

August 21, 2012
Jeremy Li

Notes taken from EMC Forum 2012

Credit: EMC Forum 2012, Long Beach Convention Center, CA

EMC (www.emc.com) held another successful one-day conference - EMC Forum 2012 on Aug. 21, 2012 at the Long Beach Convention Center. The Forum started with Welcome and Keynote by Jeremy Burton, EVP Product Operations and Marketing at 9:15 a.m., although nearly 40 to 50 attendees were still waiting for an onsite registration.

This year's theme is "**TRANSFORM: IT + BUSINESS + YOURSELF**". Click on the following link for more information.

<http://www.emc.com/campaign/global/forum2012/index.htm?pid=homepage-forum2012-06042012>

Jeremy mentioned that EMC introduced 41 and 42 new products in 2011 and 2012, respectively.

The presentation slides from his keynote, and other breakout sessions can be obtained by visiting the link ...

Below are a few Breakout Sessions I attended and notes I took

I. Active Active Datacenters, How Do I get There? Sponsor: WWT

John Berrett, National Technical Architect, World Wide Technology, Inc., (WWT) presented a session on Resilient Active/Active Datacenter (RAD) with a live demo to failover an Active/Active Datacenter to achieve six 9s objective via EMC/Cisco architecture with two second generation VPLEXs, as shown in the screenshot below:

What is the meaning of the 9's?

Disaster Avoidance - **99.9999%** Uptime (Six-Nines – only allows for 32 seconds or less of downtime per year!)

Mission Critical - 99.999% Uptime (Five-Nines - only allows for 5 minutes and 15 seconds or less of downtime per year!)

Business Critical - 99.90% Uptime (Three-Nines - only allows for 8 hours and 46 minutes or less of downtime per year!)

Business Important - 99% Uptime (Two-Nines - only allows for 3 days 15 hours and 40 minutes or less of downtime per year!)

EMC VNX Series is designed for five 9's, while VMAX is designed for six 9's. The VPLEX price tag is around millions.

ACTIVE/ACTIVE - Now a Distributed Reality



"RAD" Advantages

- Enables stretched clusters with zero RPO
- Automated recovery with near-zero RTO
- High availability within and across VPLEX Metro data centers
- *MOST importantly DR equipment is used*



COMMON ARCHITECTURE OF TODAY



Traditional challenges

- Server cycles required to mirror or cluster
- Application restart required after failover
- Remote replication RTO/RPO impact
- Network routing and different L2/L3 per site
- *Pay for expensive DR equipment that isn't used*



The VPLEX is Distributed, Dynamic, Smart and built for Storage Agnostic (e.g., a failover can occur between EMC VMAX and HP 3PAR storage). However, the

VPLEX only supports Fibre Channel (FC) protocol. Therefore, EMC Isilon storage will not support the VPLEX technology due to lacking of supporting the FC protocol. EMC VNX Series will not be able to support the VPLEX deployment at this time, even though it supports the FC protocol.

There are three types of VPLEX:

- VPLEX Local - Within a data center
- VPLEX Metro - AccessAnywhere at **synchronous** (RPO=0) distances

Two sites must have 10ms persistent link under VMware ESXi 5.0, an improvement from 5ms under VMware ESXi 4.1.

According to Cisco Webinar (Inside Cisco IT: Data Center Strategy (AMER), both datacenters - Metro Virtual Data Centers (MVDC) will act as primaries – Both DC1 and DC2 recovery will be done within the Metro Pair, while Active/Active services are designed for capable applications, and Active/Standby services are designed for databases with zero data loss option. The latency should be less than 1 ms and no more than 50 miles between two data centers.

- VPLEX Geo - AccessAnywhere at **asynchronous** distances

The VPLEX is a real-time cache coherent replication appliance and its key technology relies on Cisco **Overlay Transport Virtualization (OTV)**, which extends Layer 2 VLANs across multiple data centers over any network, as illustrated in the screenshot below.

With the OTV, the Ethernet traffic between sites is encapsulated in IP. This is called “MAC in IP”. The OTV is a MAC routing scheme where each Cisco Nexus 7000 switch maintains its MAC address table for every device across Cisco OTV domain. When an OTV edge device identifies a layer 2 frame targeted for a remote destination, it encapsulates the frame inside the IP packet, transmits it via a layer 3 network. When it arrives to the remote site, the edge device will unwrap the layer 2 frame and forward it to a final destination. MAC address table at each edge device are automatically populated and shared among all Cisco OTV devices (Dynamic encapsulation based on MAC routing table). No Pseudo-Wire or Tunnel state is maintained.

A Special Note on Cisco Nexus 7000 Switch:

1. The flagship Cisco Nexus 7000 Aggregation Layer switch uses its old chips from the Cisco 6500 series, which are heavy power consumption, especially on the 10GeE ports. Dell switches, formerly known as Force 10, has consumption of about 30 Watts per 10G port, while Cisco was 130 Watts per 10GeE port.

The Nexus 5000 switches via Cisco's acquisition are much better in the power consumption by relying on more up to date technology, although Dell's Force 10 stackable switches are still drawing less power than Nexus 5000.

2. Cisco, traditionally, requires a customer to use 2N redundancy, while Dell's Force 10 promotes N + 1 redundancy for more economical deployments to achieve a similar result.

Overlay Transport Virtualization (OTV)

Real Problems Solved by OTV

- ✓ Extension over any transport (IP, MPLS)
- ✓ Failure boundary preservation
- ✓ Site independence/isolation
- ✓ Resiliency/multi-homing
- ✓ Built-in end-to-end loop prevention
- ✓ Multi-site connectivity (inter- and intra-Data Center)
- ✓ Optimal multicast bandwidth utilization (no head-end replication)
- ✓ Operational simplicity
- ✓ Ease of site adds/drops

Overlay Transport Virtualization - OTV delivers a virtual L2 transport over any L3 Infrastructure

Overlay - A solution that is *independent of infrastructure technology* and services, flexible over various inter-connect facilities

Transport - Transporting services for *Layer 2 and Layer 3* Ethernet and IP Traffic

Virtualization - Provides *virtual connections (connections that are in turn virtualized and partitioned into VRFs, VLANs*

World Wide Technology, Inc.

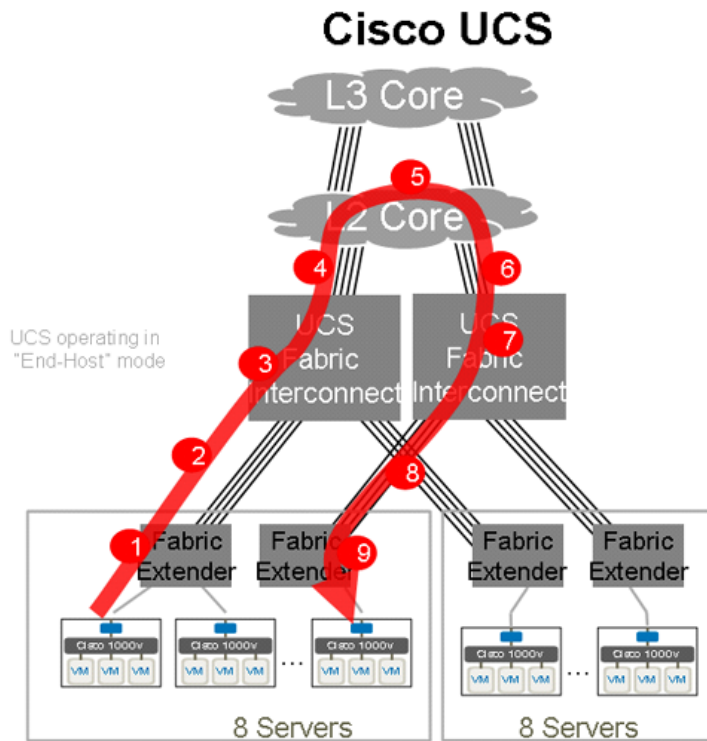
The OTV is designed for Disaster Avoidance. You can click on the following link <http://www.cisco.com/en/US/netsol/ns1153/index.html> for details.

The upcoming new vSphere 5.1 release (include more than 100 enhancements and new features with guaranteed application service levels) announcement from VMWORLD 2012 in San Francisco from Aug. 26 to 30 may further increase the latency from 10ms to 15ms or higher. This will further make the implementation of this Active/Active Datacenter Failover in reality and affordability.

II. Finding a Clue Why FCoE Has Bandwidth Bottleneck from Cisco UCS

I met and asked a Cisco Account Executive about the FCoE's (Fibre Channel over Ethernet) Bandwidth Bottleneck issue from Cisco UCS 1) if there is a heavy load (oversubscribed) in its system starting from its blade servers to top of the rack (ToR) or Fibre Interconnect (FI) switch; 2) the traffic goes from North to South, as illustrated in the screenshot below vs. East to West from other storage

vendors; 3) the FCoE consists of seven (7) virtual IP lanes (Layer 3) and only one (1) virtual FC lane; 4) the IP always has higher priority than one virtual FC lane in its UCS system. He immediately told me that the FCoE can use all lanes for the FC traffic dynamically as needed.



I asked an EMC senior engineer (SE) for a similar question right after an afternoon vBlock session and was told that he never heard of the issue. He asked me to seek the answer from a Cisco expert at the Cisco booth.

A Cisco expert told me that it is eight classes, instead of eight virtual lanes and the FC always has higher priority than IP. The eight classes are defined as below:

0 – System traffic (Always has highest priority)

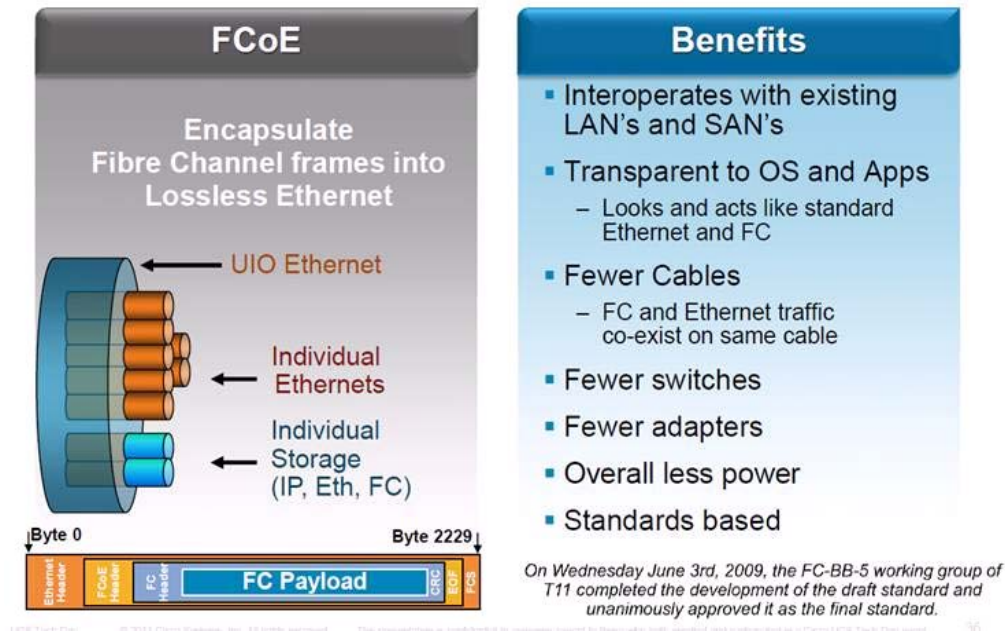
- 1 – FC traffic (Always has highest priority over IP)
- 2 – IP traffic (Always has lower priority over FC)
- 3 to 7 – Other traffic can be dynamically assigned to meet any traffic demand.

Although I still do not understand it, I opened my notes taken from Cisco one-day UCS Tech-Day Training and found the following diagram regarding FCoE.

As you can see, seven (7) virtual lanes are designated to Ethernets (Layer 2), while the remaining virtual lane (8th) can be assigned to IP, Ethernet or FC.

Consolidation with Unified Fabric

Fibre Channel with simpler infrastructure and lower cost



I also read the NetApp White Paper regarding FCoE (Click on the link below http://www.bladenetwork.net/userfiles/file/PDFs/WP_NetApp_Enhanced_Ethernet.pdf for details)

Here are the experts - "as well as increase traction for the upcoming Fibre Channel over Ethernet (FCoE) standard. As data volumes increase, **bandwidth is likely to become something of a bottleneck**, but storage success is based on more than raw throughput. Robustness, cost reduction, and ease of use are key goals for all organizations, and the convergence between SAN and LAN, made possible by storage over Ethernet, and lossless Ethernet will be a major step toward accomplishing these goals.

PRIORITY FLOW CONTROL (IEEE 802.1QBB)

Converged Enhanced Ethernet (CEE) capable products will enable lossless Ethernet fabrics by using IEEE 802.1Qbb Priority Flow Control (PFC) to pause traffic based on the priority levels. 802.1Qbb allows eight virtual lanes to be created in an Ethernet link, with each virtual lane assigned a priority level. During periods of heavy congestion, lower priority traffic can be paused, while allowing high-priority and latency-sensitive tasks such as data storage to continue.

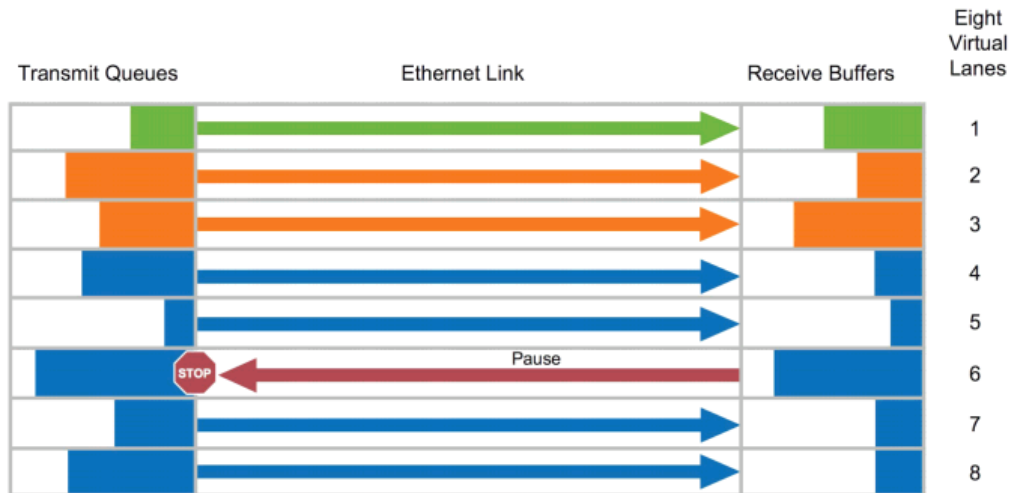


Figure 1) Priority Flow Control allows up to eight prioritized virtual lanes per Ethernet link.

IEEE 802.1Q (Virtual LAN) can be used to partition the physical Ethernet fabric to create high levels of security by isolating traffic types and to enhance quality of service (QoS) by configuring guaranteed bandwidth and latencies per VLAN. Using VLANs and 802.1Qbb flow control, several high-performance lanes of lossless Ethernet can be established on a single 10 Gigabit Ethernet fabric.

Not all data traffic has the same priority. For instance, storage traffic is generally higher priority than other network traffic, such as e-mail or instant messaging. For this reason, among others, it is common to have a dedicated storage network where security, QoS, and performance can be managed independently from the LAN. PFC, coupled with VLANs, allows LANs, SANs, and other application-specific networks to coexist on the same wire, while remaining isolated from each other logically to ensure I/O security. The result is improved iSCSI performance and savings associated with sharing a common wire and switch fabric. ”

In order to build a FCoE network, the followings are the requirements:

- Need a lossless Ethernet network that will support FCoE
- Need a Fibre Channel Forwarder (FCF) switch, e.g., Cisco Nexus 5000
- Need converged network adapters (CNAs) installed on each host server, which can connect to a FC or FCoE target, which can connect to the block storage (SAN) for the server.
- The current CNAs for FCoE support up to 4 Gbps of Fibre Channel traffic. The HBA driver and hardware on the shipping CNA is identical to that on the 4-Gbps Fibre Channel HBAs. The next generation of CNAs will support the 8-Gbps Fibre Channel standard.

Click the following link for details:

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/white_paper_c11-501770.html

Note:

- Cisco Nexus 5000 switch (Unified fabric) can be either Ethernet or FC switch and is almost always used as a top of rack (ToR) IP switch, but can be licensed to be used for the FCoE.

- Brocade 8000 [Prior to Transparent Interconnection of Lots of Links (TRILL)] and VDX switches, such as 6710 and 6720, can achieve the same. VDX uses TRILL and can be scaled into large Ethernet fabrics.

In my opinion, Cisco UCS might have a network bandwidth bottleneck issue when heavy loads (oversubscribed) are applied to its UCS system. The UCS system should be running very well in non-oversubscribed environments.

A further research is needed to confirm the hypothesis of the network bandwidth bottleneck issue mentioned above and must be verified by Cisco.

However, a help is under way, if the issue does exist due to oversubscribed, because Cisco just released 40 Gigabit Modules (40 Gigabit Ethernet interfaces) that meet connectivity requirements for high-performance computing, cloud services, and aggregation layers. Click the following link <http://www.cisco.com/en/US/products/ps11708/index.html> for details.

III. Building an Efficient and Flexible Virtual Infrastructure

a) EMC Avamar

It was designed from ground zero for protecting Virtual Machines (VMs) and is tightly integrated with the forthcoming vSphere 5.1, where VMware Data RecoveryPoint (VDR), a new feature, is added.

The VDR, a free utility, is based on Avamar technology and provides 90% to 99% inline deduplication. The recovery process is 30% faster due to utilizing changed block tracking CBT - with five clicks for quick backup and a few clicks for quick restore.

Avamar can:

- Provide faster image-level backup via VMware's CBT
- Use variable length deduplication technique to reduce the backup size
- Use bare-metal (VMDK) image-level for a quick restore.
- Protect individual physical servers.
- Support agentless deployment via VMware's best effort method, a one-way communication from a backup server to target server. However an agent must be installed on each VM for applications such as Exchange and SQL server in order to perform a reliable restoration due to application consistency checking required.

Note:

- Avamar might be able to replace a complicated CommVault backup solution in the future due to its tight integration with EMC's Data Domain, a

secondary disk-to-disk backup storage with high-speed and inline deduplication capability.

- Avamar was designed for medium and big enterprises, while Veeam was designed for small and medium size environments.

Notes: Veeam V6 is an excellent Backup & Recovery product. But, it has some challenges with its replication technology and the VMware VDR, a free and built-in powerful backup and restore application in vSphere 5.1.

Click <http://www.lacaaea.com/vendors/Veeam-evaluation07272012.pdf> for details.

b) VMware v-Director

The v-Director is used for multi-tenant purpose and deliver cloud infrastructure on-demand. Therefore, customers can consume virtual resources with maximum agility with VMware vCloud Director. For example, an organization can use v-Director to isolate each department virtualization environment

IV. VCE Vblock: The Simplest and Fastest Path to Your Cloud

This session is a general overview of the VCE (virtual computing environment) Vblock (www.vce.com). The VCE Company is formed by Cisco and EMC with some investments from VMware and Intel in a goal of accelerating the adoption of converged infrastructure and quickly building a private cloud. The presentation never mentions how many customers have adopted the vBlock across the globe since the product was released a few years ago, although the installation and provisioning of the vBlock should bring the initial acquisition and deployment of the product from months to days, including one technical support phone number to address multi-vendors' issues all together.

The vBlock basically consists of Cisco blade servers and switches with EMC VMAX as a disk storage. The VMAX has three types of models:

VMAX 40K (It was just released two months ago)
VMAX 20K
VMAX 10K (Formally called VMAXe)

Again, this session lacks of technical details. It does not have a single diagram to show the audience how the blade servers connected to SAN switches via FI switches etc.

According to some researches found, the vBlock may have some network bandwidth bottleneck issue when it is oversubscribed due to Cisco's FCoE

implementation. Please refer to the session – “Finding a Clue Why FCoE Has Bandwidth Bottleneck from Cisco” described in section II above for details.

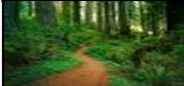
IV. **Network for the Virtual Enterprise Americas Forum Series Sponsor: Brocade** (Source: Brocade)

Although I did not attend this session, I chatted with Brocade staff at the Brocade booth and learned that Brocade switches consume much less power than Cisco's, as illustrated in the table below and its SAN FC switches own 71% market share today, while Cisco has 25% [e.g., HP BladeSystem (Matrix) is shipped with 90% Brocade OEM FC switches.]

Brocade SAN switch has the following additional features Cisco's does not have:

- 16Gbps capability
- InterSwitch Links (ISLs) with compression and encryption
- D-port testing
- Bottle Neck detection
- Local Switching (note: recently, Cisco supports local switching)
 - Cisco sent all traffic over the backplane prior to supporting local switching
 - The benefit of local switching is to prevent the backplane from over subscribed
 - when two devices are talking on the same switch (ASIC), the traffic does not go through the backplane
 - The average latency is 600ns via local switching; it is 1.2us when the traffic goes through the backplane
- 2 Tbps Inter Chassis Link (ICL) kits
- Frame based trucking
 - Brocade uses trucking to balance data at the Frame level
 - The trunk groups are master less, that is, if one of the links in the trunk group goes down, it does not cause a fabric rebuild in order to re-elect a trunk master.
- 512G per slot
- SAN Health tool

Energy Efficiency: Going Green

	Brocade DCX 8510-8 (384p×16 Gbps)	Brocade DCX 8510-4 (192p×16 Gbps)	Brocade DCX (384p×8 Gbps)	Cisco* MDS 9513 (528p×8 Gbps)	Cisco* MDS 9506 (192p×8 Gbps)
Watts	2117	1154	1265	4455	1798
Cooling (BTU/hr)	7227	3940	4317	15,204	6136
Watts/Port	5.5	6	4.9	8.4	9.4
Watts/Gbps	0.3	.3	.3	1.1	1.2
CO2 Emissions/yr metric tons	7.8	4.3	4.7	16.39	6.61

Try the Power Consumption Calculator at www.brocade.com/power

© 2012 Brocade Communications Systems, Inc. Company Proprietary Information

March 2012

Brocade provides Power Calculator, shown in the link below:

<http://www.brocade.com/data-center-best-practices/competitive-information/power.page>

Brocade advantages over Cisco – Highlights Include:

- 2X performance due to 16 Gbps vs. 8 Gbps
- ½ foot print
- ½ power consumption
- Less CO2 Emissions
- Lower price (up to ½ price of Cisco's)

I talked to a Cisco expert at Cisco booth regarding power on Cisco switches. He acknowledged that it is true that Brocade switches use less power because Cisco switch has more ASIC chips, which requires more powers, but the ASIC chips provide more functions.

Conclusion:

EMC Forum is an excellent forum and provides an opportunity for many IT professionals to learn how EMC cloud computing solutions transform IT, business, and yourself; in addition, networking with other IT professionals and exchanging some best practices etc.

Reference:

1. Fibre Channel Over Ethernet (FCoE) - Questions and Answers from live Webcast - <https://supportforums.cisco.com/docs/DOC-15882>
2. Data center bridging - http://en.wikipedia.org/wiki/Data_center_bridging
3. Converged Enhanced Ethernet (CEE) - http://www.bladenetwork.net/userfiles/file/PDFs/WP_NetApp_Enhanced_Ethernet.pdf

Recommended Reading:

1. How New York City is going to Consolidate 50 Data Centers from 40 City Agencies into One Location:
<http://www.informationweek.com/news/government/state-local/229219575>
2. NASA uses Amazon's cloud computing in Mars landing mission
Source: Los Angeles Times (Use a Search Engine to find the article)
3. The New York Public Library is powered by Google Cloud
<http://www.nypl.org/collections/articles-databases/google-book-search>
4. Debunking the Myth of the [Single-Vendor Network](#) (Source: Gartner)
<http://www.dell.com/downloads/global/products/pwcnt/en/Gartner-Debunking-the-Myth-of-the-Single-Vendor-Network-20101117-published.pdf>

Acknowledgement

Thanks for WWT for the courtesy of its presentation slides, where some of them are used in my notes.