

Business Ready Configurations: Reference Architecture for VMware vSphere 4 with PowerEdge Blade Servers and Dell/EMC CX4 Fibre Channel Storage

April 2010

Sitha Bhagvat; Balasubramanian Chandrasekaran

Dell Virtualization Solutions Engineering

<http://www.dell.com/virtualization/businessready>



Information in this document is subject to change without notice.

© Copyright 2010 Dell Inc. All rights reserved.

Reproduction in any manner whatsoever without the written permission of Dell Inc. is strictly forbidden.

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

Dell, the *DELL* logo, *PowerEdge*, *PowerConnect*, and *OpenManage* are trademarks of Dell Inc. *Microsoft* is a registered trademark of Microsoft Corporation; *VMware*, *vSphere*, *vCenter*, and *VMotion* are registered trademarks or trademarks (the "Marks") of VMware, Inc. in the United States and/or other jurisdictions. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

Table of Contents

1	Introduction	5
2	Audience and Scope	5
3	Overview	5
3.1	VMware vSphere 4	6
3.2	Dell PowerEdge Blade Servers	7
3.3	Dell/EMC CX4-120 Fibre Channel Arrays	8
4	Specification	10
5	Reference Architecture	12
6	Understanding the Dell Blade Architecture	13
7	Network Architecture	15
7.1	Traffic Isolation using VLANs	15
7.2	Load Balancing and Failover	16
7.3	Uplinks	17
7.4	Inter Switch Links	17
7.5	vNetwork Distributed Switch	17
7.6	VMware Traffic Shaping in Virtual Switches	17
7.7	VMware Fault Tolerance	18
8	Storage Architecture	18
8.1	Storage Connectivity	18
8.2	PowerPath/VE Multipath Plugin	19
8.3	Volume Size Considerations	20
8.4	Array RAID Configuration	20
9	Scalability	21
9.1	Scaling Compute Resources	21
9.2	Scaling Storage Resources	21
10	Management	21
10.1	Virtualization Management	22
10.2	Systems Management	23
10.3	Storage Management	24
10.4	Switch Management	24
11	Deploying and Configuring the Solution	24
12	References	27
13	Feedback	27

Table of Figures

Figure 1: Overview of Dell Blade Server, Dell/EMC SAN, and VMware vSphere	6
Figure 2: CX4-120 Physical View	9
Figure 3: Reference Architecture for a Large Configuration	13
Figure 4: Adapter and I/O Modules Connections in Chassis for M610	14
Figure 5: Overview of Network Configuration	14
Figure 6: LAN Configuration	15
Figure 7: Storage Connectivity	19
Figure 8: Management Components	22
Figure 9: VMware vCenter (High Availability) Configuration	23

1 Introduction

Business Ready Configurations for Virtualization are pre-configured, easy to order, bundled solutions that include Dell™ PowerEdge™ servers, storage and networking combined with industry leading hypervisors. This white paper provides a detailed reference architecture for VMware vSphere 4 ESXi with Dell PowerEdge blade servers, Dell/EMC CX4 Storage array, Brocade Fibre Channel switches and PowerConnect network switches. Based on extensive engineering work in architectural design and certification, customers can quickly and confidently deploy these engineering proven architectures into production environments, thereby helping to eliminate much of the costly and time consuming trial and error often encountered during complex infrastructure design and implementation. This solution provides scalability and manageability with no single point-of-failure. Network architectures, storage configurations, and best practices necessary for deploying and configuring vSphere are discussed in detail.

2 Audience and Scope

The intended audience for this white paper is IT administrators, IT managers, and channel partners who are planning to deploy VMware virtualization using Dell PowerEdge 610 blade servers, Dell/EMC CX4 Storage Arrays and 10Gb Ethernet network environments. This white paper provides an overview of the recommended servers, storage, networking, software, and services. It can be used to plan, scope, and procure the required components to set up a virtualization infrastructure. Based on customer requirements, further customization of the recommended architecture may be required.

3 Overview

Figure 1 provides a high-level product overview of VMware vSphere, Dell PowerEdge blade servers, and Dell/EMC CX4 arrays.



VMware vSphere 4

- VMotion
- VMware HA, DRS and DPM
- Distributed Virtual Switch
- VMware Fault Tolerance

Dell PowerEdge Blade Servers

- Energy efficient PowerEdge M1000e enclosure
- 11th generation M610 blade server
 - Intel Xeon 5500/5600 processor
 - Intel QuickPath memory technology
 - Unified Server Configurator
- Flex Address
- CMC and iKVM for enclosure management

Dell CX4-120

- Two 8Gbps Fibre Channel Ports per SP
- Front end expansion using Ultra-Flex modules
- PowerPath/VE for vSphere
- Virtualization aware NaviSphere

Figure 1: Overview of Dell Blade Server, Dell/EMC SAN, and VMware vSphere

3.1 VMware vSphere 4

VMware vSphere 4 includes the hypervisor (either ESX or ESXi) as well as an optional vCenter Server which is responsible for configuring, deploying, and managing VMware hosts. Capabilities for the ESX/ESXi *Enterprise Plus* license level include:

VMware VMotion: VMware VMotion technology provides real-time migration of running virtual machines from one server to another with no disruption or downtime. VMotion can be used to achieve manual load balancing and near-zero downtime maintenance.

VMware Distributed Resource Scheduler (DRS): VMware DRS technology enables VMotion to automatically achieve load balancing according to resource requirements.

VMware Distributed Power Management (DPM): When virtual machines in a DRS cluster need fewer resources, such as during nights and weekends, DPM consolidates workloads onto fewer servers and powers off the rest to reduce power consumption. When virtual machine resource requirements increase (such as when users log into applications in the morning), DPM brings powered-down hosts back online to ensure service levels are met.

VMware High Availability (HA): VMware HA provides high availability at the virtual machine level. Upon server failure, VMware HA automatically re-starts virtual machines on other physical servers.

VMware Fault Tolerance: VMware Fault Tolerance provides continuous availability for applications in the event of server failures by creating a live shadow instance of a virtual machine that is in virtual lockstep with the primary instance.

VMware vNetwork Distributed Switch: VMware's vNetwork Distributed Switch spans many ESX/ESXi hosts and aggregates networking to a centralized cluster level. The vNetwork Distributed Switch abstracts configuration of individual virtual switches and enables centralized provisioning, administration and monitoring through VMware vCenter Server.

VMware Update Manager: VMware Update Manager automates patch management; enforcing compliance to patch standards for VMware ESX/ESXi hosts, as well as Microsoft or Linux virtual machines. Secure patching of offline virtual machines, as well as automated snapshots and rollback capabilities, helps to reduce risk. Integration with DRS allows for non-disruptive VMware ESX/ESXi patching with little-to-no administrative overhead.

VMware Storage VMotion: VMware Storage VMotion enables real-time migration of running virtual machine disks from one storage array to another with no disruption or downtime. It minimizes service disruptions due to planned storage downtime previously incurred for rebalancing or retiring storage arrays.

Host Profiles: Host Profiles standardize and simplify the deployment and management of VMware ESX host configurations. It captures and stores validated configurations including networking, storage and security settings. It also makes it easy to monitor host compliance.

For more information on VMware vSphere, see <http://www.vmware.com/products/vsphere/>.

3.2 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: 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.

I/O Modules: The enclosure provides three redundant fabrics using six I/O modules. The modules can be populated with Cisco Catalyst® blade switches, Dell PowerConnect™ switches, and pass-through modules. Fibre Channel and InfiniBand® switch modules are also supported.

Chassis Management: The Dell PowerEdge M1000e blade enclosure features an embedded Chassis Management Controller (CMC) module enabling IT administrators to manage multiple enclosures from a single console. It is one of the only blade solutions to include an integrated KVM switch enabling easy setup and deployment. Through the CMC, the enclosure supports FlexAddress technology which enables the blade enclosure to lock the World Wide Names (WWN) of the Fibre Channel 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 Fibre Channel 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 iDRAC Enterprise in 11th-generation Dell PowerEdge blade servers. It includes 1 GB of managed and persistent storage that embeds systems management features directly on the server, thus eliminating the media-based delivery of system management tools and utilities previously needed for systems management. 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 blade servers, see www.dell.com/blades.

3.3 Dell/EMC CX4-120 Fibre Channel Arrays

Fourth generation Dell/EMC CX4 arrays deliver new levels of performance, flexibility, ease-of-use, data management, and resource utilization. The Dell/EMC CX4 family forms the core of a scalable enterprise infrastructure that is easy to deploy, manage, and upgrade. The CX4-120 is ideal for heavy transaction processing such as high speed large block data transfers, backup-to-disk, audio/video streaming, replication and departmental applications. Figure 2 shows a physical layout of the CX4-120 including the storage processors, the standby power supplies, and the Disk-Array Enclosures (DAEs). CX4-120 Features:

- Supports up to 120 hard drives
- Stores up to 72TB of raw storage capacity with Fibre Channel drives or 235TB with SATA II drives
- Customized connectivity with up to 256 hosts in a SAN
- All of these connectivity options are available on the same array: 8Gbit Fibre Channel (FC), 4Gbit FC, 10Gbit iSCSI and 1Gbit iSCSI

UltraFlex™ Modular I/O technology enables easy addition of more ports to the array, either Fibre Channel (FC) or iSCSI, as needs change. The innovations built into the new Dell/EMC CX4 arrays extend beyond just flexible I/O to include virtual provisioning, support for enterprise flash drives (EFD), enhanced tiering options, features, such as drive spin down and variable speed fans that enable greener storage by helping to reduce power consumption, and design improvements for performance and availability. In addition, management and reporting features on the CX4 have been enhanced to help reduce the time and effort required for managing storage in virtualized server environments.

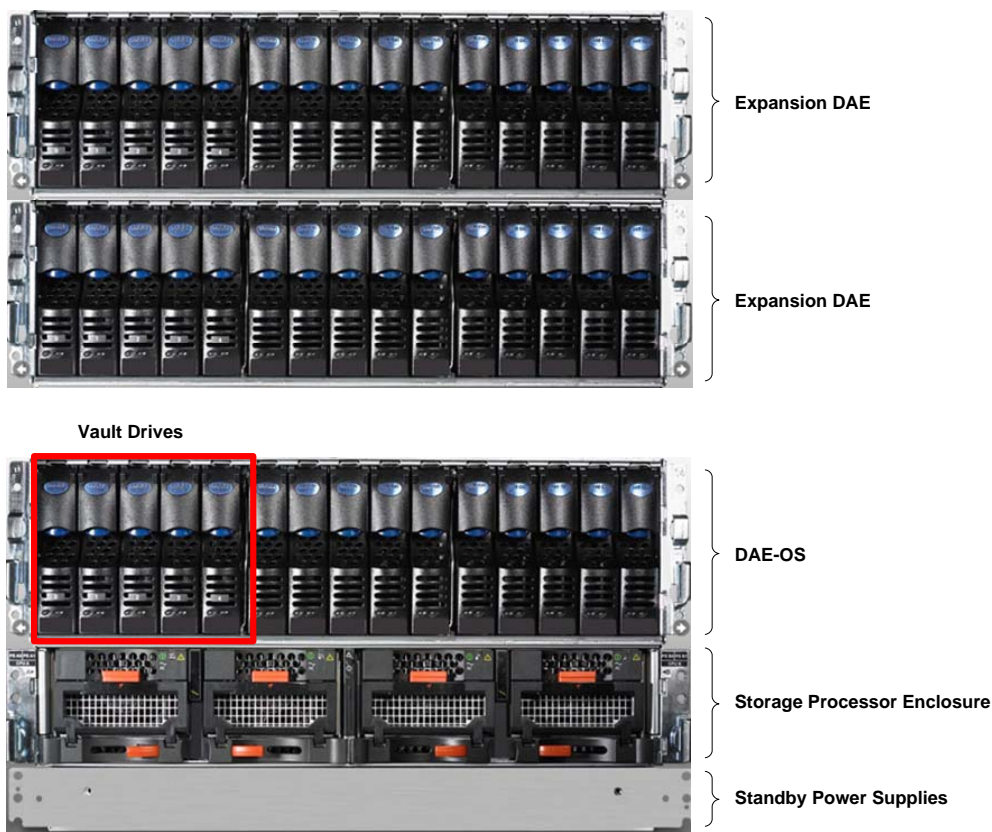


Figure 2: CX4-120 Physical View

Cost Effective Scalability

The CX4-120 array can scale from 5 to 120 drives; connect to as many as 256 highly available hosts in a SAN and store up to 72TB of data. The CX4 helps ease growing pains allowing you to easily add more capacity, whether it is 10K or 15K FC, 7.2K or 5.4K (energy efficient) SATA or flash drives, as your storage needs change. In addition, the new enterprise flash drives (EFD) option adds a new level of performance for critical applications where 1 EFD can equal the throughput of up to 30 Fibre Channel drives. Plus, if you outgrow your CX4-120, data-in-place upgrades are available to move up to an array with higher capacity and performance within the CX4 family.

4 Specification

The following table provides the specifications for three pre-configured solutions:

Table 1 Solution Specification

	Entry Configuration	Medium Configuration	Large Configuration
Solution Summary			
Solution ID	1065337.1	1065372.1	1065381.1
Blade Chassis	(1) M1000e with (2) PowerConnect M8024 10 Gigabit switch I/O modules and (2) Brocade M5424 Fibre Channel Switch modules	(1) M1000e with (2) PowerConnect M8024 10 Gigabit switch I/O modules and (2) Brocade M5424 Fibre Channel Switch modules	(1) M1000e with (2) PowerConnect M8024 10 Gigabit switch I/O modules and (2) Brocade M5424 Fibre Channel Switch modules
ESX Blade Server	(4) x M610 with 48GB memory in each blade	(6) x M610 with 48GB memory in each blade	(10) x M610 with 48GB memory in each blade
Storage Device	(1) CX4 – 120 with (10) 450GB 15k drives	(1) CX4 – 120 with (25) 450GB 15k drives	(1) CX4 – 120 with (40) 450GB 15k drives
Management Blade Server	(1) x M610 with VMware vCenter Server	(1) x M610 with VMware vCenter Server	(1) x M610 with VMware vCenter Server
Chassis Configuration			
I/O module for A1	Ethernet Pass-through Module	Ethernet Pass-through Module	Ethernet Pass-through Module
I/O module for B1	PowerConnect M8024	PowerConnect M8024	PowerConnect M8024
I/O module for C1	PowerConnect M8024	PowerConnect M8024	PowerConnect M8024
I/O module for A2	None	None	None
I/O module for B2	Brocade M5424	Brocade M5424	Brocade M5424
I/O module for C2	Brocade M5424	Brocade M5424	Brocade M5424
Management	Redundant Chassis Management Controllers (CMC)	Redundant Chassis Management Controllers (CMC)	Redundant Chassis Management Controllers (CMC)
KVM	Integrated Avocent keyboard, video, and mouse (iKVM) switch	Integrated Avocent keyboard, video, and mouse (iKVM) switch	Integrated Avocent keyboard, video, and mouse (iKVM) switch
ESX Blade Server Configuration			
Blade Server Model	M610	M610	M610
Processor	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache
Memory	48 GB (12 x 4 GB, DDR3)	48 GB (12 x 4 GB, DDR3)	48 GB (12 x 4 GB, DDR3)

Reference Architecture for VMware vSphere 4 with PowerEdge Blade Servers and Dell/EMC CX4

Add-in Controllers	1 Broadcom® 57711 Dual Port 10 GbE I/O Card for M-Series Blades 1 Dual Port Emulex LPe1205-M Adapter	1 Broadcom 57711 Dual Port 10 GbE I/O Card for M-Series Blades 1 Dual Port Emulex LPe1205-M Adapter	1 Broadcom 57711 Dual Port 10 GbE I/O Card for M-Series Blades 1 Dual Port Emulex LPe1205-M Adapter
Local storage and controller	Diskless configuration	Diskless configuration	Diskless configuration
Hypervisor	VMware ESXi 4.0 U1 - Enterprise Plus	VMware ESXi 4.0 U1- Enterprise Plus	VMware ESXi 4.0 U1 - Enterprise Plus
Storage Configuration			
Storage Device	(1) Dell EMC CX4-120 with two SPEs	(1) Dell EMC CX4-120 with two SPEs	(1) Dell EMC CX4-120 with two SPEs
Drives	(10) 450GB 15K FC drives	(25) 450GB 15K FC drives	(40)450GB 15K FC drives
Storage Capacity	18 Terabyte capacity	24.75 Terabyte capacity	45 Terabyte capacity
PowerPath Licenses	Includes (8) socket licenses	Includes (12) socket licenses	Includes (20) socket licenses
Management Blade Server			
Blade Server Model	(1) M610	(1) M610	(1) M610
Processor	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache	(2) x Intel Xeon (Nehalem) E5520, 2.26Ghz, 8M Cache
Memory	8 GB	8 GB	8 GB
Network Controller	(1) Broadcom 57711 Dual Port 10 GbE I/O Card for M-Series Blades	(1) Broadcom 57711 Dual Port 10 GbE I/O Card for M-Series Blades	(1) Broadcom 57711 Dual Port 10 GbE I/O Card for M-Series Blades
Virtualization Management	VMware vCenter 4.0, 3Yr Upgrade Subscription	VMware vCenter 4.0, 3Yr Upgrade Subscription	VMware vCenter 4.0, 3Yr Upgrade Subscription
Software and Services			
Services	<ul style="list-style-type: none"> 3 Year ProSupport for IT and Mission Critical 4HR 7x24 Onsite Pack M1000e Blade Chassis On-site installation service 	<ul style="list-style-type: none"> 3 Year ProSupport for IT and Mission Critical 4HR 7x24 Onsite Pack M1000e Blade Chassis On-site installation service 	<ul style="list-style-type: none"> 3 Year ProSupport for IT and Mission Critical 4HR 7x24 Onsite Pack M1000e Blade Chassis On-site installation service

How to order: To order one of the solutions discussed in this paper, contact your Dell Sales Representative with the *Solution ID* provided in the table above. Customization can be made to these configurations to fit specific needs, but may invalidate some of the best practices discussed in this paper.

These solutions can also be directly ordered online:
<http://www.dell.com/virtualization/businessready>.

5 Reference Architecture

This section describes the high-level reference architecture for the solution.

Design Principles

The following principles were used to design this architecture.

1. **Optimal hardware configuration for virtualized workloads:** Each solution is designed with an optimal hardware configuration to support virtualized workloads. Each blade server is configured with sufficient memory and network adapters required for virtualization.
2. **Redundancy with no single point-of-failure:** Redundancy is incorporated in every aspect of the solution, including server availability features, networking and storage.
3. **Scalability:** The solutions are highly scalable and are based on the architecture of the M1000e chassis and Dell/EMC CX4 storage arrays. Guidelines to deploy additional blade servers and storage to increase the compute and storage capacity respectively are included.
4. **Isolated, redundant, and high-performance network design:** One of the key aspect of this architecture is the 10 Gb Ethernet network which is designed to support isolation of various traffic types required in a virtualized environment. This network is designed to have no single point-of-failure and have optimal performance through NIC teaming and load balancing.
5. **Integration into an existing data center:** The architectures assume that there is a existing 10 Gb infrastructure network. In addition, it is assumed that there is network infrastructure in place to support the out-of-band hardware management network.

In all three configurations, one blade is used for management (VMware vCenter) with the remainder devoted to an ESXi Cluster. Figure 3 provides a high-level reference architecture for the large configuration. Entry and Medium configurations have similar reference architecture with varying number of blades and storage arrays.

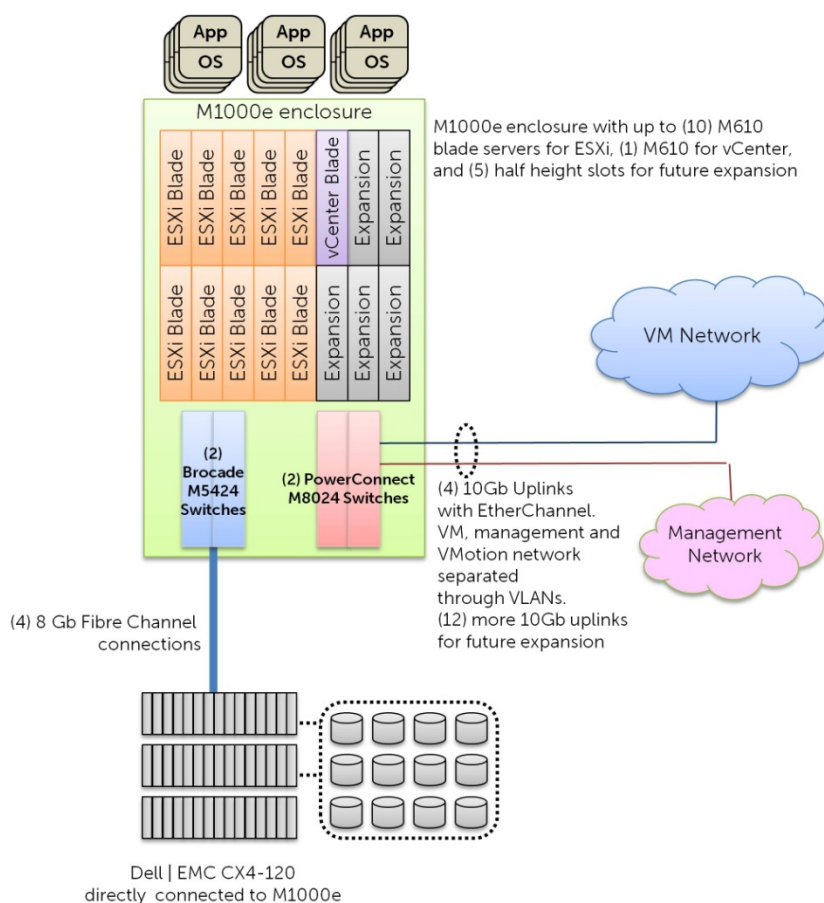


Figure 3: Reference Architecture for a Large Configuration

6 Understanding the Dell Blade Architecture

The Dell blade chassis has three separate fabrics referred to as A, B, and C¹. Each fabric has two I/O modules, making a total of six I/O modules 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, or a Fibre Channel switch². In this solution, fabric B contains PowerConnect M8024 10 Gigabit Ethernet I/O modules. Fabric C is populated with Brocade M5424 Fibre Channel Switches.

Each M610 blade server has a dual port onboard 1 GbE network adapter and two optional dual port mezzanine I/O card slots. For this solution, Broadcom 57711 10GbE mezzanine cards populate the Fabric B slots and Emulex LPe1205-M mezzanine cards occupy the Fabric C slots. Figure 4 illustrates how these adapters are connected to the I/O modules in the chassis.

¹ <http://support.dell.com/support/edocs/systems/pem600/en/hom/HTML/about.htm#wp1212100>

² Fabrics B and C support InfiniBand, which is outside the scope of this white paper.

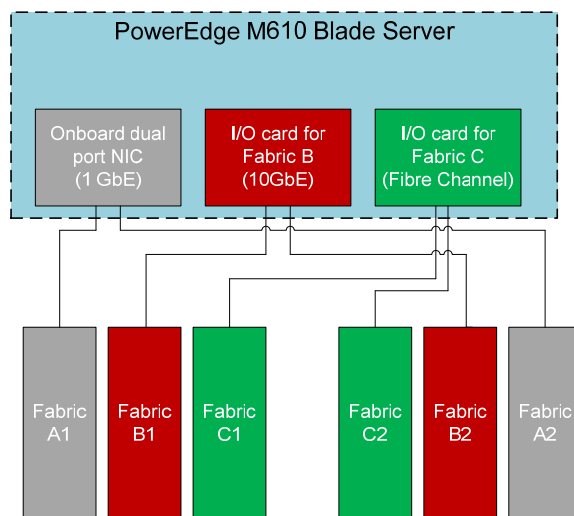


Figure 4: Adapter and I/O Modules Connections in Chassis for M610

The two onboard Broadcom NetXtreme II™ 5709 1 GbE NICs embedded in the PowerEdge M610 blade servers connect to Fabric A, which has been configured with a pass-through module. These ports can be used for installation and troubleshooting purposes. Figure 5 provides an overview of the chassis configuration for this solution.

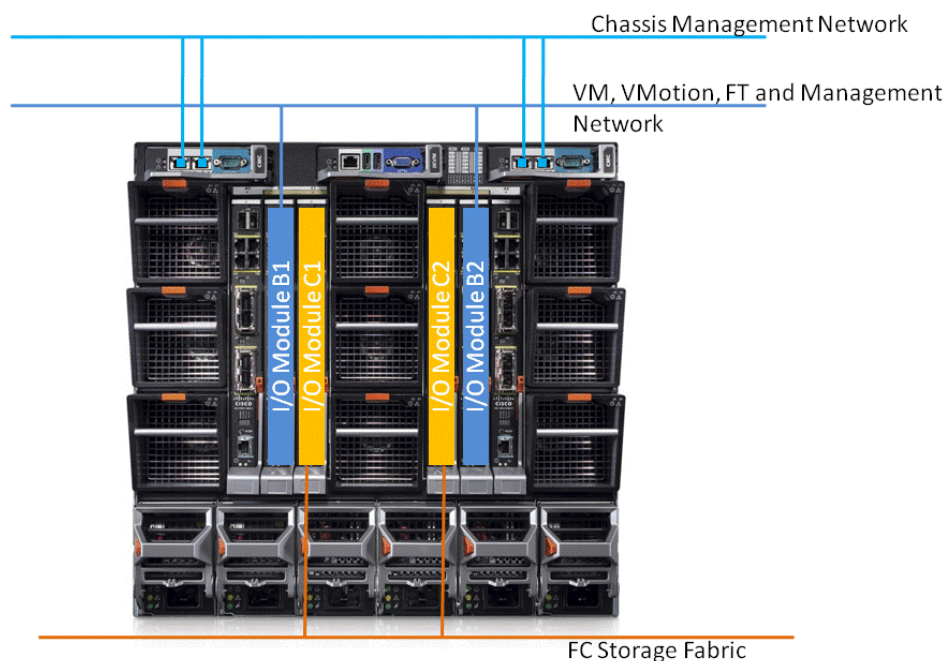


Figure 5: Overview of Network Configuration

7 Network Architecture

This section provides the network best practices for implementing VMware vSphere 4 on Dell blade servers over a 10 Gigabit environment.

vSphere network traffic is typically comprised of three types: virtual machine traffic, management traffic, and VMotion traffic. A virtual switch (vSwitch0) is created and two 10 GbE physical network adapters are connected as uplinks to this virtual switch. This creates a team of two network adapters, enabling NIC failover and load balancing. These network adapters are in turn connected to PowerConnect M8024 I/O modules in slots B1 and B2. Three port groups are created on vSwitch0: one for the virtual machine network, another for the VMotion network, and the third for the service console (for ESX) or management (for ESXi). When virtual machines are created, the virtual network adapters are connected to the virtual machine network port group.

Figure 6 illustrates the virtual switch configuration with port groups, and how the virtual switch is connected to physical network adapters and in turn to the I/O modules.

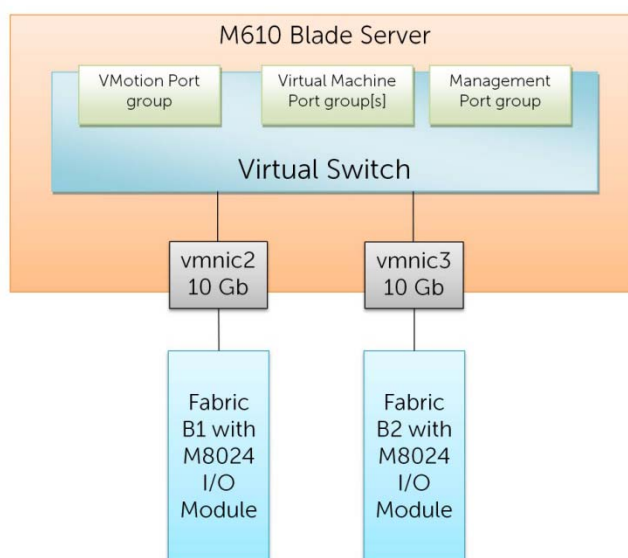


Figure 6: LAN Configuration

7.1 Traffic Isolation using VLANs

The traffic on the LAN is separated into three VLANs; one VLAN each for management, VMotion, and virtual machine traffic. Network traffic is tagged with the respective VLAN ID for each traffic type in the virtual switch. Routing between VLANs is dependent on specific customer requirements and is not covered by this paper. If desired, the M8024 modules can be configured to provide the routing function.

Trunking must be enabled for all ports on the two M8024 switches so that all the VLANs can share the same physical connection. VMotion traffic is unencrypted, so it is highly recommended to isolate this traffic from the rest of the network³.

7.2 Load Balancing and Failover

Load balancing enables sharing of network traffic between the physical network ports in a team, thereby generating higher throughput. The VMware virtual switch has three options to configure load balancing:

- **Route based on the originating virtual switch port ID (default configuration):** This is also known as vSwitch port-based load balancing. For outbound traffic, the physical adapter port is selected based on the *hash of the virtual port*. This means that a given virtual network adapter will use only one physical adapter at any given time to transmit network packets unless a failover to another adapter has occurred. Replies are received on the same physical adapter. This configuration offers the lowest use of CPU resources for routing.
- **Route based on source MAC hash:** Here, the physical adapter is selected for transmit based on a *hash on the source MAC address*. This means that a given virtual network adapter will use only one physical adapter at any given time to transmit network packets unless a failover to another adapter has occurred. This mode is rarely used in practice. Replies are received on the same physical adapter.
- **Route based on IP hash:** Here, the physical adapter is selected for transmit based on a hash of the source and destination IP address. This method distributes traffic the most evenly, but comes at the cost of very high CPU utilization. Each packet transitioning across the vSwitch must be inspected by the VMkernel to make routing decisions. Link Aggregation should be used with this setting and the attached switch must be configured for 802.3ad link aggregation. The physical switch load balancing algorithm will then determine which physical adapter is selected to receive the packets. To achieve proper load balancing, both the virtual switches and the physical switches need to be configured. Note that the virtual switch and the physical switch hashing algorithms work independent of each other. This is the only load balancing method that supports Link Aggregation. Since this requires Link Aggregation and both *vmnic2* and *vmnic3* connect to different I/O modules, it is not possible to configure this option in the recommended architecture.

The default load balancing configuration of route based on the originating virtual switch port ID is recommended. Multiple VMs will use different physical adapter ports to transmit and receive traffic. Service console (for ESX) or management network (for ESXi) has only one port ID or MAC address. Hence, the service console or management network will use only one physical adapter port for communicating, unless there is a failover to another physical adapter port.

All the physical network adapters will be configured as *active* by default when load balancing is enabled.

For VMotion, it is recommended to configure one physical adapter (for example, *vmnic2*) as the active adapter and the other adapter is configured as standby for all the blade servers. This ensures VMotion traffic will use only one I/O module and the traffic can be contained within the chassis, unless there is a failover.

³ <http://blogs.vmware.com/security/2008/02/keeping-your-vm.html>

7.3 Uplinks

Each of the two M8024 I/O modules has eight 10 Gb Ethernet ports. Some of these ports can be used for Inter Switch Links (ISL). The rest of the ports can be used for connection to the existing network infrastructure in the datacenter.

Example: two 10 Gb ports in each PowerConnect I/O module can be used as uplinks. The two ports must be configured as a Link Aggregation to provide redundancy and load balancing.

VLAN and routing must be configured on these uplinks depending on the existing infrastructure requirements.

7.4 Inter Switch Links

For a stand-alone configuration without uplinks, the two I/O modules can be connected to each other using ISLs. The number of ISL connections will depend on the network requirements of the infrastructure. Using two ISLs will provide bandwidth of 20 Gigabits per second (Gbps) between the I/O modules which should be sufficient for most cases. The two links are configured using Link Aggregation.

When connecting to existing infrastructure switches, using ISL between I/O modules will potentially create loops for using Spanning Tree Protocol (STP). Care must be taken to provide routing in the existing infrastructure for VMotion and FT traffic. This depends on the existing infrastructure and outside the scope of this paper.

7.5 vNetwork Distributed Switch

If the Enterprise Plus license level is available, it is highly recommended to use the vNetwork Distributed Switch in place of the standard virtual switch. The vNetwork Distributed Switch provides a centralized management interface for configuring virtual networks across multiple ESX/ESXi servers in the cluster.

Additional features are supported in the vNetwork Distributed Switch:

- Network VMotion (Port state follows VM)
- Bi-directional Traffic Shaping
- Private VLAN

For more information on the vNetwork Distributed Switch:

http://www.vmware.com/files/pdf/VMW_09Q1_WP_vSphereNetworking_P8_R1.pdf

7.6 VMware Traffic Shaping in Virtual Switches

Traffic shaping is a feature available in virtual switches that can be used for rate limiting. The standard vSwitch supports traffic shaping on outbound traffic. The vNetwork Distributed Switch supports traffic shaping on both inbound and outbound activity. Traffic shaping has three parameters: average bandwidth, peak bandwidth and burst size. Refer to VMware's [ESXi Configuration Guide](#) (section "Traffic Shaping") for more information on these parameters.

Traffic shaping can be used to rate limit different LAN traffic types: VM traffic, VMotion traffic, and management traffic. Some important things to consider when implementing traffic shaping:

- Traffic shaping does not guarantee bandwidth to a certain traffic type, but provides a limit on the maximum bandwidth that can be transmitted.
- Traffic shaping can be used to limit the bandwidth of the non-critical traffic. This will enable the critical traffic to get sufficient bandwidth.
- Traffic shaping can be configured per port group on the virtual switch. The same traffic shaping configuration will be applied to all of the ports within that port group. (For example, if the Virtual Machine port group is set to an average bandwidth of 2 Gbps, all the virtual network adapters connecting to that port group will have traffic shaping set to an average bandwidth of 2 Gbps).
- It is highly recommended to set the traffic shaping parameters to less than 4Gbps for consistent behavior. Refer to VMware documentation and Knowledge Base articles for more information.

7.7 VMware Fault Tolerance

VMware Fault Tolerance (FT) can be implemented in the proposed solution with a few modifications. A new port group can be created in the virtual switch for the FT logging network to which traffic shaping can be applied.

In section **Error! Reference source not found.**, we recommended configuring *vmnic2* as the active adapter and *vmnic3* as standby in all of the blade servers for VMotion traffic. Similarly, for VMware FT, we recommend configuring *vmnic3* as the active adapter, and leaving *vmnic2* as standby. This will ensure VMotion and VMware FT use different physical adapters for communication.

It is important to note the bandwidth requirement for the FT logging network using the following formula:

$$\text{VMware FT logging bandwidth} \approx (\text{Avg disk reads (MB/s)} \times 8 + \text{Avg networkinput (Mbps)}) \times 1.2 \text{ [20\% headroom]}$$

Requirements for VMware FT can be found in the VMware white paper: [VMware Fault Tolerance Recommendations and Considerations on VMware vSphere 4](#).

8 Storage Architecture

In this section, we discuss the best practices for configuring the FC SAN in ESX/ESXi.

8.1 Storage Connectivity

Each Brocade M5424 I/O module contains eight 8 Gbps ports. Each Service Processor (SP) on the CX4-120 array has four ports. For redundancy, each SP is connect to both of the I/O modules in the blade chassis. Connections from each I/O module are split between both of the SPs to allow for failover in case of SP failure. Figure 7 illustrates how a Dell | EMC CX4-120 array is connected to the Dell PowerEdge M1000e chassis. Additional Fibre Channel ports can be added to the array using UltraFlex I/O modules.

The storage array for this reference architecture is directly connected to the Brocade blade Fibre channel switches located on the back of the M1000e chassis. If your environment requires the use of intermediate Fibre Channel switches this configuration will need to be adjusted.

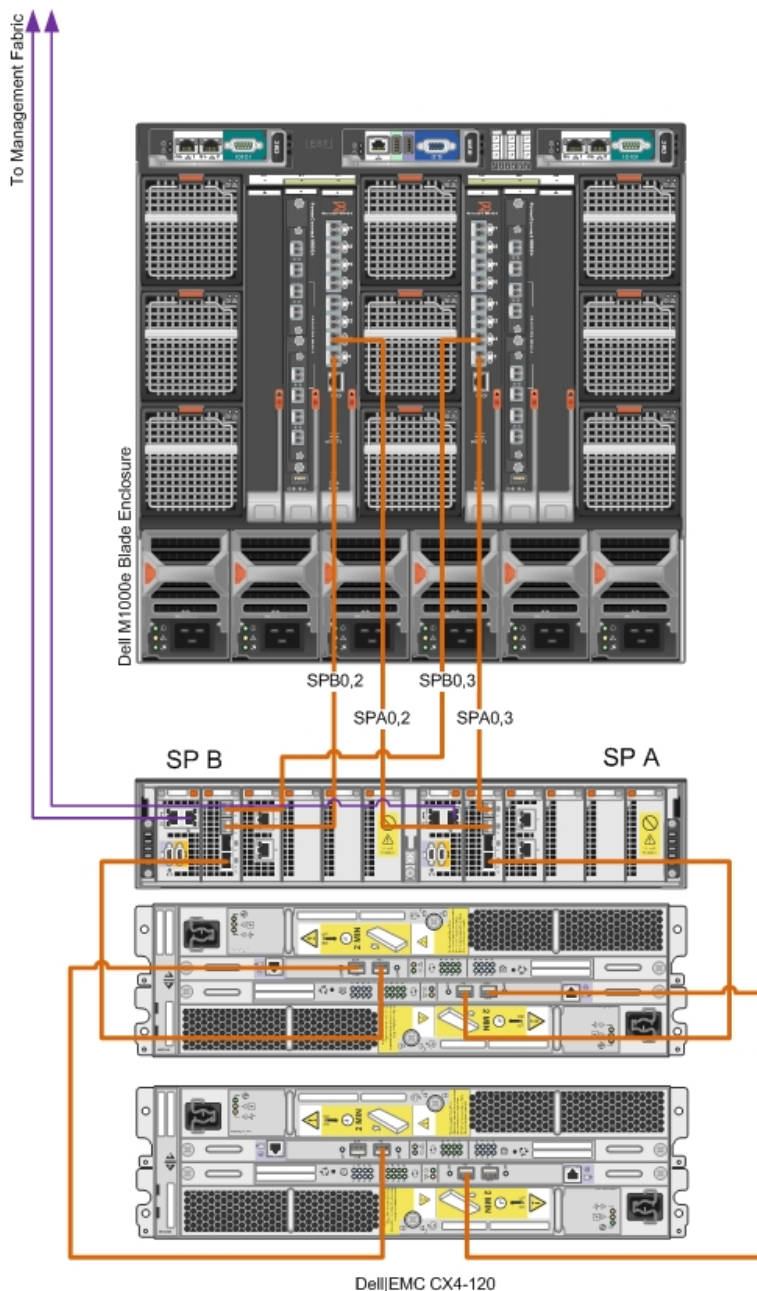


Figure 7: Storage Connectivity

8.2 PowerPath/VE Multipath Plugin

VMware introduced vStorage APIs in vSphere 4, which enabled storage vendors to develop Multipath Extension Modules (MEM) that plug into vSphere. The MEM modules can provide advanced multipathing capabilities for storage arrays. PowerPath/VE is a MEM that works with ESX/ESXi to provide enhanced path management capabilities for Dell/EMC storage. The Enterprise Plus license level is required to support 3rd party plug-ins for more advanced path management. PowerPath/VE installs in ESX/ESXi and provides the following capabilities:

- Dynamic load balancing: PowerPath/VE enables utilization of all of the available paths to access the storage device, dynamically load balancing between these paths.
- Dynamic path failover and recovery: PowerPath/VE redistributes I/O traffic on path failures and ensures that the failed path is not used. Once the path is available, the I/O traffic is redistributed again to utilize the restored path.
- Device prioritization: Single or several devices can be set to higher priority to improve their I/O performance at the expense of lower priority devices
- Automated performance optimization: PowerPath/VE automatically identifies the type of storage array and sets the highest performing optimization mode by default. For the Dell/EMC CX4 storage array, the mode is set to CLAROpt (CLARiiON Optimized)
- Performance monitoring: Performance of all the I/O paths can be reported using PowerPath/VE remote server (*rpowermt*).

For more information on PowerPath/VE refer to the white paper: [EMC PowerPath/VE for VMware vSphere](#).

PowerPath/VE requires separate licensing for each ESX/ESXi host. There are two types of licenses for PowerPath/VE software: *served* and *unserved*. Served license utilizes a license server to store, distribute, and manage the license. Unserved licenses are tied to each ESX/ESXi server and are administered using remote PowerPath server (*rpowermt*). Both the license server and the remote server can be installed as virtual machines. For more information on licensing refer to the [EMC PowerPath Licensing Guide](#).

8.3 Volume Size Considerations

Volume sizes depend on the customer environment and the type of workloads. Before you use the Dell | EMC CX4 in the reference configuration, the array must be divided into RAID Groups. RAID groups are made up of physical hard drives with a certain type of RAID configuration. Before they can be used, these disks groups must be further divided into logical unit numbers or LUNs, and then assigned to Storage Groups.

LUNs must be sized to accommodate not only the virtual machine hard drive, but also the size of the virtual memory of the virtual machine with enough space for any snapshots of the virtual machine. Depending on the environment, you may decide to create multiple ~500 GB volumes with multiple VMs. It is important to include space for the Guest OS memory cache, snapshots, and VMware configuration files when sizing these volumes.

8.4 Array RAID Configuration

The storage array RAID configuration is highly dependent on the workload in your virtual environment. The Dell/EMC CX4 storage arrays support four RAID types – RAID 1, RAID 5, RAID 6, and RAID 10. The RAID configuration will depend on workloads and customer requirements. In general, RAID 10 provides the best performance at the expense of storage capacity. RAID 10 generally provides better performance in random I/O situations and requires additional overhead in the case of a drive failure scenario. RAID 5 provides the most storage capacity at the expense of a little lower performance and availability, and is not recommended since the availability of RAID 6.

9 Scalability

As workloads increase, the reference configuration can be scaled to provide additional resources. In this section, guidelines are provided for adding blade servers, blade chassis and storage arrays to the reference configuration.

9.1 Scaling Compute Resources

Adding blade servers: If the cluster has reached or is about to exceed its acceptable maximum resource utilization, additional blades can be added to the cluster. For quick addition of a new blade and for the cluster to be uniform, host profiles (a feature of Enterprise Plus) can be used. A host profile is created from an existing ESXi blade server on the same cluster. It will capture the networking, storage and security settings from the validated configuration. An ESXi blade server with basic vSphere installation can be added to the cluster first. A host profile is then applied to this newly added blade server.

Adding blade chassis: If the chassis is full of blades and more compute resources are needed, an additional blade chassis can be added to the solution. Each new chassis will be uplinked to the external network infrastructure.

9.2 Scaling Storage Resources

The CX4-120 can support up to 15 expansion disk enclosures with up to 120 total disks. Scaling beyond 120 disks requires a higher model of the CX4 series array be utilized such as the CX4-240, CX4-480 or CX4-960. Upgrades from a lower CX4 series array to a higher model is supported if the solution scales beyond initial expectations.

10 Management

This section describes the various components used to manage the proposed solution.

Figure 8 displays the various management components as they relate to each part of the architecture.

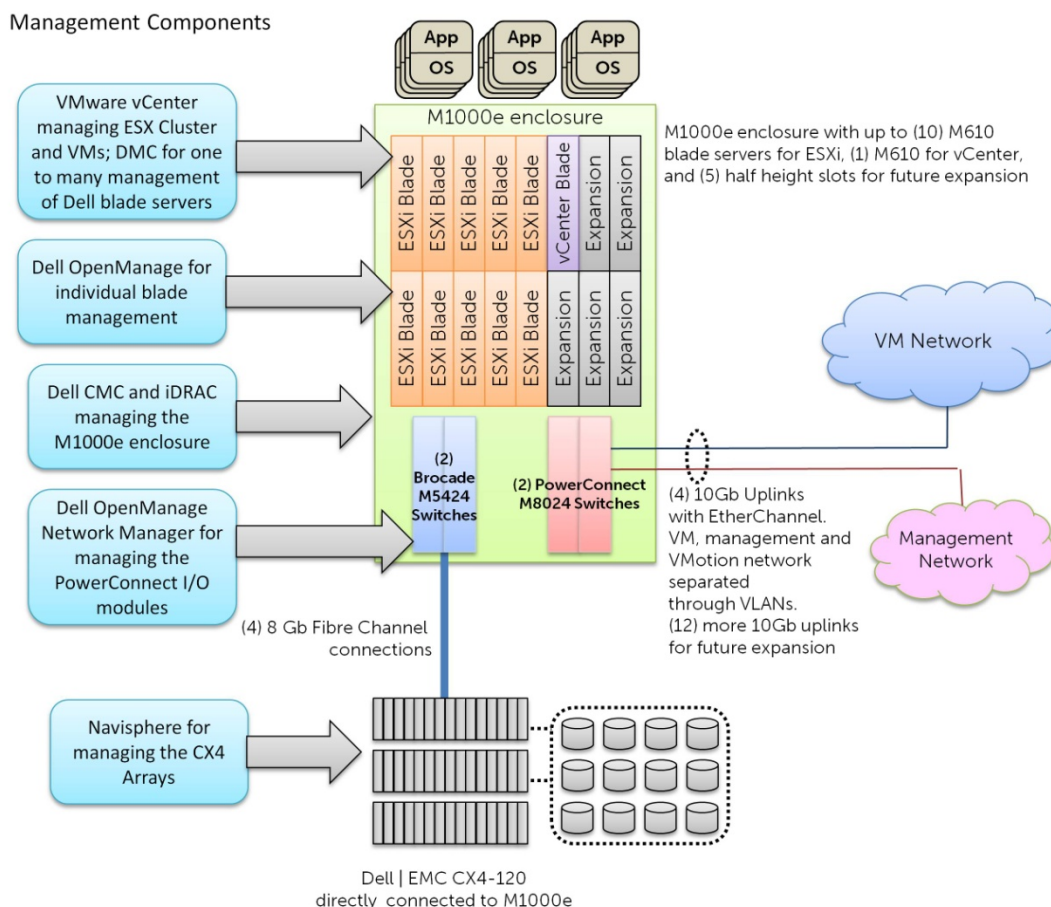


Figure 8: Management Components

10.1 Virtualization Management

10.1.1 VMware vCenter

VMware vCenter provides a single control and management center for VMware ESX/ESXi Servers. Capabilities include the ability to manage all vSphere hosts and virtual machines, utilize advanced features such as VMotion, HA, and DRS, monitor performance and resource optimization, and use standard templates to provision new servers in minutes. For more information on vCenter, see <http://www.vmware.com/products/vcenter/>.

Best Practices for vCenter Configuration

Install vCenter on the dedicated Management blade. It is highly recommended to use an external database for vCenter. After installation, it can easily be accessed remotely using VI Client.

For information on Setup and Configuration, see the VMware vCenter Guide at <http://www.vmware.com/products/vcenter-server/resource.html>.

vCenter Configuration- High Availability

Best practice recommendations for configuring VMware vCenter in a high availability configuration:

- Install two instances of VMware vCenter in a linked mode configuration on two separate virtual machines. Make sure that the virtual machines are on two separate ESX hosts to ensure maximum availability. This requires an additional management blade to be added to the configuration.
- Enable mirroring on the external database. Configure the ODBC client on the vCenter virtual machines, to reflect the primary and mirrored databases.
- If the databases are on virtual machines, make sure they are hosted on separate ESX hosts. This can also be done by adding a rule to the DRS configuration to ensure that the virtual machines are hosted on separate hosts at all times.

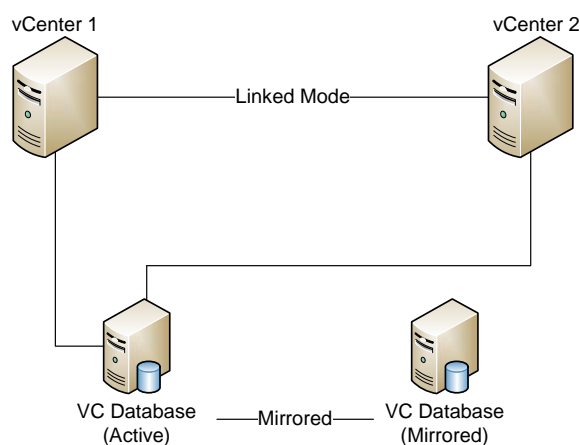


Figure 9: VMware vCenter (High Availability) Configuration

10.2 Systems Management

The following tools and software can be used to manage the hardware:

Dell OpenManage: Dell OpenManage Server Administrator (OMSA) can be installed on the VMware ESX/ESXi hypervisor and used to manage the hardware. For more information on Dell OpenManage and its capabilities, see www.dell.com/openmanage.

Deployment and change management using Lifecycle Controller: Dell PowerEdge M610 blade servers come with the Unified Server Configurator (USC). 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 Chassis Management Controller (CMC) provides a single, secure interface to manage the inventory, configuration, monitoring, and alerting for chassis components (iKVM, CMC), I/O modules, servers, and the integrated Dell Remote Access Controller (iDRAC). It also provides excellent real-time power management, monitoring, and alerting capabilities. The Dell M1000e 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.

Dell Management Console: The Dell Management Console (DMC), powered by Symantec, is a systems management solution designed to provide a central console offering features from basic hardware management to advanced enterprise functionality. The DMC creates a foundation for more advanced management functionality and provides a single view into the deployment, inventory, monitoring, and updating of your IT infrastructure - including servers, storage, desktops, notebooks, network devices, printers, and other non-Dell devices.

10.3 Storage Management

The Dell/EMC CX4 storage array can be used with EMC Navisphere Manager. Navisphere provides the following capabilities:

- Provision, monitor, and configure CX4 networked storage systems from a single console with Navisphere Manager.
- With Navisphere Analyzer, collect performance information in graphical form to identify performance patterns and trends before problems occur. Capabilities include trend analysis, reporting, and system capacity management.
- Script/automate common storage management tasks with the Navisphere Agent/CLI as your host-to-array communication bridge.
- Access and manage all Dell/EMC platform applications including Access Logix, MirrorView, SnapView, and SAN Copy - via the Navisphere suite.

10.4 Switch Management

Dell OpenManage Network Manager is a comprehensive, centralized network management solution designed to:

- Minimize the time, complexity and cost of network management
- Increase overall network efficiency
- Effectively manage corporate networks across platforms
- Maximize your IT investment

11 Deploying and Configuring the Solution

This section provides guidelines for the initial deployment and configuration of the solution.

Rack, Stack, and Cable:

1. Rack the chassis with blade servers and I/O modules.
2. Rack the Dell/EMC CX4 storage array.
3. Connect power to the chassis and storage.
4. Consider PDUs. This document provides detailed information on how to configure the PDUs: http://www.dell.com/downloads/global/products/pedge/en/pe_m1000e_selection_whitepaper.pdf
5. Cable servers and storage.
 - a. Cable the server and storage for Ethernet connectivity as per the architecture.
 - b. Connect the switches with the given configuration.

Configuring the Chassis

Configure the Chassis Management Controller (CMC) using the web interface or the front LCD panel on the chassis. By default, the CMC will get a DHCP address, which you can view from the LCD.

1. Use a static chassis management IP address (preferable).
2. Enable FlexAddressing.
3. Configure the iDRAC IP address for the blades.
4. Configure the PowerConnect Switch Management IP.

Configure Networking

1. Configure the network switch modules.
2. It is recommended that you create a configuration file for the network switches. These files are written in a sparse format using the range command for interfaces.

Configuring the Server

Configure the server either locally or by using the iDRAC Console Redirection remotely. Perform the following:

1. Enable VT in the BIOS.
2. For ESXi, it is recommended to have diskless servers.
3. Configure USC for self-update and firmware repository.
4. Boot the ESXi system.
5. Assign a password, management IP address, and DNS name to ESXi using the DCUI (Direct Console User Interface).

Configuring Storage

The Dell | EMC CX4 will require onsite setup initially. Once the initial setup has been completed, you will be able to follow these steps to configure the array.

1. Install/Run the Navisphere Initialization Utility on the management server.
2. Connect to the Navisphere Manager (browse to the IP address of one of the storage processors in a web browser).

Note: *Java Runtime Environment (JRE) is required to support Navisphere Manager*

3. Set Read/Write Cache settings on the array (recommend providing the maximum supported memory to write cache (775 MB) and splitting the remaining read cache between the two storage processors (486MB each)).
4. Enable Storage Groups on the array.
5. Configure the 8Gbit Fibre Channel ports as necessary.
6. Enable any optional array-based advanced features that were purchased (such as virtual provisioning).

Configuring vCenter

Installing SQL database is required before installing VMware vCenter. See the “ESXi Embedded and vCenter Server Setup Guide”,

http://www.vmware.com/pdf/vsphere4/r40/vsp_40_esxi_e_vc_setup_guide.pdf to install vCenter on the management server.

Deployment Steps

1. Create a cluster in vCenter with DRS and DPM features enabled.
2. Add the ESX servers to the cluster.
3. Place the servers in maintenance mode.
4. Configure the vNetwork Distributed Switches (if available) and add to the server.

12 References

VMware Documentation

- ESXi Configuration Guide
http://www.vmware.com/pdf/vsphere4/r40_u1/vsp_40_u1_esxi_server_config.pdf
- VMware Virtual Networking Concepts
http://www.vmware.com/files/pdf/virtual_networking_concepts.pdf
- What's new in vSphere Networking
http://www.vmware.com/files/pdf/VMW_09Q1_WP_vSphereNetworking_P8_R1.pdf
- VMware Fault Tolerance recommendations and considerations for vSphere
http://www.vmware.com/files/pdf/fault_tolerance_recommendations_considerations_on_vmw_vsphere4.pdf

PowerEdge Blade Server

- Blade Server Manual <http://support.dell.com/support/edocs/systems/pem/en/index.htm>

Fibre Channel Documentation

- EMC CX Series initial configuration
http://www.dell.com/downloads/global/products/pvaul/en/cxseries_initialconfig.pdf
- Brocade M5424 Documentation
<http://www.dell.com/content/products/productdetails.aspx/switch-brocade-m5424?c=us&cs=555&l=en&s=biz>
- EMC Clariion Practices for Best Performance and Availability
<http://www.emc.com/collateral/hardware/white-papers/h5773-clariion-perf-availability-release-28-firmware-wp.pdf>
- EMC PowerPath/VE for VMWare vSphere <http://www.emc.com/collateral/software/white-papers/h6340-powerpath-ve-for-vmware-vsphere-wp.pdf>

13 Feedback

We are interested in knowing your feedback to this document. If you identify any errors, including typographical, grammatical or architectural errors, or if you have any questions, please email us at solutionfeedback@dell.com.