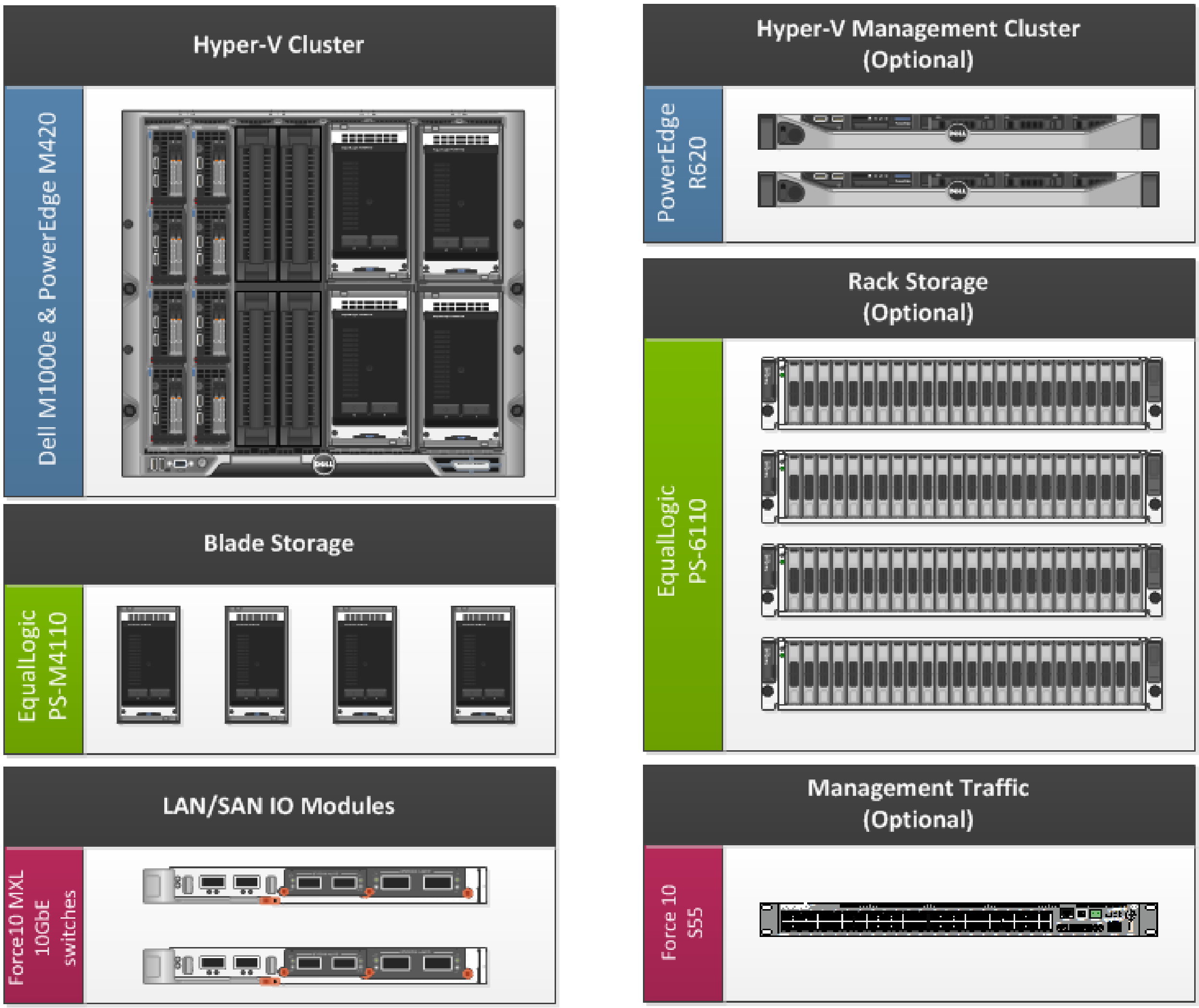


Table 1 below describes the key solution components and the roles served.

Table 1: Component Logical Groups

Component	Description	Role
Hyper-V Cluster	Dell PowerEdge M420 blade servers running Windows Server 2012 Datacenter Edition and Hyper-V role enabled	Host virtual machines (VMs)
Storage	Up to four Dell EqualLogic PS-M4110X and/or PS-6110X storage enclosures	Provide shared storage for the Hyper-V cluster to host the VMs
LAN/SAN Traffic Switches	Two Force10 MXL 10GBE switches for the chassis	Support VM, Live Migration, Management, iSCSI and Cluster traffic
Management Cluster (optional)	Dell PowerEdge R620 servers running Microsoft® Windows Server® 2012 Datacenter Edition	Host optional EqualLogic SAN-HQ and Dell OpenManage Essentials
OOB management Switch (optional)	One Force10 S55	Provide OOB management connectivity

3.1 Microsoft Windows Server 2012

Microsoft® Windows Server® 2012 is Microsoft’s flagship server operating system which provides the Hyper-V® virtualization platform. Hyper-V provides a virtualization platform that can consolidate Windows® and Linux workloads enabling IT managers the ability to more fully utilize their available hardware resources.

Dynamic Memory: Dynamic Memory allows Virtual Machines (VM) to dynamically use physically available memory. Allowing systems to boot with minimal memory, then expand and contract memory on demand, Dynamic Memory further increases the optimization of physical resources and the density of VM workloads.

Live Migration: Integrated as part of Windows Server Failover Clustering, Live Migration provides customers with the ability to move VMs from one host within a cluster to another with near zero down time.

3.2 Dell OpenManage Essentials

The Dell OpenManage™ Essentials Console provides a single, easy-to-use, one-to-many interface through which to manage resources in multivendor operating system and hypervisor environments. It automates basic repetitive hardware management tasks — like discovery, monitoring and updates — for Dell servers, storage and network systems. OME employs the embedded management of PowerEdge™ servers — Integrated Dell Remote Access Controller 7 (iDRAC7) with Lifecycle Controller — to enable agent-free remote management and monitoring of server hardware components like storage, networking, processors and memory.

OpenManage Essentials helps you maximize IT performance and uptime with capabilities like:

- **Automated discovery, inventory, and monitoring** of Dell PowerEdge™ servers, EqualLogic™ and PowerVault™ storage and PowerConnect™ switches
- **Agent-free server monitoring** as well as BIOS, firmware and driver updates for Dell PowerEdge servers, blade systems and internal storage
- **Control of PowerEdge servers** within Windows®, Linux®, VMware® and Hyper-V® environments
- Interface with additional optional Dell systems management solutions including:
- Repository Manager to facilitate and secure precise control of system updates
- KACE® K1000 Appliance service desk to provide actionable email alerts describing the status of Dell servers, storage and switches
- Dell ProSupport™ phone home services for your data center resources

For more information on OpenManage Essentials, see Dell.com/openmanageessentials.

3.3 Dell PowerEdge Blade Servers

Blade Modular Enclosure: The Dell PowerEdge M1000e is a high-density, energy-efficient blade chassis that supports up to sixteen half-height blade servers, or eight full-height blade servers, and six I/O modules. A high-speed passive mid-plane connects the server modules to the I/O modules, management, and power in the rear of the chassis. The enclosure includes a flip-out LCD screen (for local configuration), six hot-pluggable/redundant power supplies, and nine hot-pluggable N+1 redundant fan modules.

Blade Servers: The Dell PowerEdge M1000e Blade Server Chassis supports the Dell PowerEdge M420 blade servers based on Intel® Xeon® E5 series processors. Dell's embedded management houses the tools and enablement pieces for management directly on the server, allowing administrators to perform a complete set of provisioning functions from a single, intuitive interface. Zero-media low-touch deployment capabilities enable IT administrators to provision workloads in an efficient, secure, and user-friendly manner. The enclosure supports all Dell PowerEdge 12th generation blade servers, but the PowerEdge M420 was selected as the optimum blade server for this solution.

I/O Modules: The enclosure provides three redundant fabrics using six I/O modules. The modules can be populated with Ethernet switches, Fibre Channel (FC), and pass-through modules.

Chassis Management: The Dell PowerEdge M1000e has integrated management through a redundant Chassis Management Controller (CMC) module for enclosure management and integrated keyboard, video, and mouse (iKVM) modules. Through the CMC, the enclosure supports FlexAddress Plus technology which enables the blade enclosure to lock the World Wide Names (WWN) of the FC controllers and Media Access Control (MAC) addresses of the Ethernet controllers to specific blade slots. This enables seamless swapping or upgrading of blade servers with Ethernet and FC controllers without affecting the LAN or SAN configuration.

Embedded Management with Dell's Lifecycle Controller: The Lifecycle Controller is the engine for advanced embedded management and is delivered as part of iDRAC7 Enterprise in Dell PowerEdge 12th generation blade servers. Embedded management includes:

- Unified Server Configurator (USC) aims at local 1-to-1 deployment via a graphical user interface (GUI) for operating system install, updates, configuration, and for performing diagnostics on single, local servers. This eliminates the need for multiple option ROMs for hardware configuration.
- Remote Services are standards-based interfaces that enable consoles to integrate, for example, bare-metal provisioning and one-to-many OS deployments, for servers located remotely. Dell's Lifecycle Controller takes advantage of the capabilities of both USC and Remote Services to deliver significant advancement and simplification of server deployment.
- Lifecycle Controller Serviceability aims at simplifying server re-provisioning and/or replacing failed parts and thus reduces maintenance downtime.

For more information on Dell Lifecycle Controllers and blade servers, see http://content.dell.com/us/en/enterprise/dcsm-embedded-management and Dell.com/blades.

3.4 Dell Force10 MXL Switches

The Force10 MXL is an ultra-low-latency 10/40 GbE switch module for the M1000e chassis purpose-built for applications in high-performance data center and computing environments. Leveraging a non-blocking, cut-through switching architecture, the MXL delivers line-rate L2 and L3 forwarding capacity with ultra-low latency to maximize network performance. The compact MXL design provides a chassis switch with 1/10 GbE (SFP+) ports as well as up to six 40 GbE QSFP+ uplinks to simplify the migration to 40 Gbps in the data center core. (Each 40 GbE QSFP+ uplink can support four 10 GbE ports with a breakout cable). Powerful Quality of Service (QoS) features coupled with Data Center Bridging (DCB) support make the MXL ideally suited for iSCSI storage environments. In addition, the MXL incorporates multiple architectural features that optimize data center network flexibility, efficiency, and availability, including Force10's stacking technology, reversible front-to-back or back-to-front airflow for hot/cold aisle environments, and redundant, hot-swappable power supplies and fans. For more information on Force10 switches, see Dell.com/force10.

3.5 Dell EqualLogic PS-M4110 and PS-6110 Storage Arrays

Dell EqualLogic PS-M4110 storage array is a blade storage solution which connects to the blade MXL switches with 10GbE redundant interconnects. Both the PS-M4110 and the PS-6110 have 2 storage processors in an active-passive configuration.

Key features encompass the following:

- **Storage Virtualization** – Storage is virtualized at the disk level to create a flexible pool of storage resources shared by all servers all the time.
- **Thin Provisioning** – Allocation is completely separated from utilization so any size volume can be created at any time, yet capacity is only consumed when data is written.
- **Replication** – All data can be replicated on separate storage groups. These groups are usually recommended to be at different physical locations with robust networks between them.
- **Storage Management** – All storage resources are managed through a single point-and-click interface, providing a complete view of the entire storage environment.

For more information on Dell EqualLogic, see Dell.com/EqualLogic. Contact your Dell sales representative for more information on EqualLogic storage configurations and sizing guidelines.

3.6 Dell Force10 S55 - Optional

The Dell Force10 S-Series S55 1/10 GbE ToR switch is designed for high-performance data center applications. The S55 leverages a non-blocking architecture that delivers line-rate, low-latency L2 and L3 switching to eliminate network bottlenecks. The high-density S55 design provides 48 GbE access ports with up to four modular 10GbE uplinks in 1-RU to conserve valuable rack space. The S55 incorporates multiple architectural features that optimize data center network efficiency and reliability, including reversible front-to-back or back-to-front airflow for hot/cold aisle environments and redundant, hot-swappable power supplies and fans.

3.7 PowerEdge R620 Management Server - Optional

The Dell PowerEdge R620 uses Intel® Xeon® E5-2600 series processors in a 1U rack mount form factor. These servers support up to ten 2.5" drives and provide the option for an LCD located in the front of the server for system health monitoring, alerting, and basic management configuration. An AC power meter and ambient temperature thermometer are built into the server, which can be monitored on this display without any software tools. The server features two CPU sockets and 24 memory DIMM slots supporting 2, 4, 8, 16 or 32GB DIMMs.

Energy-efficient design features include power-supply units sized appropriately for system requirements, innovative system-level design efficiency, policy-driven power and thermal management, and highly efficient standards-based Energy Smart components. For more information, see the *PowerEdge R620 guides at* Dell.com/PowerEdge.

4 Design Principles

This section covers the design principles, requirements, and solution capabilities incorporated in the Converged Blade Data Center solution architecture.

4.1 Optimal hardware configuration for virtualization

The solution is designed with an optimal hardware configuration to support virtualization. Each blade server is configured with sufficient memory and network adapters required for virtualization.

4.2 Redundancy with no single point of failure

The solution is designed so that there is no single point-of-failure and redundancy is incorporated into all mission critical components of the solution. Management applications are not architected with this level of redundancy because the mission critical workloads will continue to operate in the event of a management application failure. Network redundancy for the mission critical components is achieved with redundant network interface controllers (NICs) and redundant switches. NIC teaming for LAN and MPIO for SAN are used to provide failover across the redundant network interfaces.

For network traffic, NIC ports are teamed in such a way that they avoid any single point of failure. Hyper-V High Availability (HA) is provided by Windows Server 2012 Failover Clustering. The solution also includes redundant power supplies connected to separate PDUs.

5 Prerequisites and Datacenter Planning

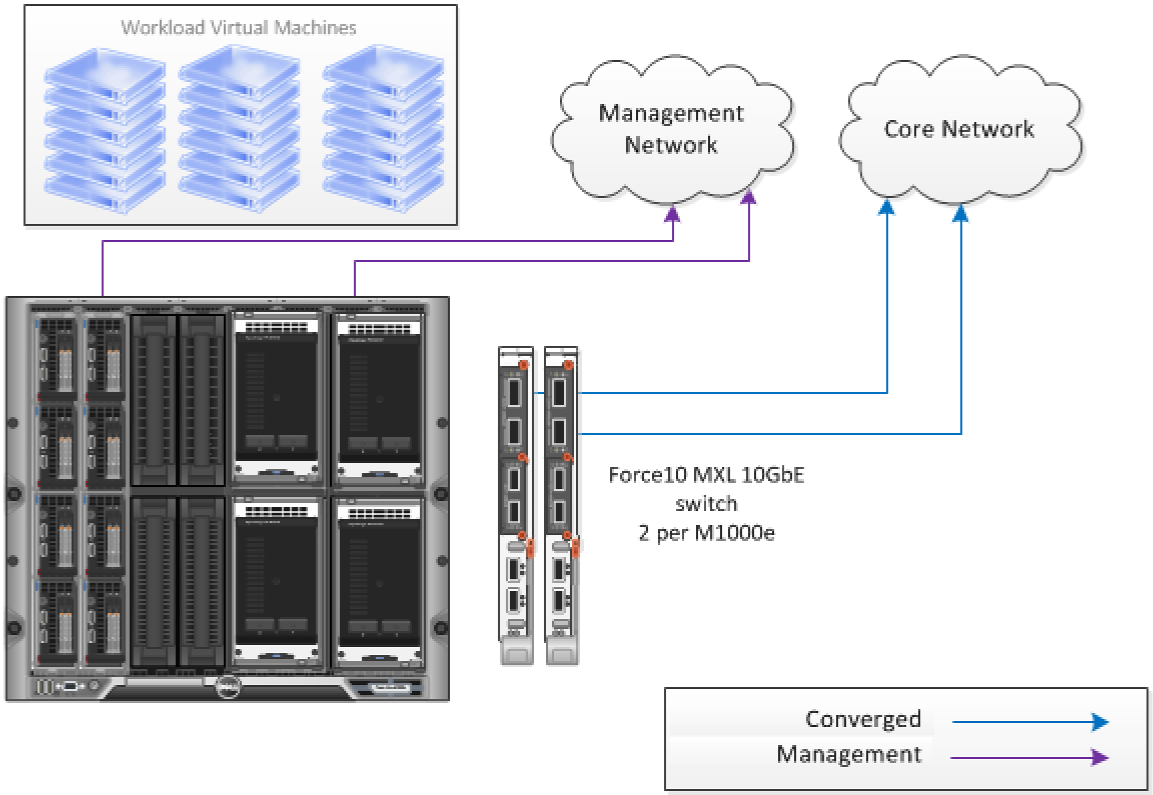
To support the architecture, the following components are required to be present in the customer environment:

- An existing 10 Gb Ethernet infrastructure with which to integrate. In addition, it is assumed that there is a 1 Gb Ethernet network infrastructure in place to support the management network.
- Additional components, such as Force10 network cables and transceivers, are needed to uplink the solution to the customer network. The components to be added will depend on customer networking and uplink requirements.
- Active Directory® (AD) Domain Services (AD DS) – An AD DS domain must be available on the network. The Hyper-V hosts will be joined to an existing or new domain. Cluster Services also require AD DS. Consult with your Dell Sales and Services representatives for more details.
- Domain Name Server (DNS) – DNS must be available on the management network.
- Sufficient power and cooling to support the solution must be present. Detailed power, weight, and cooling requirements for the datacenter are defined in the *Microsoft Windows 2012 for Dell Converged Blade Data Center Specification.*

6 Architecture

This solution consists of a Dell PowerEdge M1000e chassis populated with PowerEdge M420 blade servers running Windows Server 2012 Data Center Edition. Figure 3 below depicts the high-level reference architecture for the solution including solution components and redundant connectivity for each I/O fabric.

Figure 3: Network Topology (Logical View)


Workload Virtual Machines
Management Network
Core Network
Force10 MXL 10GbE switch 2 per M1000e
Converged
Management

6.1 Dell Blade Network Architecture

The Dell blade chassis has three separate fabrics referred to as A, B, and C. Each fabric has two I/O modules, making a total of six I/O modules slots in the chassis. The I/O modules are A1, A2, B1, B2, C1, and C2. Each I/O module can be an Ethernet physical switch, an Ethernet pass-through module, FC switch, or FC pass-through module. Each quarter-height blade server has a LAN on Motherboard (LOM).

The Chassis Fabric A contains two 10 GbE Force10 MXL switches and is used for all traffic.

PowerEdge M420 blade servers use an embedded Broadcom 57810-k Dual port 10Gb to connect to Fabric A. The MXL modules uplink to a core network for LAN connectivity.

The network traffic on each blade includes iSCSI as well as traffic for the parent partition (hypervisor), Live Migration, cluster heartbeat and cluster shared volume, and child partitions (virtual machines). A Smart Load-balancing (SLB) with Failover team is created using the two 10 GbE ports, and VLAN adapters are created on top of the team for each traffic type. A Virtual Network Switch is then created and bound to the Virtual Machine VLAN adapter.

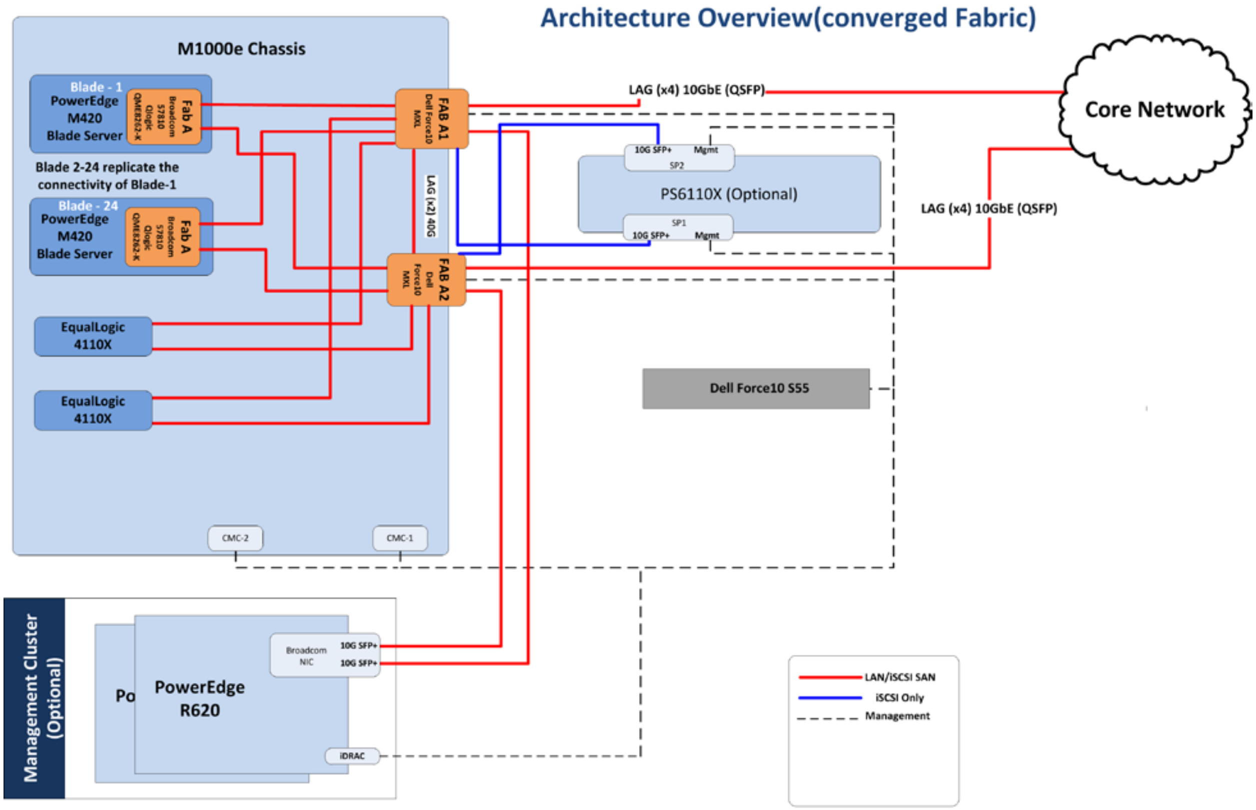
6.2 Server / Blade Network Connectivity

Each host network adapter in both the compute and management hosts utilizes network teaming technology to provide highly available network adapters to each layer of the networking stack. The teaming architecture closely follows the *Hyper-V: Live Migration Network Configuration Guide*, but extends further to provide highly available networking to each traffic type used in the architecture.

The 10GbE embedded NIC is utilized in single function mode, network partitioning (NPAR) is not utilized in this configuration. Each port has the hardware iSCSI initiator enabled.

As mentioned previously, each PowerEdge M420 blade server is configured with a Broadcom BCM57810 bNDC providing two 10GbE ports. These ports are wired to the MXL modules in Fabric A. Meanwhile, each PowerEdge R620 rack server is configured with a Broadcom BCM57810 Add-in NIC providing two 10Gb SPF+ ports, and they are also connected with the two Force10 MXL switches. The two Force10 MXL switches are configured with Inter Switch Links (ISL) using two 40 Gbps QSFP+ links. Link Aggregation Groups (LAGs) are created between the two 40 Gbps QSFP+ ports, providing a path for communication across the switches. Network connectivity for the M420 is illustrated in Figure 4 below.

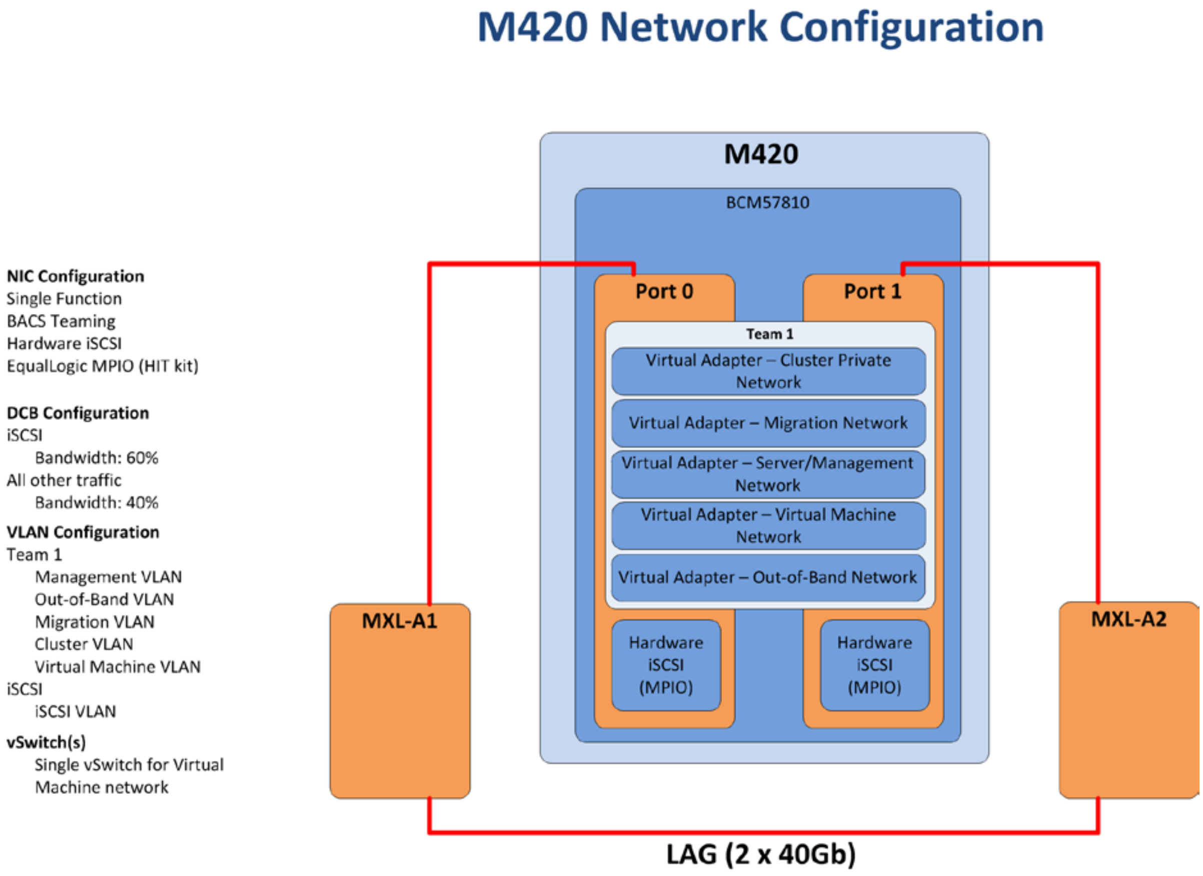
Figure 4: Dell Blade LAN Connectivity Overview



On both the PowerEdge M420 and PowerEdge R620 servers, each of their 10GbE ports is configured for single function mode. Then, on each server, four NIC teams are created and configured as switch independent teams using the Broadcom Advanced Control Suite utility.

Figure 5 illustrates the network configuration on a PowerEdge M420 blade server and the configuration of the optional PowerEdge R620 management servers. Different VLAN IDs are assigned to these teamed NICs to segregate the traffic on the host and provide the segmentation necessary for cluster management, Live Migration (Migration), cluster private (Cluster), virtual machine, and other types of traffic as described in Table 2 below. The VLAN configuration used in this solution is listed in Table 3.

Figure 5: Teaming and VLAN Configuration on a PowerEdge M420 Server



M420 Network Configuration

NIC Configuration
Single Function
BACS Teaming
Hardware iSCSI
EqualLogic MPIO (HIT kit)

DCB Configuration
iSCSI
  Bandwidth: 60%
All other traffic
  Bandwidth: 40%

VLAN Configuration
Team 1
  Management VLAN
  Out-of-Band VLAN
  Migration VLAN
  Cluster VLAN
  Virtual Machine VLAN
iSCSI
  iSCSI VLAN

vSwitch(s)
  Single vSwitch for Virtual Machine network

M420
BCM57810
Port 0
Port 1
Team 1
Virtual Adapter – Cluster Private Network
Virtual Adapter – Migration Network
Virtual Adapter – Server/Management Network
Virtual Adapter – Virtual Machine Network
Virtual Adapter – Out-of-Band Network
Hardware iSCSI (MPIO)
Hardware iSCSI (MPIO)
MXL-A1
MXL-A2
LAG (2 x 40Gb)

Table 2: Traffic Description

Traffic Type	Use
Compute Node Hypervisor (Management)	Supports virtualization management traffic and communication between the host servers in the cluster.
Migration	Supports migration of VMs between the host servers in the cluster.
Tenant VM	Supports communication between the VMs hosted on the cluster and external systems.
Compute Cluster Private	Supports internal cluster network communication between the servers in the cluster.
Out-of-Band Management	Supports configuration and monitoring of the servers through the iDRAC management interface, storage arrays, and network switches.
iSCSI Data	Supports iSCSI traffic between the servers and storage array(s). In addition, traffic between the arrays is supported.
Management Node Hypervisor	Supports virtualization management traffic and communication between the management servers in the cluster.

Management Cluster Private	Supports private cluster traffic for the management clusters.
Management Live Migration	Supports migration of VMs between the management servers in the cluster.
Management VM	Supports the virtual machine traffic for the management virtual machines.

Table 3: Sample VLAN and subnet configuration

Traffic Type	Sample VLAN	Sample Subnet
Out-of-Band Management	24	10.110.4.0/24
Compute Node Management	20	10.110.0.0/24
Compute Live Migration	21	10.110.1.0/24
Compute Cluster Private	22	10.110.2.0/24
Management VM Network	23	10.110.3.0/24
iSCSI	25	10.110.5.0/24
Management Hypervisor	26	10.110.6.0/24
Management Cluster LM	27	10.110.7.0/24
Management Cluster Private	28	10.110.8.0/24
SQL Clustering	30	10.110.10.0/24
VM Network	32	10.110.12.0/22

6.3 Server / Blade Storage Connectivity

In this configuration, each PowerEdge M420 uses an internal RAID controller PERC H310 and is connected to two SAS SSD configured in RAID-1. Each optional PowerEdge R620 management server uses an internal RAID controller PERC H710 and is connected to two SAS HDDs configured in RAID-1. This RAID volume hosts Windows Server 2012 for the hypervisor OS.

Each server uses its embedded Broadcom network adapters with hardware iSCSI enabled to connect to the iSCSI network.

6.4 Server /Blade HA and Redundancy

The PowerEdge M420 blade chassis enclosure PowerEdge M1000e is designed with redundant power supplies and redundant fans. Each PowerEdge M420 uses a PERC H310 RAID controller and two hard drives configured in a RAID-1 which hosts the parent operating system.

The design of PowerEdge R620 servers includes high availability and redundant features such as redundant fans and power supplies that are distributed to independent power sources. The servers also use PERC H710 controllers with two hard disks configured with RAID-1 to prevent server crashes in the event of single disk failures.

6.5 Management

6.5.1 Management Components

This section describes in more detail the various management components that were introduced in the Overview section.

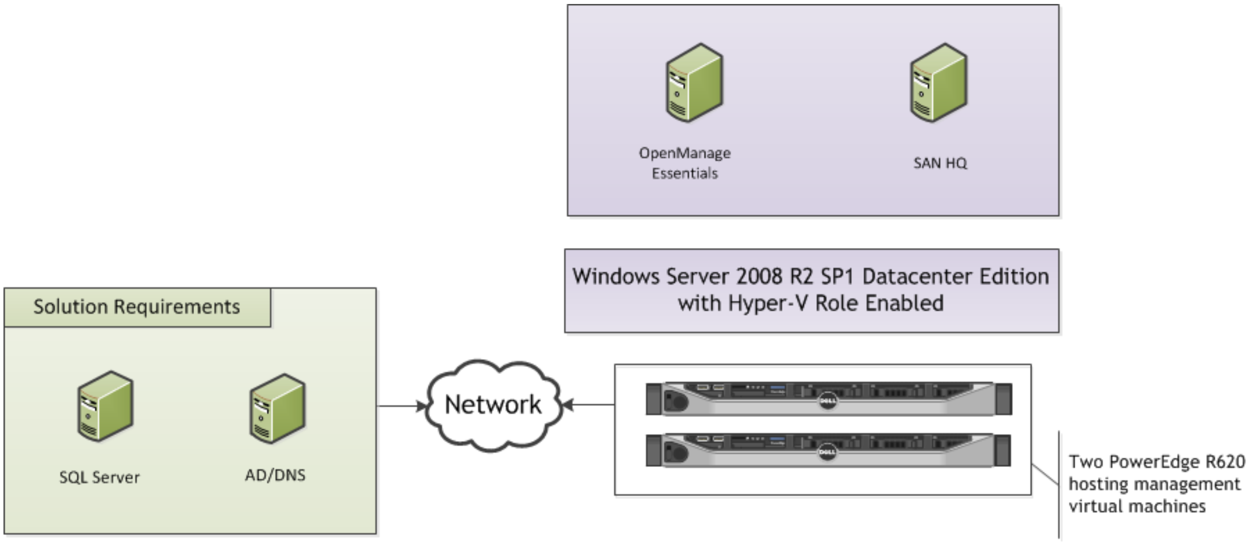
SAN HQ: SAN HQ simplifies storage management by providing a single, centralized console for the administration of multiple local and remote EqualLogic systems. Users can configure storage, monitor storage capacity and disk utilization in real time, and generate comprehensive enterprise storage usage and performance reports.

Dell Lifecycle Controller: This helps reduce operating costs by simplifying deployment and management. Key features include diagnostics, self-update (UEFI, Driver Pack update), firmware updates (BIOS, NIC FW, RAID Controllers), and hardware configuration.

Out-of-band CMC and iDRAC: The CMC provides a single, secure interface to manage the inventory, configuration, monitoring, and alerting for chassis components (iKVM, CMC), I/O modules, servers, and iDRAC. It also provides excellent real-time power management, monitoring, and alerting capabilities. The Dell chassis provides users with system-level power limiting, slot-based prioritization, and dynamic power engagement functionalities. The iDRAC on each server provides the flexibility to remotely manage the server through Console redirection and Virtual CD-ROM/DVD/Floppy/Flash capabilities.

The management applications discussed above are installed as virtual machines in the PowerEdge R620 server management cluster. Figure 6 below illustrates the Virtual Machines used in the management cluster along with optional components.

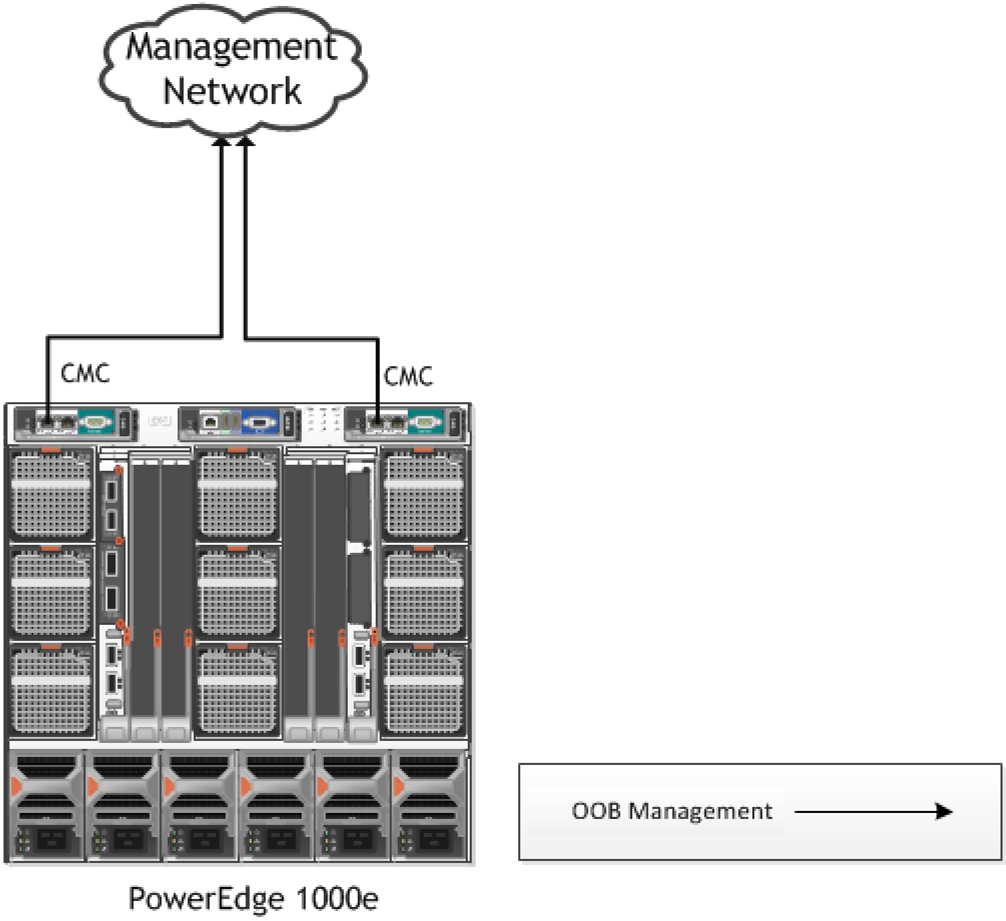
Figure 6: Optional Management Cluster


OpenManage Essentials
SAN HQ
Windows Server 2008 R2 SP1 Datacenter Edition with Hyper-V Role Enabled
Solution Requirements
SQL Server
AD/DNS
Network
Two PowerEdge R620 hosting management virtual machines

6.5.2 Management Connectivity

The Force10 S55 switch is used as a 1GbE out-of-band management switch. Each of the solution components are connected to the S55, as shown in Figure 7. The S55 switch is uplinked to each of the S4810 switches for core network connectivity.

Figure 7: Connectivity of management components.



6.6 Storage Architecture

6.6.1 Storage Options

This solution utilizes EqualLogic iSCSI storage, both blade storage and rack mount arrays. iSCSI is used for hypervisor connectivity and provides a SAN fabric for the storage traffic. DCB is used to guarantee the hypervisors dedicated bandwidth to provide the VMs a very low latency and high bandwidth option for storage connectivity. iSCSI also provides an interface for the VMs to have direct access to enable in-guest clustering.

6.6.2 SAN Storage Protocols

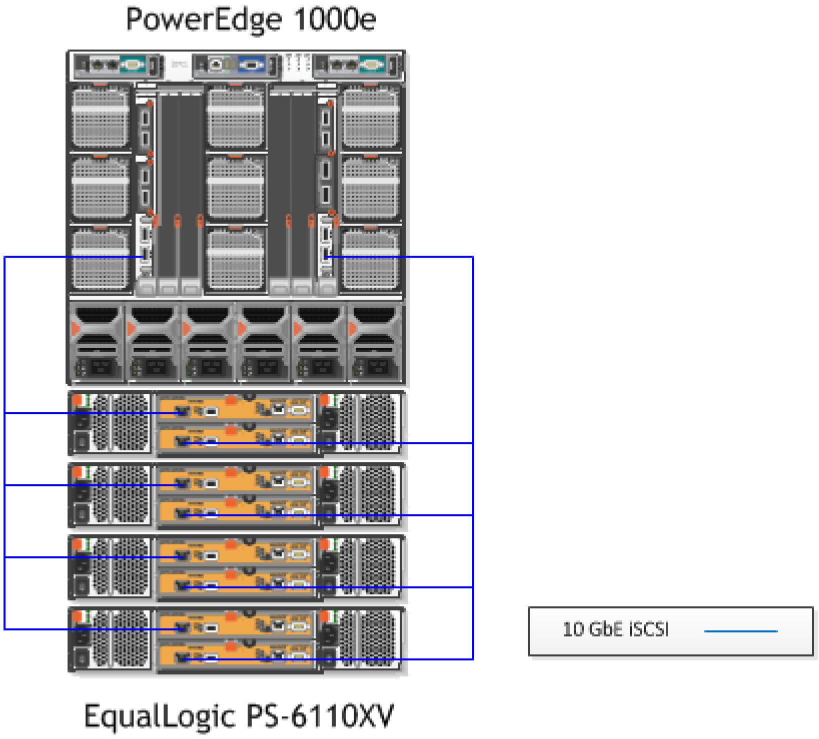
This Dell solution utilizes iSCSI protocols. For Hyper-V, iSCSI-capable storage provides an advantage in that it is the protocol that can also be utilized by Hyper-V guest virtual machines for guest clustering. This requires that VM storage traffic and other network traffic flow over the same interface, however this contention is mitigated through the use of VLANs and DCB on the network adapter, EqualLogic storage, Force10 MXL switches, and the operating system.

6.6.3 Storage Network

To support the Fabric Management guest cluster, the PowerEdge R620 server is also configured with iSCSI connectivity to the EqualLogic by using its dual port Broadcom BCM57810 10GbE Add-in card. Both ports are configured in single function mode and feature a virtual function dedicated to iSCSI traffic on the converged network. The connectivity to the EqualLogic iSCSI front-end is established via the two Force10 MXL switches. To provide the fully redundant and independent paths for storage I/O, MPIO is enabled by the iSCSI initiator on the host. The iSCSI traffic on PowerEdge R620 is segregated by the implementation of VLAN. QoS is provided by DCB and can be tailored to any percentage of the whole.

In the 24-blade configuration, two 40 GbE ports are used from each Force10 MXL switches to connect to each active and standby controller in every PS-6110XV.

Figure 8: Optional SAN Connectivity Overview



6.6.4 Performance

Dell EqualLogic PS-M4110X and PS-6110X, with the dual-controller configuration and 10 GbE connections, provide high bandwidth for data traffic. This bandwidth is complemented with a large variety of drives in multiple speeds and sizes. The Series 40 also uses virtual port IQNs and WWNs, thereby enabling higher throughput and fault tolerance.

6.6.5 Drive Types

Dell EqualLogic storage enclosures feature both 10 GbE and 1 GbE network connections. The storage arrays chosen for this reference architecture utilize 10 GbE.

The Microsoft Windows 2012 for Dell Converged Blade Data Center solution is based upon the number of compute nodes. As an example the 8 blade configuration utilizes four 14 bay 2.5" SAS drive enclosures (PS-M4110XV). The 2.5" drives selected are 15K RPM SAS, with 300GB drives in both the PS-M4110XV. These drives are selected for best IOPS performance.

6.6.6 RAID Array Design

Dell EqualLogic storage supports RAID 5, 6 and 10. To optimize performance from an IOPS stand point, only RAID 10 was used for sizing and is recommended for virtual machine storage. If additional LUNs are desired for other data types, RAID 5 or 6 may be a better fit.

6.6.7 Cluster Shared Volumes

Cluster Shared Volumes (CSV) is the storage volumes of choice for Hyper-V clusters. Developed by Microsoft exclusively for Hyper-V, they enable multiple Hyper-V cluster nodes to simultaneously access VMs. The CSVs are used throughout this solution for both the fabric and fabric management servers.

6.6.7.1 CSV Limits

The below limitations are actually imposed by the NTFS file system and are inherited by CSV.

Table 4: CSV Limits

CSV parameter	Limitation
Maximum Volume Size	256 TB
Maximum # Partitions	128
Directory Structure	Unrestricted
Maximum Files per CSV	4+ Billion
Maximum VMs per CSV	Unrestricted

6.6.7.2 CSV Requirements

- All cluster nodes must use Windows Server 2012.
- All cluster nodes must use the same drive letter for the system disk.
- All cluster nodes must be on the same logical network subnet. Virtual LANs (VLANs) are required for multi-site clusters running CSV.

- NT LAN Manager (NTLM) authentication in the local security policy must be enabled on cluster nodes.
- SMB must be enabled for each network on each node that will carry CSV cluster communications.
- “Client for Microsoft Networks” and “File and Printer Sharing for Microsoft Networks” must be enabled in the network adapter’s properties to enable all nodes in the cluster to communicate with the CSV.
- The Hyper-V role must be installed on any cluster node that may host a VM.

6.6.7.3 CSV Volume Sizing

Because all cluster nodes can access all CSV volumes simultaneously, standard LUN allocation methodologies are used and are based on performance and capacity requirements of the workloads running within the VMs themselves. Generally speaking, isolating the VM Operating System I/O from the application data I/O is a good start, in addition to application-specific I/O considerations such as segregating databases and transaction logs and creating SAN volumes and/or Storage Pools that factor in the I/O profile itself (i.e., random read and write operations vs. sequential write operations).

CSV’s architecture differs from other traditional clustered file systems, which frees it from common scalability limitations. As a result, there is no special guidance for scaling the number of Hyper-V Nodes or VMs on a CSV volume other than ensuring that the overall IO requirements of the expected VMs running on the CSV are met by the underlying storage system and storage network.

Each enterprise application that is planned to run within a VM may have unique storage recommendations and even, perhaps, virtualization-specific storage guidance. That guidance applies to use with CSV volumes as well. It is important to keep in mind that all VM’s virtual disks running on a particular CSV will contend for storage I/O.

Also worth noting is that individual SAN LUNs do not necessarily equate to dedicated disk spindles. A SAN Storage Pool or RAID Array may contain many LUNs. A LUN is simply a logic representation of a disk provisioned from a pool of disks. Therefore, if an enterprise application requires specific storage IOPS or disk response times, all the LUNs in use on that Storage Pool must be considered. An application which would require dedicated physical disks, were it not virtualized, may require dedicated Storage Pools and CSV volumes running within a VM.

6.6.8 High Availability

In order to maintain continuous connectivity to stored data from the server, each storage member in a storage group has redundant I/O paths. This provides for continuous connectivity with no single point of failure. Each PS-M4110 has 2 controllers that each has a connection to each MXL switch. In this implementation, if one port fails, the member will use the other port on the same controller.

6.6.9 Multi-pathing

For Windows Server 2012, both EqualLogic Host Integration Tools (HIT) and the built-in generic Microsoft DSM (MSDSM) provide functionality for Dell EqualLogic. The multi-pathing solution uses the Round Robin load balancing algorithm to utilize all available paths.

6.6.10 iSCSI

For some workloads, iSCSI connectivity must be presented to tenant VMs to create guest clusters. The traffic separation is implemented by DCB and VLANs.

6.6.11 Encryption and Authentication

Challenge Handshake Authentication Protocol (CHAP) is available for use on EqualLogic storage.

6.6.11.1 Jumbo Frames

In the Converged Blade Data Center configuration, jumbo frames are enabled for all devices. This includes the server network interface ports, the network switch interfaces, and the EqualLogic interfaces.

6.6.11.2 Thin Provisioning

Particularly in virtualization environments, thin provisioning is a common practice. This allows for efficient use of the available storage capacity. The LUN and corresponding CSV may grow as needed, typically in an automated fashion to ensure availability of the LUN (auto-grow). However, as storage becomes over-provisioned in this scenario, very careful management and capacity planning is critical.

6.6.11.3 Volume Snapshots

SAN Volume snapshots are a common method of providing a point-in-time, instantaneous backup of a SAN Volume or LUN. These snapshots are typically block-level and only utilize storage capacity as blocks change on the originating volume. Some SANs provide tight integration with Hyper-V, integrating both the Hyper-V Microsoft Volume Shadow Copy Service (VSS) Writer on Hosts and Volume Snapshots on the SAN. This integration provides a comprehensive and high-performing backup and recovery solution.

6.6.11.4 Storage Tiering

Tiering storage is the practice of physically partitioning data into multiple distinct classes based on price, performance, or other attributes. Data may be dynamically moved among classes in a tiered storage implementation based on access activity or other considerations.

6.7 Scalability

As workloads increase, the solution can be scaled to provide additional compute and storage resources independently.

Scaling Compute and Network Resources: This solution is configured with two MXL network switches. Up to two PowerEdge M1000e chassis can be utilized in this configuration. In order to scale the compute nodes beyond two chassis, two S4810 switches need to be added as a ToR switch solution. Additional switches can either be stacked together and/or connected to this distribution switch based on customer needs, it is recommended that VLT be used with S4810 switches.

Scaling Storage Resources: EqualLogic storage can be scaled seamlessly and independent of the compute and network architectures. Additional enclosures can be added to the existing controllers. New volumes can be created or existing volumes can be expanded to utilize the capacity in the added enclosures. This design is currently limited to four additional enclosures in the single rack configuration and eight additional enclosures in the two rack configurations.

7 Delivery Model

This reference architecture is currently only available in this white paper. Dell Services will deploy and configure the solution tailored to the business needs of the customer based on the architecture developed and validated by Dell Engineering. For more details or questions about the delivery model, please consult with your Dell Sales representative.

8 Additional Reading and Resources

8.1 Dell PowerEdge Server Documentation and Hardware/Software Updates

For Drivers and other downloads: Visit http://support.dell.com, click "Start Here" for Enterprise IT, "Select a product", then enter a server service tag or select the server model and operating system version.

8.2 Dell Force10 Switch Documentation and Firmware Updates

Visit http://support.dell.com or click here for directions to sign onto your Force10 Online Support account.

8.3 Microsoft® Hyper-V Documentation

Install the Hyper-V Role and Configure a Virtual Machine:
http://technet.microsoft.com/library/hh846766.aspx

Supported Guest Operating Systems:
http://technet.microsoft.com/library/hh831531.aspx

Failover Clustering Overview:
http://technet.microsoft.com/en-us/library/hh831579.aspx

Data Center Bridging:
http://technet.microsoft.com/en-us/library/hh849179.aspx