# HP Advanced Data Center Network Architectures Reference Guide

Applying HP Converged Infrastructure to Data Center Networks

Technical white paper

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview of the Guide</td>
<td>3</td>
</tr>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Key Business and Technology Drivers for a New Data Center Network Architecture</td>
<td>4</td>
</tr>
<tr>
<td>Large-scale Data Center Consolidation</td>
<td>4</td>
</tr>
<tr>
<td>BladeSystems and Server Virtualization Technologies</td>
<td>5</td>
</tr>
<tr>
<td>New Application Deployment and Delivery Models</td>
<td>5</td>
</tr>
<tr>
<td>Data Center Deployment Models</td>
<td>5</td>
</tr>
<tr>
<td>Key Data Center Networking Requirements</td>
<td>6</td>
</tr>
<tr>
<td>Virtualization Scale-out: Layer 2 Performance, Scaling, High Availability, and Multi-site Extension</td>
<td>6</td>
</tr>
<tr>
<td>Securing the Virtual Server Edge</td>
<td>29</td>
</tr>
<tr>
<td>Managing and Provisioning the Virtual Server Edge</td>
<td>37</td>
</tr>
<tr>
<td>Converged Network Infrastructure: Unifying Data and Storage Networks</td>
<td>45</td>
</tr>
<tr>
<td>HP Data Center Reference Architecture (DCRA)</td>
<td>59</td>
</tr>
<tr>
<td>Demystifying Data Center Reference Architecture</td>
<td>60</td>
</tr>
<tr>
<td>Evolving Network Designs</td>
<td>61</td>
</tr>
<tr>
<td>Blade Server 1-Tier Design</td>
<td>64</td>
</tr>
<tr>
<td>Simplified 2-Tier Design (ToR)</td>
<td>67</td>
</tr>
<tr>
<td>Optimized 3-Tier Design</td>
<td>70</td>
</tr>
<tr>
<td>Summary</td>
<td>73</td>
</tr>
<tr>
<td>HP Networking Services</td>
<td>74</td>
</tr>
<tr>
<td>Glossary of Acronyms</td>
<td>75</td>
</tr>
<tr>
<td>For More Information</td>
<td>78</td>
</tr>
</tbody>
</table>
Table of Figures

Figure 1: Virtual machine footprint ................................................................. 8
Figure 2: Traffic from a virtual NIC is treated in one of two ways ................... 14
Figure 3: The SR-IOV process ........................................................................ 15
Figure 4: MultiChannel connectivity with virtual machines ............................. 17
Figure 5: Advantage of using vMotion on a Layer 2 network ............................. 18
Figure 6: Traditional 3-tier vs. collapsed virtualized design ............................... 20
Figure 7: Illustrates how data centers can be linked to support a large L2 network 21
Figure 8: STP ................................................................................................. 22
Figure 9: Virtual private LAN services ............................................................ 24
Figure 10: Network topology with and without IRF ........................................... 25
Figure 11: An example of the Spanning Tree Protocol ....................................... 26
Figure 12: Integrated policy management ......................................................... 30
Figure 13: Security perimeters around individual assets in the network in the data center 32
Figure 14: Virtualized servers ......................................................................... 33
Figure 15: Increased load on a physical server or a physical server failing ....... 34
Figure 16: IPS and SSL in the converged network infrastructure ........................ 35
Figure 17: Application VMs ........................................................................... 36
Figure 18: Configuration of over 2,700 devices from 30 manufacturers .......... 41
Figure 19: Applications consuming network bandwidth ................................. 42
Figure 20: Applications consuming network bandwidth 2 ................................ 43
Figure 21: Network out-of-band management in a data center .......................... 44
Figure 22: Converged infrastructure ............................................................... 45
Figure 23: Typical NAS configuration on a LAN ............................................. 47
Figure 24: Generic network transport architecture .......................................... 47
Figure 25: Typical SAN networks ................................................................... 48
Figure 26: iSCSI is SCSI over TCP/IP ............................................................. 51
Figure 27: FCoE packet .................................................................................. 52
Figure 28: Host bus adapters compared to converged network adapters (CNAs) 53
Figure 29: The evolution and convergence of network protocols ........................ 55
Figure 30: Flex-10 device .............................................................................. 57
Figure 31: Flex-10 with Fibre Channel or accelerated iSCSI ............................... 58
Figure 32: DCRA example methodology ......................................................... 60
Figure 33: Top of Rack placement ................................................................... 61
Figure 34: End of Row .................................................................................... 63
Figure 35: Blade Server 1-tier Architecture Rack View ..................................... 65
Figure 36: Blade Server 1-tier Logical ............................................................. 66
Figure 37: Simplified 2-tier Architecture Rack view .......................................... 68
Figure 38: Simplified 2-tier Logical ................................................................. 69
Figure 39: Optimized 3-tier Architecture Rack View ........................................ 71
Figure 40: Optimized 3-tier Logical ............................................................... 72

Table of Tables

Table 1: Data center deployment models and corresponding key networking imperatives 6
Table 2: Virtualization drives bandwidth & utilization ..................................... 8
Overview of the Guide

This guide is intended for technology decision-makers, solution architects, and other experts tasked with improving data center networking. It can serve as a baseline for network planning and design projects.

It is said, “You cannot chart your course without first knowing whence you came.” This also applies to data center architecture. However, many technical guides take the opposite approach. They attempt to sway the reader towards specific technical directions based on the merits of a current technology or standard. That approach often loses the reader, because it does not provide a context for why the new technical approach was developed in the first place.

This document will frequently reference technology trends in the data center which have and are being driven through virtualization and standards. It will also introduce issues that confront data center architects in this fast-paced, results-driven, and security-minded industry.

As mentioned before, technical documents often promote a vendor’s products or vision. This document takes a slightly different approach. Rather than put HP’s vision for converged network infrastructure first, this guide instead presents the building blocks for that vision. It does this by first identifying the most important IT trend today – virtualization of resources at all levels. It then moves forward by introducing HP-supported technologies that enhance virtualized computer networks. Finally, it will provide Data Center Reference Architecture examples for different types of virtualized server deployments using a layered approach.

The Data Center Reference Architecture guide is less of a discussion on specific HP equipment, and more of an overall focus on two things—virtualization and HP Networking technologies that support virtualization. It provides another level of detail to complement the HP Converged Infrastructure Reference Architecture Solution Block Design Guide http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA2-6453ENW.pdf

HP believes that the overriding key to success in networks supporting virtualization is simplification. This document provides guidance on network simplification for virtualized deployments, which do not sacrifice performance or deployment flexibility.

The major concept areas that will be covered are as follows:

- Virtual server networking
- Securing the virtual edge
- Managing the virtual edge
- Converged network infrastructure

This approach allows data center architects and IT teams to develop new and more flexible data center models and methodologies. By doing so, IT can meet new demands head-on, rather than forcing businesses to adapt to technology limitations.
Introduction

The primary driver for evolving today’s enterprise data center is efficiently deploying resources to support business applications. Apparent “trends” in consolidation and distribution of infrastructure are side effects of addressing this core goal. In fact, enterprise data center network architects and managers are now expected to build networks that can concurrently consolidate and geographically distribute resources. These include physical and application servers, data storage, management platforms, and knowledge workers at the same time. This evolution did not happen overnight. It has been fueled by the accelerating needs of businesses to be more agile, to do more with less, and to increase their IT efficiency.

The current trend in network architecture is virtualization. Virtualization encompasses the ability to operate multiple servers concurrently on top of a single server hardware platform, sharing CPU, disk, interface, and network services. In this case, each virtual server operates as an independent entity on a single physical server. Virtualization does not dismiss the need for network convergence. Early deployments of virtualized servers operated statically on the platform they were deployed on. Today, virtualized servers are flexible in their deployment, with the ability to move to other physical server platforms. Moving with the virtualized server is their configured storage, data, and multi-modal communication functions. This type of infrastructure can now significantly reduce equipment, network facility, and operational expenses.

In an effort to address this opportunity, traditional data center models have stretched to support autonomous data and storage networks with separate interface cards, switches, and cabling plants. This methodology has proven to be ineffective and costly when implementing and operating both centralized and geographically dispersed data centers.

To reduce that complexity, data center architects adopted designs utilizing network convergence, where data and storage I/O are merged onto a single network. Network convergence has provided substantial gains, but falls short in providing a complete answer for overall IT infrastructure efficiency.

This converged approach can eliminate physical clutter and complexity, while making more efficient use of networking resources. However, the simplification, in many cases, is only true at the surface level. Data center networks today are more complex below the surface than ever before. Virtualization has presented a substantial number of challenges which are driving a new evolution in data center network architecture that far exceeds the initial need for a converged network infrastructure.

HP’s approach, detailed below, is to reduce the complexity of data center networks, especially now that the primary requirement is networking virtual instances.

Key Business and Technology Drivers for a New Data Center Network Architecture

Large-scale Data Center Consolidation

For many enterprise customers, the data center IS the business. Mission-critical applications and services provide the foundation for day-to-day operations and delivery of end-customer services. The data center must deliver unquestioned availability and meet stringent service level agreements. Exploiting server virtualization and low-cost computing power, customers are deploying more and more sophisticated applications on a larger scale. To reduce their sheer complexity and improve operations in these deployments, customers are seeking to consolidate fragmented, dispersed facilities into fewer, centralized locations.
Today's data center networks must be designed to deliver much higher levels of performance, scalability, and availability than before to meet service-level agreements and maintain continuity of operations. Beyond sheer performance, these data center networks must quickly recover from hardware- or software-related faults and protect against server, storage, network, and application vulnerabilities to maintain performance and minimize service disruptions.

**BladeSystems and Server Virtualization Technologies**

The adoption of increasingly powerful multi-core-processor servers, higher-bandwidth interfaces, and BladeSystems is dramatically increasing the scale of data center deployments. Now, tens of thousands of virtual machines are commonly deployed in a single data center to consolidate infrastructure and streamline operations. These large-scale solutions are dramatically increasing network performance requirements at the server edge and across the extended network. Likewise, virtualization and vMotion/Live Migration tools for moving virtual servers are introducing high-volume machine-to-machine traffic flows. This impacts existing administrative practices, creating a new “virtual edge” which blurs the traditional boundaries between network and server administration.

**New Application Deployment and Delivery Models**

Traditional client-server software and infrastructure deployment models are being displaced by new application architectures and service delivery models which are reshaping the data center. Web 2.0 mashups, SOA solutions, and other federated applications are being widely deployed to deliver integrated, content-correlated, context-specific information and services to end-users within the enterprise and beyond. These deployments drive new, bandwidth-intensive, traffic flows within the data center and demand low-latency, high-performance server-to-server and intra-server, virtual machine-to-virtual machine connections. At the same time, cloud computing and XaaS initiatives are introducing more stringent service level and security demands and driving requirements for a more agile and dynamic infrastructure.

**Data Center Deployment Models**

The adoption of more virtualized, dynamic application environments is impacting traditional enterprise and hosted/multi-tenant data center designs. These methods enable new cloud-based delivery models that drive a whole new set of technology requirements across servers, storage, and networking domains. These increasingly popular-use models let enterprises provision applications more flexibly within traditional internal infrastructures, and enable hosted application and service providers to build entire businesses based on delivering services via a public cloud model. Given the range of use cases and options, customers often deploy a combination of architectures to address varied requirements and to optimize operations.

Table 1 summarizes some of the most important networking focus areas that emerge as customers pursue these diverse deployment models. While all these imperatives play some role across all the deployment models regardless of market or industry, certain initiatives figure more prominently in specific use cases.
Table 1: Data center deployment models and corresponding key networking imperatives

<table>
<thead>
<tr>
<th>Deployment Model</th>
<th>Characteristics</th>
<th>Key Networking Focus Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Enterprise Data Center</td>
<td>DC services are a critical complement to the company’s core business</td>
<td>Converged networking&lt;br&gt;Virtualization scale-out&lt;br&gt;Managing/provisioning the virtual server edge</td>
</tr>
<tr>
<td></td>
<td>Complex application environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Security, cost and flexibility are key</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evolving towards private cloud over time</td>
<td></td>
</tr>
<tr>
<td>Traditional Multi-tenant Data Center</td>
<td>DC services are the company’s core business</td>
<td>Virtualization scale-out&lt;br&gt;Securing the virtual server edge&lt;br&gt;Managing/provisioning the virtual server edge</td>
</tr>
<tr>
<td></td>
<td>Complex application environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Security, SLAs and flexibility are key</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evolving towards public cloud over time</td>
<td></td>
</tr>
<tr>
<td>Multi-tenant XaaS/ Cloud Computing Data Center</td>
<td>DC services may be the company’s core business</td>
<td>Virtualization scale-out&lt;br&gt;Securing the virtual server edge&lt;br&gt;Managing/provisioning the virtual server edge</td>
</tr>
<tr>
<td></td>
<td>Heavy use of blade servers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cost, latency and scalability are key</td>
<td></td>
</tr>
<tr>
<td>High-performance Computing Data Center</td>
<td>DC services may be the company’s core business</td>
<td>Low latency</td>
</tr>
<tr>
<td></td>
<td>Heavy use of blade servers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cost, latency, performance and scalability are key</td>
<td></td>
</tr>
</tbody>
</table>

Key Data Center Networking Requirements

Virtualization Scale-out: Layer 2 Performance, Scaling, High Availability, and Multi-site Extension

Overview

Virtualization is now broadly accepted in data centers. With it, the operating systems, applications, and servers can work in a non-dedicated or loosely coupled manner, interacting as needed to meet enterprise needs. Virtualization provides several key benefits:

- **Higher Efficiency/Reduction in Capital and Operational Expenditure** – As multi-core/multi-processor systems provide a substantial amount of computing performance, virtualization allows for the more efficient use of a given server and its resources. Capital expenditure on hardware and software, as well as operation and support costs, are reduced.

- **Agility and Flexibility** – Data center managers must keep operations running efficiently by rapidly deploying resources for business growth and for stress periods. Application developers look for flexibility in the use of CPU and memory utilization. Virtualization provides the ability to do on-the-fly migration of virtual servers or applications across physical servers in the data center.

- **Resiliency** – Virtualization technology provides the ability to restack or shuffle applications in support of business continuity or even for routine maintenance.

Virtualization creates the need for new data center network architecture methodologies. Traditional data centers with standalone servers were defined by rack space, switch port capacity, IP addressing, and sub-netting requirements. Bandwidth and packet forwarding performance have always been key elements, but now there are fewer physical servers and more processor cores within a given data center footprint. See Figure 1 (“Virtual machine footprint”).
With a virtualized environment, a complete enterprise information system infrastructure can be deployed in less space than in the past. With virtualization, the number of servers requiring bandwidth in a given rack can easily quadruple the bandwidth requirements over more traditional data center topologies. See Table 1 (“Virtualization drives bandwidth & utilization”).

Now, more than ever, data center architecture designs require thorough investigation and planning focused on bandwidth requirements, ethernet switch performance, shared data storage resources, broadcast traffic, and geographic redundancy.

It is also important to note that as data centers become more virtualized, and networks converge, the need for designs providing performance and capacity are becoming imperative. But the reality is that data centers are not ‘greenfield’ implementations. Data centers were previously designed with traditional three-tier network architectures. Three-tier architectures operate on a foundation of IP peering and single purpose rack mount application servers. This status quo is being turned upside down with the rapid deployment of virtual servers. The fact is, data no longer primarily moves between servers inside and outside of the data center, but now moves horizontally within and across data center virtual server and network boundaries. Traditional three-tier network architectures focused around Layer 3 IP are not well suited to support these virtualized deployments.

HP believes that in order to support low latency, high performance applications on virtual servers with converged I/O technologies, the solution should operate on collapsed Layer 2 networks, with Layer 3 routing in the core or aggregation layers. In addition, maintaining a larger Layer 2 domain provides the maximum flexibility to allow VM mobility with technologies like vMotion or Live Migration. HP is also driving technologies and standards focused around Converged Enhanced Ethernet, and TCP in support of iSCSI and FCIP (Fibre Channel over IP). It is only logical that with rack and processing utilization gains of 300 -1000% with blade servers and VMs in a single rack, an enterprise can house their entire IT demands within that single rack. Therefore, server to server and virtual machine to virtual machine communications are going to be best suited with flatter L2 networks. HP is continuing to meet these demands with its L2 network and IRF switching foundation strategy.
Figure 1: Virtual machine footprint

![Diagram showing traditional and contemporary design with server footprints and network configurations.

Table 2: Virtualization drives bandwidth & utilization

<table>
<thead>
<tr>
<th>Server Type</th>
<th># of Physical Servers</th>
<th># of Virtual Servers</th>
<th>Server Bandwidth</th>
<th>Rack Bandwidth</th>
<th>Bandwidth &amp; Footprint Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rack Servers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Single Quad Core Xeon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• (10) VMs per Server</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Virtualization</td>
<td>36</td>
<td>0</td>
<td>300Mbps</td>
<td>10.8Gbps</td>
<td></td>
</tr>
<tr>
<td>Virtualized</td>
<td>36</td>
<td>360</td>
<td>3Gbps</td>
<td>108Gbps</td>
<td>+900%</td>
</tr>
<tr>
<td>Blade Servers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Single Quad Core Xeon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 16 Blades per Chassis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• (10) VMs per Blade Server</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 3 Chassis per Rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Virtualization</td>
<td>48</td>
<td>0</td>
<td>300Mbps</td>
<td>14.4Gbps</td>
<td></td>
</tr>
<tr>
<td>Virtualized</td>
<td>48</td>
<td>480</td>
<td>3Gbps</td>
<td>144.0Gbps</td>
<td>+900%</td>
</tr>
</tbody>
</table>
Server Deployment Considerations

Virtualization of data, web-based and multimedia applications have driven the requirement for new servers, new platforms, and associated network storage. This has resulted in unplanned growth in server complexity. IT departments are being charged with the improvement of computing efficiencies in their data centers. The response has been server consolidation utilizing virtualization technologies implemented on blade servers.

Rack Servers vs. Blade Enclosures

In a typical rack server environment, each rack is connected to two networks: The Ethernet general purpose network (LAN) and the storage network (SAN). The Ethernet network has the following features:

- A single rack of servers may connect to Top of Rack switches (typically deployed in pairs for redundancy) which uplink to a redundant aggregation network layer.
- Several racks of servers may connect to end-of-row switches which consolidate the network connections at the edge before going into the aggregation network layer.

These types of networks have different interconnect implementations. Each server on the rack needs to have one set of interconnects to the LAN, and one set to the SAN. In addition, each server has its own power requirements, cooling requirements, and management requirements. Each network has its own management process, as well as its own security and performance requirements. Some rack servers have now instituted splitting the physical frame into two separate server units. This type of design provides expanded capacity and processing power per rack unit, but lacks the network consolidation that blade servers possess.

Advantages

- **Simplicity** – One physical system sharing common CPU(s), memory, and storage controller.
- **Commodity pricing** – Through standards and competition.
- **Industry acceptance** – IT technicians have been working with these systems for years, and know how to size them, install them, configure them, and support them.
- **More expansion slots available for network and storage adapters** – Larger rack-mount servers (of more than 5U) have up to seven I/O expansion slots. Rack servers provide a large number of Network Interface Cards (NICs) or storage controllers in your virtual host for load balancing and fault tolerance.
- **Traditional servers have a greater internal capacity for local disk storage** – If running several VMs on local disk storage, rack servers are a better choice, as they have more drive bays for internal disk storage.
- **Processing power** – Rack servers can support 8 or more CPU sockets. This works well when using a smaller number of powerful servers for virtual hosts.
- **Infrastructure** – Rack Servers come complete, no need for integrated Ethernet switches or blade chassis. Once at full capacity, purchase another rack server, not a whole new blade chassis and components.
- **Serial, parallel, and USB I/O ports** – For connecting external storage devices and optical drives. They can also accept hardware dongles used for licensing software.
- **Backup devices** – Rack servers can support tape backup devices inside the host chassis.

Disadvantages

- **Network consolidation** – Adds required ethernet ports to the network.
- **Green data center** – These types of systems do not optimize the computing power per rack unit, they utilize more electrical power per rack unit, and add to the overall requirement for cooling systems.
Blade servers can bring about a reduction in the external switch count and can significantly reduce the number of cables in a rack. The blade enclosure can include server-to-network interconnect modules, which can replace the access layer switch at either the Top of Rack or End of Row placements.

With an Edge-Core SAN architecture, SAN designers selecting blade enclosures can use these approaches:

- Use pass-through modules in the enclosure while maintaining an external edge switch.
- Replace an existing external Fibre Channel edge switch with an embedded blade network module. This reduces the number of physical connections at the server edge to both the network and storage media. Because the blades share the chassis power supply, power and cooling requirements are reduced.

Advantages

- **Data center space** – Blade servers can increase rack density. Compared with traditional servers, this means up to 50% more servers in a standard 42U rack.
- **Disk expansion blades** – Rack servers do not possess this feature, so they must connect to external SCSI array chassis or networked storage.
- **Consume less power** – They are more energy efficient and require less cooling. The amount of power that blades consume is dependent on how full the blade chassis is. A fully-loaded chassis will consume far less power than an equivalent amount of traditional servers. If the chassis is not full, it will still consume less power than an equal amount of rack servers.
- **Integrated Ethernet switches** – Ethernet cables plug into a chassis with a single connector, which makes cabling neater and eliminates the clutter of cables common with rack servers.
- **Booting from SANs** – Blade servers require minimal resident disk storage. Virtual hosts can boot from the storage area network (SAN), so no internal disk is required. The host performs a Preboot Execution Environment (PXE) start up from the network, and then connects to the SAN disk to access the files to continue the boot process.

Disadvantages

- **Local expansion slots** – Relatively limited number of expansion slots for storage and/or internal blade Network Interface Cards.
- **Hard disk capacity** – Blade servers typically have a limited amount of localized hard disks (zero to four drives), and are dependent on SAN and/or NAS (Network Attached Storage).
- **Expansion cost** – Blade server chassis have substantial capacity, but once full, the next incremental CPU blade requires another chassis and the additional integrated chassis components.
- **Blade flexibility** – As new blades are developed by chassis vendors, not all new blades are downward compatible with the existing chassis, so it is important to verify this with the vendor before making the investment in a specific chassis.

**Server Virtualization Design Considerations**

Virtualization can enable an organization to create more agile IT services that can respond faster to business needs. Virtualization is not a single technology; rather, it is a way of designing technology solutions. Many different aspects of IT infrastructure, such as servers, storage, networks, and clients can be virtualized.
**Server Virtualization Technologies – vSphere (VMware), Hyper-V (Microsoft®), XEN (Citrix)**

Virtualization is the separation of a resource, application, or service from the underlying physical components of that service. For example, virtual memory allows applications to access more memory than is physically installed. Similarly, server virtualization can give applications the appearance that they have access to an entire server, including memory, CPU, and hard drives, while in reality, they may be sharing the physical hardware with other operating systems and applications. Each application and operating system combination is called a virtual machine.

Data center administrators have the ability to utilize physical servers optimally by placing virtual machines across their physical system assets and infrastructure.

A key benefit of virtualization is that it provides the ability to run multiple operating system instances on the same physical hardware simultaneously and to share hardware resources such as disk, network, and memory. Running multiple instances that are independent of each other on a single physical hardware structure is known as partitioning.

**VMware**

VMware offers multiple components. The most important are mentioned below:

- **Hypervisor** – VMware ESX Server. Runs on physical servers and abstracts processor, memory, storage, and networking resources.
- **VMware Virtual Machine File System** – A file system for use by virtual machines.
- **Virtual Center Management Server** – Allows for configuring, provisioning, and managing virtualized IT infrastructure.
- **VMware vMotion™** – Allows live migration of virtual machines running on one physical server to another with no down time and complete transactional integrity.
- **VMware High Availability** – In the event of server failure, High Availability automatically starts affected virtual machines on other servers with spare capacity.
- **VMware Consolidated Backup** – Allows for a centralized, agent-free backup of virtual machines.

One of the most important components in the context of this document is called “virtual switch” or “vSwitch”. VMware provides a “vSwitch” as a software component running under the control of the hypervisor. It provides the necessary network connectivity between virtual machines running on the server, as well as between the virtual machines and the outside world.

Each virtual switch provides modular support for functionalities such as L2 forwarding, VLAN tagging and stripping, and L2 security. The required modules for a switch are loaded at run time.

In many ways, a virtual switch behaves like a physical switch. For example, a virtual switch maintains an L2 forwarding table, which contains associations between the port number and the Media Access Control address (MAC address) of any device connected to that port. As each ethernet frame arrives at the virtual switch, it looks up the destination MAC and forwards the frame to the associated port. VLANs can also be set up on the virtual switch. Ports can be defined as either an access port (where the port is associated with a single VLAN) or as a trunk port, which allows multiple VLAN packets to traverse. A trunk port is typically used to connect two VLAN-aware devices, such as two switches, or a switch and a router. If the two switches were connected using access ports (same VLAN), then only traffic for that VLAN will pass between the two switches. If multiple VLANs are defined on both switches, a trunk port is needed to allow traffic for these VLANs to traverse both switches.

vSwitch ports also support promiscuous mode. Here, all traffic received on one port is copied onto another port regardless of VLAN membership or destination MAC. This is useful for network monitoring and sniffing applications, as well as intrusion detection systems (IDS) which require L2 insertion.
Unlike a physical switch, there is no need for frame flooding to map a MAC address with a switch port. Because each virtual switch port is connected to a virtual NIC in the virtual machine, this association begins when the virtual machine is started.

Unlike physical switches, VMware vSwitches cannot be interconnected on the same server. As a benefit, ethernet loops cannot exist in the virtual topology. Therefore, the Spanning Tree Protocol is not required and is not present in a vSwitch. Additionally, virtual switches cannot share the same physical Ethernet NIC, and cannot have an entry in one table of a virtual switch port and on another virtual switch. The L2 hash tables are associated with each NIC, and cannot be duplicated on other NICs.

An important concept in virtual switches is port groups, which are templates that hold specific sets of specifications for a port. A port group ensures that a particular virtual machine has the same type of connectivity on any machine on which it is run. This is particularly important for vMotion, where a live migration of a VM from one physical server to another can take place without affecting application services.

The typical specifications carried by a port group include virtual switch name, VLAN IDs, L2 security policies, and traffic shaping parameters. A virtual machine can be associated with a particular port group, which then connects the virtual NIC to an appropriate port on the switch.

The connection between a virtual network and a physical network is via the physical Ethernet adapters on the host. These are called uplinks in VMware terminology, and the virtual ports connected to them are called uplink ports. If a virtual switch has multiple VLANs configured, the link between the uplink port and the physical switch port must be in trunk mode.

In order to support VLANs, an element in either the virtual or the physical network has to be able to tag the ethernet frames with 802.1q tags.

There are three ways that this can be done:

- **Virtual Switch Tagging** – A port group is defined for each VLAN, and the virtual machine is associated with the relevant port group. The port group tags all outbound frames and strip tags off all inbound frames.

- **Virtual Machine Guest Tagging** – An 802.1q VLAN driver is installed in the virtual machine, which tags all outgoing frames. The tags pass through unchanged through the vSwitch.

- **External Switch Tagging** – An external physical switch is used for tagging.

There are some security policies which can be applied to the vSwitch to prevent virtual machines from seeing traffic destined for other nodes in the virtual network. By default, promiscuous mode is turned off. The administrator can also specify a MAC address lockdown, which prevents individual virtual machines from changing their MACs, thus preventing a node from seeing unicast traffic to other nodes. The administrator can also block forged transmissions to prevent virtual machines from spoofing other nodes.

**Server-to-Network Edge Using Virtual Machines**

Virtual machines and virtual switches running on a physical server introduce new a complexity at the server edge and dramatically impact associated networks.

**Challenges with virtual machines**

- Managing virtual machine sprawl and associated virtual networking.
- Performance loss and management complexity of integrating software-based virtual switches with existing network management: These are significant challenges not fully addressed by any vendor today. HP is working with other industry leaders to develop standards to simplify and solve these challenges.
**Virtual Ethernet Bridges**

With the growing use of virtual machines and their associated virtual switches, a new level of complexity has been introduced at the server edge. Network management has to consider a new virtualized network with virtual switches. Typically, two different groups are now involved at the edge—the traditional network administrators, who are responsible for the physical network switches, and the server administrators, who are responsible for managing and configuring the virtual switches, which must be configured and managed manually.

Virtual switches do not have the traffic monitoring capabilities of physical access switches, so troubleshooting VM-to-VM traffic may be an issue. Additionally, they may lack some of the advanced security features provided by physical switches, and even when these are provided, they may not be fully compatible with their physical counterparts. This makes it difficult to provide consistent end-to-end security policies.

As the number of VMs in a server increases, the traffic through the virtual switches also increases. This drives greater CPU resource utilization on the physical servers used to handle the traffic.

A vNetwork Distributed Switch (VDS) has been introduced with VMware vSphere 4.0 to respond to these complexities. A VDS treats the network as an aggregated resource, and provides a single abstracted view of all the individual vSwitches present in the virtual network. This abstracted view presents a single, large switch that spans multiple servers across the data center. Port groups that were earlier associated with a single vSwitch now become Distributed Virtual Port Groups that span multiple hosts and provide support for features such as vMotion.

The virtual switches in earlier versions of VMware had to be managed and configured separately. With a VDS, only a single switch needs to be configured and managed.

**Distributed Virtual Uplinks (dvUplinks)** are a new concept with the VDS that provides an abstraction for the virtual NICs on each host. The NIC teaming, load balancing, and failover policies on the VDS and the Distributed Virtual Port Groups are applied to the dvUplinks rather than to individual virtual NICs. Each virtual NIC on a virtual machine is then associated with a dvUplink, providing consistency in the way teaming and failover occur.

---

**VM-to-VM Switching**

With multiple Virtual Machines residing on a single physical server, there are requirements for efficient VM-to-VM traffic switching on the same server, and the ability to enforce physical switch policies on the virtual machines.

There are two models for enabling VM-to-VM switching:

- **Virtual Ethernet Bridge (VEB)** – A Virtual Ethernet Bridge is a virtual ethernet switch. It can reside on a single physical server and provide a virtual network between virtual machines on that server, or it can be used to connect the virtual machines to an external network. Most VEBs are software-based. However, with the adoption of the PCI Single Root IO Virtualization (SR-IOV) standard, hardware-based virtual switches are built into NICs. Hardware-based VEBs provide better performance than software-based VEBs.

- **Software-based VEBs** – The hypervisor creates Virtual NICs for each virtual machine, then creates one or more virtual switches that connect the virtual NICs to the physical NICs. Traffic received by a physical NIC is passed to the correct virtual NIC based on the configuration information held by the hypervisor.

Traffic from a virtual NIC is treated in one of two ways, as shown in Figure 2 (“Traffic from a virtual NIC is treated in one of two ways”).

- If the destination is external to the physical server or to a different vSwitch, the vSwitch forwards the traffic to the physical NIC.
• If the destination is on the same vSwitch on the same physical server, the vSwitch forwards the traffic back to the target virtual NIC.

Figure 2: Traffic from a virtual NIC is treated in one of two ways

Advantages of Software-based vSwitches
- Good performance for horizontal, inter-VM traffic: Internal VM-to-VM traffic involves only memory copy operations, and the performance is limited by available CPU cycles and memory bandwidth.
- Software-based vSwitches are standards-based, and can interoperate with a wide variety of external network infrastructure.

Disadvantages of Software-based vSwitches
- Because the host server’s CPU is used by the virtual switch, more CPU cycles are used for higher traffic. This limits the number of virtual machines that can be supported on a single server.
- Lack of standard network monitoring capabilities such as flow analysis, statistics, and remote diagnostics: VM-to-VM traffic on the same vSwitch is not visible outside the physical server.
- No way to enforce policies: External switches support port security, QoS, and access control lists. Even if vSwitches support some of these features, their management and configuration is not consistent with the way external switches are managed, thus making it difficult to create end-to-end policies in the data center.

Issues with management scalability – As the number of virtual machines increases, the number of vSwitches also increases. Because each vSwitch must be managed separately, this adds to management overhead. Distributed vSwitches somewhat alleviate this problem, but do nothing for management visibility beyond the virtualized server. Single pane of glass management systems will have to address this issue with added functionality in their software.
• **Hardware-Based VEBs** – The performance issues with software-based VEBs can be alleviated by moving the VEB functionality to the NIC. **Single Root I/O Virtualization (SR-IOV)** is the technology that makes this possible. SR-IOV is a standard developed by PCI-SIG that allows multiple PCI devices to be shared among virtual machines, and is implemented in hardware. SR-IOV creates many virtualized instances of a physical PCI device, and each virtualized instance is assigned to a VM. This allows the VM to access the PCI device directly for IO operations, without hypervisor mediation. This reduces the number of CPU cycles required to move traffic between VMs and from VMs to the external world.

In order for SR-IOV to work on a server, an SR-IOV NIC is required. This would have the virtual functions built into the hardware, making the virtualized instances possible. Additionally, the physical server’s BIOS must be capable of recognizing the SR-IOV NIC. The hypervisor must be capable of loading the drivers to work with the SR-IOV NIC, and the guest operating systems must be able to perform IO directly with the SR-IOV NIC. **Figure 3, (“The SR-IOV process”)** illustrates this.

![Figure 3: The SR-IOV process](image)

**Advantages of using SR-IOV NICs**

- Reduced CPU utilization compared with software based vSwitches.
- Increased network performance due to direct I/O between a guest OS and the NIC.

**Disadvantages of using SR-IOV NICs**

- As with software-based VEBs, SR-IOV NICs are not visible from the network. Because of limits on cost-effective NIC silicon, they often have fewer capabilities than software-based VEBs.
- Similar to software VEBs, they do not offer advanced policy enforcement.
- SR-IOV NICs usually have very small address tables, and they do not learn. This may increase the amount of flooding, especially if there is a large amount of VM-to-VM traffic.
- Guest operating systems need to be able to support SR-IOV. This is currently available only in open source operating systems.
Virtual Ethernet Port Aggregator (VEPA)
The IEEE 802.1 Working Group has agreed to base the IEEE 802.1Qbg EVB standard on VEPA technology because of its minimal impact and minimal changes to NICs, bridges, existing standards, and frame formats (which require no changes).

VEPA is designed to incorporate and modify existing IEEE standards so that most existing NIC and switch products could implement VEPA with only a software upgrade. VEPA does not require new tags and involves only slight modifications to VEB operation, primarily in frame relay support. VEPA continues to use MAC addresses and standard IEEE 802.1q VLAN tags as the basis for frame forwarding, but changes the forwarding rules slightly according to the base EVB requirements. In doing so, VEPA is able to achieve most of the goals envisioned for EVB without the excessive burden of a disruptive new architecture such as VN-tags.

Software-based VEPA solutions can be implemented as simple upgrades to existing software virtual switches in hypervisors. In addition to software-based VEPA solutions, SR-IOV NICs can easily be updated to support the VEPA mode of operation. Wherever VEBs can be implemented, VEPAs can be implemented as well.

VEPA enables a discovery protocol, allowing external switches to discover ports that are operating in VEPA mode and exchange information related to VEPA operation. This allows the full benefits of network visibility and management of the virtualized server environment.

There are many benefits to using VEPA:

- A completely open (industry-standard) architecture without proprietary attributes or formats.
- Tag-less architecture which achieves better bandwidth than software-based virtual switches, with less overhead and lower latency (especially for small packet sizes).
- Easy to implement, often as a software upgrade.
- Minimizes changes to NICs, software switches, and external switches, thereby promoting low cost solutions.

MultiChannel Technology

During the EVB standards development process, scenarios were identified in which VEPA could be enhanced with some form of standard tagging mechanism. To address these scenarios, an optional “multichannel” technology, complementary to VEPA, was proposed by HP and accepted by the IEEE 802.1 Working Group for inclusion into the IEEE 802.1Qbg EVB standard. MultiChannel allows the traffic on a physical network connection or port (like an NIC device) to be logically separated into multiple channels as if they are independent, parallel connections to the external network. Each of the logical channels can be assigned to any type of virtual switch (VEB, VEPA, and so on) or directly mapped to any virtual machine within the server. Each logical channel operates as an independent connection to the external network.

MultiChannel uses existing Service VLAN tags (“S-Tags”) that were standardized in IEEE 802.1ad, commonly referred to as the “Provider Bridge” or “Q-in-Q” standard. MultiChannel technology uses the extra S-Tag and incorporates VLAN IDs in these tags to represent the logical channels of the physical network connection.

MultiChannel supports:

- Multiple VEB and/or EVB (VEPA) virtual switches share the same physical network connection to external networks: Data center architects may need certain virtualized applications to use VEB switches for their performance and may need other virtualized applications to use EVB (VEPA) switches for their network manageability, all in the same physical server.
- Direct mapping of a virtual machine to a physical network connection or port while allowing that connection to be shared by different types of virtual switches: MultiChannel technology allows
external physical switches to identify which virtual switch, or direct mapped virtual machine, traffic is coming from.

- MultiChannel also allows direct mapping of a virtual machine that requires promiscuous mode operation (such as traffic monitors, firewalls, and virus detection software) to a logical channel on a network connection/port. Promiscuous mode lets an NIC forward all packets to the application, regardless of destination MAC addresses or tags, thus placing the port mirroring process burden on physical ethernet switches that are better suited for this role.

Figure 4: MultiChannel connectivity with virtual machines

Optional MultiChannel capability requires S-Tags and “Q-in-Q” operation to be supported in the NICs and external switches, and, in some cases, it may require hardware upgrades, unlike the basic VEPA technology, which can be implemented in almost all current virtual and external physical switches. MultiChannel does not have to be enabled to take advantage of simple VEPA operation. MultiChannel merely enables more complex virtual network configurations in servers using virtual machines. See Figure 4 (“MultiChannel connectivity with virtual machines”).

MultiChannel technology gives data center architects the ability to match the needs of their application requirements with the design of their specific network infrastructure:

- VEB for performance of VM-to-VM traffic.
- VEPA/EVB for management visibility of the VM-to-VM traffic.
- Sharing physical NICs with direct mapped virtual machines.
- Optimized support for promiscuous mode applications.

Layer 2 and Virtual Machine Mobility

Virtualization enables the ability to move VMs on the fly from one physical server to another, whether in the same rack or across data centers. Mainstream products which enable VM movement include vMotion from VMware or Live Migration from Microsoft. These products streamline application deployment, management of VMs and physical servers. However, because VMs cannot be moved outside their L2 network, there are implications that impact the choice of L2 or L3 solutions for virtualized server deployments. See Figure 5 (“Advantage of using vMotion on a Layer 2 network”).
Advantages in choosing L2 solutions for the movement of VMs are:

- VMs or servers can be moved without requiring the IP addresses to be changed.
- Easily accommodate legacy applications/servers that require the IP address to be embedded.
- Eases the management of VMs/servers across a virtual data center or hosting during data center consolidation.

Risks:

- Topology changes can affect throughput.
- STP by nature is limited and can operate at losses of more than 50%. This reduces the bandwidth and, consequently, increases completion times when moving VMs.

Key Drivers for Layer 2 Networks

Application & Server Virtualization – The operational role of virtualized servers is ever-increasing in the data center. Virtualization of servers and applications is highly dependent on L2 technology. The fact that it is not possible to move virtual machines (VMs) outside of the L2 network that contains it is a key factor in this dependency. For this reason alone, the development of flexible L2 architectures in the data center is a key factor that is part of the broad advantages of virtualization.

Network Virtualization – Very much like the concepts behind application and server virtualization, the goal of network virtualization is to add flexibility and greater utilization of network resources. Today’s data center networks are in need of virtual L2 backbones that not only span one data center location, but can also operate across multiple geographically separated data centers. This demands increased control of traffic flows, efficient use of bandwidth, and reduced network elements needed for virtual backbones. A major goal of network virtualization is to further enhance the gains made with application and server virtualization. Providing a robust L2 topology with technologies like IRF can
further eliminate the Spanning Tree Protocol (STP), while at the same time support both multipath forwarding and localized failure resolution.

**Geographic Redundancy** – There are a substantial number of server and communication applications that require L2 protocols to identify operational state through heartbeats. L3 wide area networks limit the ability of servers to communicate in the same L2 broadcast domains. Extending L2 VLANs across physical network boundaries drives greater flexibility in data center and application deployment designs. Virtual L2 backbones eliminate the need for L3 tunneling to extend both L2 heartbeat traffic and private IP address space between separate data center locations. These are key factors in designing for geographic redundancy.

**Performance** – As stated earlier, data centers have an ever-increasing need for more bandwidth. The deployment of L2 Virtualized Backbones allows ethernet switch clustering. These can reach far higher theoretical throughputs than their predecessors, which relied upon either the Spanning Tree Protocol or L3 to manage traffic flow. In this new infrastructure, switches can be clustered in active/active configurations—virtually eliminating network links to servers that lie dormant in traditional designs. The introduction of technologies like Shortest Path Bridging (SPB) and Transparent Interconnect of Lots of Links (TRILL) will further enhance the virtualized network backbone paradigm.

**Legacy 3-2-1 and Flat Layer 2 Network Designs**

In the past, traditional client-server based data centers consisted of stand-alone servers supporting static applications. In this type of network structure, the traffic flowed upward and out of the network to communicate with computers and endpoints outside the data center. In comparison, the majority of network traffic in contemporary data centers is horizontal and within the data center. This is due to inter-host, intra-VM, and synchronized inter-VM servers communicating with each other within the same network and also across geographically distributed networks.

Previous network design methodology supported L3 solutions over L2 solutions for the following reasons:

- Broadcast traffic could be isolated or limited so that the network traffic is reduced.
- It is easier to scale the network with L3 solutions.
- The network is much more resilient with an L3 solution compared to an L2 solution.
- The ability to address disparate needs in various customer environments.

In the traditional three-tier data center network architecture, networks consisted of the access layer, aggregation layer, and core switches. In this legacy design, horizontal data flow between servers was restricted to travel at least two hops in the network.

With cloud computing and virtualization appearing in the data center, it is estimated that as much as 75% of the traffic flow is horizontal between servers in the data center. It is also estimated that 50% of a data center’s cabling is used in connecting network devices.

With the introduction of virtualization and the ability to move virtual machines from one physical server to another, using vMotion for example, retaining the original IP addresses has become an important requirement. If a virtual machine is moved from one data center to another, the IP address has to be changed manually. In an L2 solution, this can be achieved without having to change any IP addresses, resulting in a simpler virtual machine management environment. For a visual description, see Figure 6 (“Traditional 3-tier vs. collapsed virtualized design”).

There are challenges with L2 based solutions that directly affect the move to a converged network infrastructure:

- A significant increase in broadcast traffic, larger broadcast domains and loops.
- Changes in the topology (STP/MSTP/RSTP).
- The Spanning Tree Protocol (STP) was designed when hubs were used to interconnect networks, and its intent was to remove bridge loops in the network. STP ensures that there is a single active
path to each network device. The STP process shuts down all other alternative paths. The consequence is that there is a loss of network bandwidth due to single path links.

- Though Multiple Spanning Tree Protocol (MSTP) addresses some of the limitations, it needs careful design and implementation.

Figure 6: Traditional 3-tier vs. collapsed virtualized design

---

**VLANs**

VLANs are usually used in data centers to group a set of servers or storage devices into logical networks on the LAN. VLANs help address issues such as scalability, security, and network management. VLANs are L2 broadcast domains that exist only within a set of switches. Using the IEEE 802.1q standard as an encapsulation protocol, each packet is marked with a unique VLAN tag. Tagged packets are then forwarded and flooded only to stations in the same VLAN. Tagged packets must be forwarded through a routing device to reach any station belonging to a different VLAN. VLAN grouping of a switch or a switch port can be dynamic or static. Another way of creating a VLAN group is to group certain types of traffic that are then forwarded through specific ports based on the data protocol used. For example, more demanding traffic, such as VoIP traffic, can be segmented from other traffic and grouped into a VLAN that receives a higher QoS.

**Layer 2 between Data Center Sites**

Until recently, inter-Site L3-based solutions were preferred to L2-based solutions for the following reasons:

- Easier to scale networks.
- Isolation and limitation of broadcasts.
- L3 solutions are more resilient to failures compared to L2 solutions.
Current data center deployments often require geographic redundancy. This means moving resources and services, such as VMs, from one data center to another quickly and seamlessly. To achieve this, their networks must reside in both data centers.

Key factors influencing an L2-based solution are:

- The ability to move VMs from one data center to another (Long Range vMotion).
- The ability to move servers/machines without requiring IP addresses to change virtual data center hosting.
- Data center consolidation often eliminates existing separate data centers.
- L2 application state through system heartbeats.

Linking data centers to support a large L2 network can be seen in Figure 7.

Figure 7: Illustrates how data centers can be linked to support a large L2 network

Layer 2 Networking Considerations
HP’s support for standards gives HP Ethernet products the ability to interoperate and enhance existing networks that may not be entirely deployed with HP Ethernet switches.

Below is a list of ethernet technologies that HP supports or is helping develop to address the needs of a virtualized environment:

- STP – Spanning Tree Protocol
- SPB – Shortest Path Bridging
- TRILL – Transparent Interconnection of Lots of Links
- OpenFlow
- VPLS – Virtual Private LAN Services
- IRF – Intelligent Resilient Framework
All of these enhancements are either planned to be or are currently supported on HP switches.

**The Spanning Tree Protocol**

The Spanning Tree Protocol (STP) is not a new networking protocol. It is generally the core of the 3-tier network architecture. Many of the enhancements listed in this section were developed to move away from the limitations of STP. STP tends to be inefficient and has a slow convergence period.

STP is a network protocol designed to provide a loop-free bridged ethernet. It is an L2 protocol, so it is not routable, and receives updates from other switches on their link conditions and path costs based on Bridge Protocol Data Units (BPDU). The primary function of STP is to prevent bridge loops and broadcast proliferation. STP also allows networks to be designed with spare (redundant) links, providing automatic backup paths if an active link fails.

STP is standardized as IEEE 802.1d and creates a spanning tree within a mesh network of connected layer-2 bridges, and disables those links that are not part of the spanning tree, leaving a single active path between any two network nodes.

STP can take up to 50 seconds for switches to be notified of a topology change. This is unsatisfactory for use in a converged network infrastructure. An enhancement was made to this protocol which makes the old STP standard obsolete, called Rapid Spanning Tree Protocol (RSTP). Although definitely an enhancement, it can take up to 2 seconds to respond to topology changes in an ethernet network. This is too long for latency-sensitive network data storage and VM applications.

**Figure 8: STP**

Only one path can be allowed to a bridged ethernet switch.
**Shortest Path Bridging**

An IEEE solution, it uses IS-IS as an L2 MAC in MAC learning mechanisms on ethernet networks. It uses a link state protocol to advertise both topology and logical network membership. Packets are encapsulated at the edge either in mac-in-mac 802.1ah or tagged 802.1q/802.1ad frames and only transported to other members of the logical network. Unicast and multicast are supported and all routing is on symmetric shortest paths. Many equal cost shortest paths are supported.

**Benefits:**

- The ability to use all available physical connectivity.
- Only directly affected traffic is impacted during a link failure; all unaffected traffic continues to flow.
- Rapid restoration of broadcast and multicast connectivity.
- Suitable for VM deployments due to emulation of a transparent Ethernet LAN segment.
- Suitable for geographically distributed data center connectivity by providing an ethernet VPN segment.
- SPB segments are easily bound to virtual machines.

**TRILL**

TRILL is an IETF protocol implemented by devices called Routing Bridges or R Bridges. TRILL combines the advantages of bridges and routers providing link state routing at L2. R Bridges are compatible with previous IEEE 802.1 customer bridges, as well as IPv4 and IPv6 routers and end nodes. They are invisible to current IP routers and, like routers, R Bridges terminate the bridge spanning tree protocol.

R Bridges run a link state protocol amongst themselves. A link state protocol is one in which connectivity is broadcast to all the R Bridges, so that each R Bridge knows about all the other R Bridges, and the connectivity between them. This gives R Bridges enough information to compute pair-wise optimal paths for unicast, and calculate distribution trees for delivery of frames either to destinations whose location is unknown or to multicast / broadcast groups. The link state routing protocol used is IS-IS.

**Advantages**

- Runs directly over L2.
- Easy to extend by defining new TLV (type-length-value) data elements and sub-elements for carrying TRILL information.

Temporary loop issues are alleviated by R Bridges forwarding based on a header with a hop count. R Bridges also specify the next hop R Bridge as the frame destination when forwarding unicast frames across a shared-media link, which avoids spawning additional copies of frames during a temporary loop. A reverse path forwarding (RPF) check and other checks are performed on multi-destination frames to further control potentially looping traffic.

Even though R Bridges are transparent to L3 devices, and all the links interconnected by R Bridges appear to L3 devices to be a single link, R Bridges act as link routers in the sense that in the forwarding of a frame by a transit R Bridge the outer L2 header is replaced at each hop with an appropriate L2 header for the next hop, and the hop count is decreased. Despite these modifications of the outer L2 headers and the hop count in the TRILL Header, the original encapsulated frame is preserved, including the original frame’s VLAN tag.

Multipathing of multi-destination frames through alternative distribution tree roots and ECMP (Equal Cost Multi Path) of unicast frames are supported. Networks with a more mesh-like structure will benefit to a greater extent from the multipathing and optimal paths provided by TRILL than will networks with a more tree-like structure.
OpenFlow
OpenFlow is a communications protocol that gives access to the forwarding plane of a switch or router over the network. In simpler terms, OpenFlow allows the path of network packets through the network of switches to be determined by software running on a separate server. This separation of the control from the forwarding allows for more sophisticated traffic management than feasible today using ACLs and routing protocols. OpenFlow is considered an enabler of Software Defined Networking. The OpenFlow standard is maintained by the OpenFlow Switch Consortium at Stanford University.

Virtual Private LAN Services (VPLS)
Virtual Private LAN Services (VPLS) deliver multipoint-to-multipoint connectivity between sites over a managed IP/MPLS (Multiprotocol Label Switching) network in a single bridged domain. See Figure 9 ("Virtual private LAN services"). Geographically dispersed sites can interconnect and communicate over MANs or WANs as if they are on the same LAN, promoting network resilience and response time. When combined with ethernet, VPLS transforms the MAN into a virtualized ethernet switch. This results in cost savings and operational efficiency while delivering all of the advantages of a true VPN network, and also provides the ability to migrate resources and services from one data center to another.

While MPLS is typically associated with IP VPNs, it is also the backbone transport mechanism for virtual private LAN services. Thus, it is possible to use the same infrastructure created for MPLS segmentation to create large L2 networks.

Figure 9: Virtual private LAN services

VPLS offers several benefits for connecting geographically separate data centers. Because VPLS provides a flat network, deployment of services and management of the network are simplified. It operates on L2 and makes VPLS present itself like an ethernet switch. This allows network administrators to build higher-level networks, rather than building their networks around specific IP addresses.
VPLS uses edge routers that can learn, bridge, and replicate on a per-VPLS basis. These routers are connected by a full mesh of MPLS label switched path (LSP) tunnels, enabling multipoint-to-multipoint connectivity. Because the interconnection is created by direct access between the remote sites, there is less traffic overhead and, consequently, improved performance for tasks such as disaster recovery and backup. Also, because data centers can be connected in a mesh, even if one data center loses a network link, all other sites will still be connected.

VPLS solutions also offer Class of Service (CoS) options, and the network administrator can configure VLAN tagging and define priorities.

**Intelligent Resilient Framework (IRF)**

Traditional three-tier ethernet networks were designed with hierarchical, modular L2/3 ethernet switches with the Spanning Tree Protocol, and IP to accommodate for link path management, flow control, and destination routing. These types of designs are functional, but cannot meet the demands of today’s converged network infrastructure.

Purpose-built HP networking solutions and technology are streamlining the design of next-generation data centers to ensure superior resiliency, performance, and agility. One HP innovation is Intelligent Resilient Framework (IRF), a technology that enhances ordinary ethernet switching designs, allowing substantial improvements in the way ethernet switches communicate. HP IRF provides the ability to flatten data center and campus networks, eliminating the need for multiple tiers of aggregation switches and unutilized data paths. IRF provides a framework that enhances ethernet and provides better link management, utilization, and redundancy. See **Figure 10** (“Network topology with and without IRF”).

---

**Figure 10: Network topology with and without IRF**

![Network topology with and without IRF](image)
Why Is There a Need for IRF?
In a network of interlinked switches, the Spanning Tree Protocol (STP) is used to detect and prevent loops – a highly undesirable, sometimes disastrous, situation that can occur when there are multiple active paths to the same switch. To eliminate loops, STP and its more modern variants, such as the Rapid Spanning Tree Protocol (RSTP) and the Multiple Spanning Tree Protocol (MSTP), are designed to allow only one active path from one switch to another, regardless of how many actual connections might exist in the network. If the active path fails, the protocol automatically selects a backup connection and makes that the active path.

Figure 11: An example of the Spanning Tree Protocol

While STP is fairly effective in making a network resilient and simpler to configure compared with MSTP or RSTP, network convergence can still take several seconds, affecting applications that cannot handle that length of delay. In addition, the performance of STP is poor because it blocks all parallel paths except the one it has selected as active. Even when the network is operating normally, STP actually reduces the effective bandwidth (possibly to a degree greater than 50%).

IRF is a technology that can provide a network that is fully resilient, yet also simpler to setup and manage, faster to converge, and easier to scale. IRF simplifies network operations by consolidating management of multiple, discrete devices into a single, easy-to-manage virtual switch, in every layer of the network.

IRF Operational Fundamentals
Think of IRF as a framework that ethernet operates on, rather than a protocol that operates on top of ethernet. This is not to say that you cannot still make use of the benefits of SPB or Trill in a data center bridged ethernet network, but IRF does so much more than L2 Ethernet protocols. It was designed with a lot more in mind than just the replacement of STP.

IRF technology extends network control over multiple active switches. Management of a group of IRF-enabled switches is consolidated around a single management IP address, which vastly simplifies network configuration and operations. You can combine as many as nine HP A-Series switches to create an ultra-resilient virtual switching fabric comprising hundreds or even thousands of 1GbE or 10GbE switch ports.
This framework was designed to support the growing need for converged network infrastructures that demand:

- High capacity links like 10G, supporting FCoE.
- Rapid switch to switch topology updates.
- High performance frame forwarding fabrics.
- Low latency communications.
- Ethernet trunk aggregation with flow control for mission critical data types.
- Ease of provisioning.
- Reduced management complexity.

One IRF member operates as the primary system switch, maintaining the control plane and updating forwarding and routing tables for the other devices. As mentioned before, IRF is a framework that Ethernet operates on top of. It not only enhances the L2 functionality, but also improves L3 performance. If the primary switch fails, IRF instantly selects a new primary, preventing service interruption and helping to deliver network, application, and business continuity for critical applications. Within the IRF domain, network control protocols operate as a cohesive whole to streamline processing, improve performance, and simplify network operation.

Routing protocols calculate routes based on the single logical domain rather than the multiple switches it represents. Edge or aggregation switches that are dual-homed to IRF-enabled core or data center switches “see” associated switches as a single entity, eliminating the need for slow convergence technologies such as STP. And operators have fewer layers to worry about, as well as fewer devices to configure and manage.

**Advantages of IRF**

**Design and operational simplification** – With IRF, no longer do you connect to, configure, and manage switches individually. Configurations are performed on the primary switch, and that configuration is distributed to all associated switches automatically, considerably simplifying network setup, operation, and maintenance. While all HP A-Series switches can be provisioned via the command line, adding HP Intelligent Management Center (IMC) makes management even easier. IMC lets you see and control the entire network from a single console by consolidating management of multiple, discrete devices into a single, easy-to-manage virtual switch that operates in every layer of the network.

**Flexible topology** – IRF provides a simplified, higher performing, more resilient, and flatter network design. IRF and HP A-Series switches allow enterprise networks to be designed with fewer devices and fewer networking layers – a big improvement over the low performance, high cost, and crippling latency of conventional multi-tier legacy solutions, which often rely on a variety of different operating systems and complex resiliency protocols. IRF provides a common resiliency mechanism in both Layers 2 and 3, eliminating the need for slower converging protocols like STP and VRRP. Also, unlike other L2 Ethernet network frameworks that require the use of tunneling to extend VLANs, IRF allows network operations and design to “get back to basics” by providing a simple, yet flexible architecture based on standards-based 802.1q tagging, Q-in-Q tagging, and basic but effective VLAN mapping. In addition, IRF supports advanced higher layer protocols like SPB, TRILL, and MPLS. This provides the most flexibility while maintaining industry leading resiliency and performance.

**Higher efficiency** – IRF’s loop-free, non-blocking architecture keeps all links active, enabling highly efficient, high-bandwidth connectivity throughout the switching plane. No longer is it necessary to have switch to switch ethernet trunks go unused in a standby mode in STP, or a partially utilized path with SPB or TRILL protocols. Existing networks that have legacy three-tier ethernet architectures can be rapidly integrated into the IRF framework, immediately enhancing the performance and resiliency of the network.
Scalable performance – IRF and Link Aggregation Control Protocol (LACP) work together to boost performance and increase resiliency by bundling several parallel links between switches and/or servers, allowing scalable “on-demand” performance and capacity to support the demands of converged network infrastructures. In addition, IRF is fully supported with 3rd party devices, as it leverages standards-based link aggregation for connectivity.

Faster failover – Should a network failure occur, IRF can deliver rapid recovery and network re-convergence in under 50 milliseconds – much faster than the several seconds when STP or even RSTP are utilized.

Geographic resiliency – Geographic separation is no longer just the domain of the WAN. Data centers are ever-increasing in size with extremely large campuses. With the evolution toward converged network infrastructures and large flattened L2 networks, the need to access server and I/O resources across a local campus involving substantial distances is becoming more common. IRF meets these demands, as well as providing connectivity into MANs and WANs. Within an IRF domain, the geographic location of switches does not matter. Switches can be extended horizontally, and they continue to function as a single logical unit whether they are installed locally, distributed regionally, or even situated at distant sites. Moreover, employing IRF can enhance disaster recovery by linking installations up to 70 kilometers apart, providing the same fast failover as if they were sitting side by side within the data center.

In-Service-Software-Upgrade – IRF delivers a network-based In-Service-Software-Upgrade (ISSU) capability that allows an individual IRF-enabled switch to be taken offline for servicing or software upgrades without affecting traffic going to other switches in the IRF domain.

Section Summary
Virtual networking technologies are becoming increasingly important (and complex) because of the growing virtual networking infrastructures supporting virtual machine deployment and management. Virtual I/O technologies can be implemented in the software space by sharing physical I/O among multiple virtual servers (vSwitches). They can also be implemented in hardware by abstracting and partitioning I/O among one or more physical servers. New virtual I/O technologies to consider should be based on industry standards like VEPA and VEB, and should work within existing data center frameworks.

HP is not only involved in the development of standards to make the data center infrastructure better, it is also designing the standards-based products that support it. HP has already developed products which work to simplify the server-to-network edge by reducing cables and physical infrastructure (BladeSystem). As the number of server deployments using virtual machines continues to increase, the nature of I/O buses and adapters will continue to change. HP is positioned well to navigate these changes because of the company’s skillset and intellectual property it holds in servers, compute blades, networking, IRF, storage, and virtualized I/O.

It is important to keep in mind that with the increasing adoption of blade servers and virtualization, the aggregation layer is moving in two directions (in the chassis and to the core). This is due to the fact that the traditional Top of Rack (ToR) and End of Rack (EoR) deployments are not necessarily needed, and add complexity to the network when it is not necessary. First, HP core ethernet switches support blade modules with 32 10G ports. Secondly, more traffic is traversing horizontally within a BladeSystem chassis, and each blade server chassis’ traffic can be aggregated up to the core through the use of FlexConnect fabric. This reduces the physical port capacity requirements in the core. Thirdly, LACP can be used to add substantial amounts of network bandwidth to a BladeSystem, while connecting them to aggregation and/or core ethernet switches operating within extremely resilient IRF frameworks.
Securing the Virtual Server Edge

Setting technology aside for the moment, a data center is a location where individuals consider their data safe from theft and environmental disaster. A data center’s ability to provide a reliable and safe infrastructure for applications and data is critical. If a system is unavailable for some reason, business operations will be impacted. Because data centers house critical data, it would be logical that in addition to taking other precautions to protect servers physically, information security should possess the same importance.

Data security is designed or implemented to prevent or minimize security breaches. Traditional enterprise security controls do not take into account the complexity of the network and the systems that they are designed to protect. As in any control, security can have a negative impact on performance and data center design flexibility.

Within a network, internal and external threats continue to grow while our networks become more complex. Today, there are three key security implications that increase the complexity of data center security – compliance, convergence, and consolidation.

Compliance
Compliance has gained traction as organizations deal with regulations to mitigate the risk in data loss and application downtime. Companies need to comply with regulations such as the Payment Card Industry Data Security Standard (PCI-DSS) for the retail industry, and the Health Insurance Portability and Accountability Act (HIPAA) for the healthcare industry. These regulations deal with data security and privacy, and data centers need to ensure that adequate security measures are in place.

Regulatory Compliance & Industry Controls:

- **(HIPAA) The Health Insurance Portability and Accountability Act 1996** – Applies to health plans, health care clearinghouses, and those health care providers who conduct electronically certain financial and administrative transactions that are subject to the transactions standards adopted by the Department of Health and Human Services.

- **(SOX) The U.S. Sarbanes-Oxley Act of 2002** requires CEOs and CFOs to attest to the accuracy of corporate financial documents, as well as provide IT & Networking controls and their audit as per Section 404 of the Act.

- **(GLBA) The Gramm-Leach-Bliley Act of 1999** – Banking and financial industries control. Institutions must provide clients a privacy act that explains information collection, sharing, use, and protection.

- **(PCI) The Payment Card Industry Security Standards Council** – Controls set and regulated by financial institutions for electronic credit card information and transactions.


- **(NERC) The North American Electric Reliability Corporation** – Standards and guidelines established for the national electric grid and its operation. IT and Network controls are a major component of the guidelines for security.


- **(BASEL II) The international accord on banking operations** – Of which information security and assurance are major components.

All of the outlined regulatory and industry standards affect the operations and design of security systems for data centers. As converged network infrastructure and virtualized systems become more prevalent in the data center, it is important to identify the risks, the controls, and the weaknesses. It is important to then close the gaps and continue to reassess the network’s security readiness.
Data centers need to ensure adherence to compliance standards by ensuring that adequate measures are in place to ensure data security and privacy. The steps involved in meeting this requirement are:

- Identifying the vulnerabilities and threats, and setting up the appropriate security controls.
- Quantifying the risk exposure: This includes identification of the vulnerabilities and threats, identification and definition of their compensating controls, and the calculation and documentation of risk.
- Creating an IT risk management plan.
- Ensuring that all the IT compliance requirements are met.
- Ensuring that all the personnel are trained in IT security awareness.

**Convergence**

The convergence of multiple security functions into products that support a single function has evolved to include policy management integration. The three security product functions—firewalls (legacy as well as next generation firewalls), intrusion prevention systems (IPS), and content security gateways—are converging into an integrated perimeter and data center security function. In addition to these, other security product functions such as Security Information and Event Management (SIEM) tools, and Network Behavior Anomaly Detection (NBAD) tools are also being integrated into policy management functions. Integration of these different types of products simplifies the management of network security and reduces the cost of overall security. However, integrated security products cannot compromise security for the sake of integration and convergence.

In the context of data center security, convergence refers to the integration of various tools such as firewalls, intrusion prevention systems, and monitoring systems into an integrated tool. While this may simplify management and monitoring, care should be taken to ensure that security is not compromised.
There are two elements of convergence that need to be considered:

- **Data centers can no longer rely on a single security perimeter around the enterprise network.**
  
  Individual assets, such as a virtual machine or a host in the virtualized environment, need to be protected. In order to achieve this, purpose-built intrusion prevention systems need to be considered that would inspect both ingress and egress traffic and block threats automatically. Additionally, the products should support automated updates so that newly discovered threats are addressed and the network ‘vaccines’ are updated automatically.

- **Division and convergence of security responsibilities.**
  
  Network teams are usually responsible for network security and the operations/server teams for OS security. With the introduction of virtualization, some network elements are no longer physical devices and may form part of the virtualized environment, such as the vSwitch. As this is a software component, it would usually fall within the responsibilities of the operations team. The operations team may not necessarily be aware of how to implement network security and, as a consequence, security at the vSwitch may not be enforced correctly. To address issues such as these, the network team, the server team, and the security team need to work together closely to look at all aspects of security.

**Consolidation**

The third key security trend is the consolidation of data center infrastructure. In the past, data centers were built for the purpose of bringing teams in the organization closer to the applications and data that they required, resulting in a dispersed data center infrastructure. New applications such as web applications, new protocols such as IPv6, and new traffic types such as voice and video traffic increase the complexity of the security function. The consolidation of data types and security functions onto similar network equipment in data centers requires IT teams to not only understand the technology and the management functions, but understand the impact they have across technical teams.

The complexity and criticality of data center security has only increased. Hackers not only seek to cause disruption in the data center, but in many cases cause governmental, commercial, and individual mayhem and harm. In the past, these threats were typically handled by IT security teams with a single security perimeter comprised of both firewalls and antivirus products.

However, threats are now more sophisticated and many times are caused by professionals who seek to steal data, modify it, and illegally distribute it for financial, political, or other gains. In addition, the rise of new applications and services such as file sharing, database access, and video conferencing has contributed to a new set of security weaknesses. As a result, data centers can no longer rely on a single security perimeter around the network; separate security perimeters around individual assets in the network and in the data center need to be built.
Without reviewing the advantages of converged network infrastructure and server virtualization, it is important to note an associated negative trend that goes along with the increase in virtualized data center resources. This trend is not necessarily caused by negligence, but more through a lack of understanding regarding the impact of virtualization on underlying security. Gartner states that “60 percent of virtualized servers will be less secure than the physical servers they replace through 2012.”

In an attempt to cut costs, companies are consolidating their data centers by reducing the number of geographically separate locations. In addition, the number of physical servers in each data center is being reduced due to virtualization technologies. Here, we take a look at some of the challenges introduced by virtualization.
A virtualized environment comprises a physical server (host) that hosts multiple operating systems (guests) and enables them to run concurrently. As discussed earlier, management of VMs is carried out by a hypervisor, which presents itself to the guest operating system as a virtual operating platform and monitors the execution of the guest operating system. Multiple guest operating systems share the same virtualized resources, such as the CPU and the ethernet port.

Several challenges on the security front are introduced by the hypervisor and the virtual switch:

**Hypervisor Security** – Securing the hypervisor is critical, as a compromised hypervisor could lead to a situation where access to the hosted operating systems may be made available to attackers. Another possibility is a Denial of Service (DoS) attack on the hypervisor, which could affect all the hosted virtual systems.

**Virtual Switch Security** – The virtual switch is a software component that allows communication between virtual machines and other systems. The packets are examined and sent to other virtual machines or to the physical switch, depending on the destination. In a virtual environment, a factor to be considered is that the network team may not have visibility to all of the network traffic. For example, the communication between two virtual machines on a single physical server would not reach the physical switch and, as a result, the traffic may not even be detected.

Another area to consider is mobile workloads. VMs may move from one physical server to another due to situations such as an increased load that dynamically shifts the VM onto another physical server supporting the hypervisor, or a physical server failing, as depicted in the following figure:
Due to this kind of mobility, unless adequate controls and monitoring methods are established, it may be difficult to implement policies. Policy and zone definitions must follow the mobile workloads. In particular, security policies must be maintained for the different workloads at rest, operating, and in motion.
"Don’t forget to still implement firewalls, virus & malware scanners for data, instant messaging, and files attached to instant messages."

The figure above depicts:

- An IPS device, external to the physical server that hosts multiple virtual machines at the network level.
- A virtual IPS that resides on the host.

Virtual and physical IPSs can be used separately, but the benefit of combining the two provide the advantage of monitoring traffic at the network as well as the virtual layers. One such product that provides a superior platform is HP TippingPoint’s IPS solution.

As discussed above in relation to mobile workloads, the policies should be ideally applicable to workloads at rest, while operating, and in motion. A product that can integrate disparate management tools and provide the management of resources, services, and users simplifies the task of implementing policies. One such product is the HP Intelligent Management Center (IMC).

**Virtual Machine Security**

All of the traditional security duties must be accomplished to secure any data center. On top of the standard duties, data center security teams need to secure the virtual platforms. HP TippingPoint has a suite of solutions to support a converged network infrastructure and virtual machines.

The HP TippingPoint Secure Virtualization Framework (SVF) is a suite of products designed to help prevent network threats from impacting virtualized environments. Because several virtual machines (VMs) are hosted on a single physical server, a security breach on a VM can impact the other VMs on that server. We have discussed the components of virtualization in Section 4.5, and the challenges introduced by virtualization. In this section, we take a look at the components of the Secure Virtualization Framework and how they help address the challenges.
The Secure Virtualization Framework is comprised of the following components:

IPS Platform – The IPS platform is an in-line device that examines packets as they flow through it and blocks malicious packets. It also includes DVLabs, which is the security research organization that supplies the security filters to customer IPS platforms to keep the IPS devices updated with the latest security vaccines.

Virtual Controller (vController) – The vController extends security protection from physical to virtual networks by routing traffic through an HP TippingPoint N-Platform Intrusion Prevention System (IPS) appliance. The vController prevents security attacks by inspecting all VM traffic as it moves through the network, either between VMs or from VMs to traditional servers.

Virtual IPS (vIPS) – The vIPS is a virtual appliance that provides the same IPS capabilities as the IPS platform. This leverages the resources on the host system and can provide security in cloud environments or added security in virtualized environments in combination with the IPS Platform.

Security Management System (SMS) – The TippingPoint Security Management System is a hardened appliance responsible for discovering, monitoring, configuring, diagnosing, and reporting for multiple TippingPoint systems. It features a secure Java client interface that enables analysis with trending reports, correlation and real-time graphs on traffic statistics, filtered attacks, network hosts and services, as well as IPS inventory and health.

Virtual Management Center (VMC) – The VMC is used to automatically discover every VM in the data center and deploy vController and the virtual firewall on each virtualized host. This ensures appropriate security policies are dynamically applied to and enforced by vController and the IPS Platform for all deployed/discovered VMs.
Managing and Provisioning the Virtual Server Edge

Network Infrastructure convergence and virtualization have allowed multiple networks and systems to be collapsed into unified entities. This trend, from the outside looking in, may appear to have simplified the tasks and processes associated with data center systems and network management, but in many cases it has added complexity.

With converged infrastructure, the first issue to be identified is one of responsibility and ownership. As systems and network technologies merge, the responsibility to install, provision, maintain, and monitor still exists. This is not a new topic, it has been seen in the merger of voice and data into the enterprise network, and has been addressed by multiple management experts and forums. This section will focus on the identification of the technical hurdles and the tools needed to manage systems and networks in the converged infrastructure data center.

The converged infrastructure data center will not be comprised 100% of VMs and CCE networks. Currently, there appear to be fewer physical entities in data centers to manage, and hardware consolidation is going to continue. But, keep in mind the term “convergence”: As the systems and network services converge, so must the management platforms. They will also become true converged platforms.

“Single pane of glass” management is that converged platform. This is what data centers have been trying to achieve with Simple Network Management Protocol (SNMP) and agent driven management software suites. At the same time, the roles and responsibilities of the data center network and systems managers have blurred. The lines of responsibility are eroding, and management systems are being required to evolve alongside that merging of responsibilities. Since the role of data center technical teams is changing, management systems too must change. When evaluating management platforms, there are hard questions to ask:

• Do these management systems really provide a simpler management platform to work with, or do they just provide the ability to view and receive alerts from every device in the network?
• Do these platforms deliver an interface that allows for native system management routines all aggregated together as though they are one integrated software system?
• Are these just portals that aggregate the systems and network management tools that multiple vendors provide onto a manager of managers?

A fully functional management platform must provide the following:

• A single view of all systems with access to alarms and configuration tools.
• Secure communications to the managed endpoints.
• Support for secure access to individual VMs and the underlying virtualization layer.
• Multiple types of system and network reporting schemas.
• Flexible workflow configurations that management teams can customize for their organizations.
• Flexible user, group, and role/responsibility security configurations.
• A native application look, feel, and performance.
• Availability and resilience.

These capabilities will be addressed in greater detail later in this document, with a focus on what is currently being used by data center managers and the path to “single pane” management.

Virtualization Management

One of the objectives of convergence and virtualization is to make it more manageable. The fewer distinct elements that need to be managed, the less complexity is required. Unfortunately, the management of converged and virtualized network elements did not automatically get converged at the same time. Again, HP’s major goal is to simplify the network without impact to performance and
flexibility. When we look at the virtualized network, storage, and server platforms through respective tools, they should provide a combined view of the independent pieces, thereby reducing the complexity of management.

Tools such as HP’s Intelligent Management Center (IMC) provide the data center operations team a combined view of the entire infrastructure, enabling them to provide efficient end-to-end business management to address the stringent demands of today’s mission-critical enterprise IT operations.

**Solving VMware Management**

IMC integrates with the VMware vCenter Server 4.1 and allows for centralized management of hundreds of VMware ESXi/ESX hosts and thousands of virtual machines, delivering operational automation, resource optimization, and high availability to IT environments. Using a single management client for all tasks, administrators can provision, configure, start, stop, delete, relocate, and remotely access virtual machines.

HP, along with VMware, provides solutions for creating standardized and repeatable processes. This includes managing capacity and performance, controlling configuration changes, protecting business-critical applications against outages, and self-service provisioning for development and testing environments. Users of other virtualization platform vendors must rely on third-party products to receive the same breadth of virtualization management functionality.

**Solving Server Management**

There are multiple tools that aid in the support of server management. While there are multiple tools available, the overriding goal is to fully integrate management tool capabilities through APIs and/or web-enabled applications into an integrated management system like HP’s Intelligent Management Center (IMC). VMware virtual servers and their virtual network components are integrated into IMC, providing a single interface to view, provision, move, and manage the virtualized network.

**HP Integrated Lights-Out (iLO)**

This HP tool simplifies server setup, provides access to server health information, helps optimize server power and thermal control, and enables basic remote administration functions. While iLO ships with standard features enabled, the iLO feature set (virtual KVM console, multi-user collaboration, console record and replay, and 24-hour power measurement and single server Dynamic Power Capping can be unlocked with iLO Advanced.

**Simplified server setup** – iLO makes it possible to configure server boot order, network configuration, RAID settings, and other server parameters, as well as to deploy the server operating system (OS) in minutes. This enables server administrators to setup servers in remote locations without onsite staff, while reducing travel expenses and time away from their desks.

**Embedded server health** – By leveraging iLO embedded health, system administrators can gain deep insight into server component status when the operating system is offline or distressed. With HP ProLiant G6 and G7 servers, HP delivers support for viewing configuration and status of temperature sensors, fans, and power supplies, and configuration data for CPU and memory subsystems.

**Dynamic Power Capping and thermal control** – Today more than ever, IT organizations are concerned with the cost of powering and cooling their servers. The power and thermal controls exposed through iLO allow for the reduction of server power consumption by regulating processor clock speed, and by providing an optional high-efficiency mode for power supplies. IT and facilities teams also need to make sure that they are making efficient use of existing power and cooling capacity. With iLO’s Dynamic Power Capping capabilities, IT and facilities staff can set an upper limit on power consumption and reclaim trapped power and cooling capacity created by overly conservative power budgeting policies. A company that uses faceplate values to budget server power needs can as much as triple the capacity of its data center.
Remote administration – The Integrated Remote Console provides IT staff with the ability to interact directly with the server OS or pre-OS environment without leaving their desks. In addition, IT administrators can take advantage of the multi-user console and video record and playback to collaborate with IT staff across multiple locations more effectively.

Integrating iLO with other HP Tools
iLO integrates with the HP Systems Insight Manager to enable several important features: First, a single sign-on leverages HP SIM’s role-based authentication to automatically login to iLO from the HP SIM console.

Leveraging single sign-on, HP SIM also allows for group actions with iLO virtual power and virtual UID from the SIM menu. Authorized users can control the power and UID of multiple servers by simply selecting the servers and clicking on the desired action in the HP SIM menu. HP SIM and iLO use the new WS Management standard to enable this feature.

Traffic Flow Analysis & Capacity Planning
Analyzing network traffic to determine which applications, servers, and clients are consuming network resources is a best-practice method for understanding application delivery problem root cause issues. Traditional methods of analyzing network consumption, such as link or device utilization, do not provide IT staff with adequate insight into why network resources are being consumed. HP’s Intelligent Management Center (IMC) centralized network platform provides the intelligent collection, analysis, and reporting services that maximize the insight that can be gleaned from flow data to circumvent issues, ensure quality of service (QoS), and expedite remediation.

NetFlow/sFlow/Netstream
The flow data produced by networks generally comes in one of three common formats – NetFlow, sFlow, or NetStream. Standardization around these formats makes it possible for routers and switches to send their flow data to a wide variety of collectors and analysis tools, and to be combined with flows in multi-vendor networks for wider analysis. Flow data has now become an important part of network scalability and performance, particularly in busy router cores and edge devices that handle a large number of concurrent, short-duration flows.

NetFlow is the oldest of the flow formats. It originally served as a caching algorithm in network devices, which helped optimize network efficiency. As this data was already being collected, it made sense to eventually export it for analysis and reporting purposes, which could be done without much additional overhead to the network device. NetFlow has spawned various iterations – it is now up to version 9 – as well as similar formats optimized for different purposes and processing.

sFlow was created as a standard in 2001 for high-speed networks based on sampled data rates rather than 100% packet capture. sFlow was developed exclusively as a monitoring technology. It is scalable and can provide more detailed statistics on all Layers L2-7 throughout the network. As a result, it has gained wide acceptance from network vendors. sFlow is fully implemented in all HP Networking devices today.

NetStream, a flow format created by 3Com (now HP) for its enterprise networking products, includes additional flow details. NetStream provides detailed flow information, compatible with NetFlow and is implemented in HP Networking routers and available on high-end switching platforms via an add-on module.

HP’s Intelligent Management Center (IMC), as the collection and analysis engine, can handle flow data in all of these formats as well as other formats from a wide range of devices from many manufacturers in order to provide network-wide visibility.
SNMP
With the rapid development of the Internet, two problems were introduced:

- Increasing in the number of networks and network devices makes it difficult for administrators to monitor the status of all the devices in time, and to identify and correct network faults.
- Network devices may be from different vendors, providing an independent set of management interfaces (e.g., command lines), making network management more complicated.

SNMP is an application-layer protocol between an NMS and agents introduced to solve the above-mentioned issues. It defines the standardized management framework, common languages in communication, security, and access control mechanisms used in monitoring and managing devices in a network. The administrators can query device information, modify device parameters, monitor device status, and enable automatic detection of network faults and generation of reports by using SNMP.

SNMP provides the following advantages:

- A TCP/IP-based standard protocol, with UDP as the transportation layer protocol.
- Automatic network management: Administrators can search and modify information, find and diagnose network problems, plan for network growth, and generate reports on network nodes.
- A combination of simple request-reply mode and active notification mode, timeout, and retransmission mechanism.
- Few packet types and simple packet format, which facilitates resolution and implementation.

HP’s Intelligent Management Center (IMC) has a built-in SNMP collector that allows it to receive SNMP messages and statistics from any SNMP device in the network, including switches, routers, servers, NAS, etc.

Single Pane of Glass Management
HP’s Intelligent Management Center (IMC) is a next-generation management software which provides the data center operations team with a comprehensive platform that integrates network technologies and provides full fault, configuration, accounting, performance, and security management functionality.

Built from the ground up to support the Information Technology Infrastructure Library (ITIL) operational center of excellence IT practices model, IMC’s single-pane management paradigm enables efficient end-to-end business management to address the stringent demands of today’s mission-critical enterprise IT operations.

Configuration Management – Backup
Configuration management can be defined as the ability to control changes to the operating status of a managed device, as well as the ability to detect and/or prevent unauthorized changes in the operational status of a device. Maintaining an accurate inventory of last known hardware, software, and configuration information enhances this function.

To manage a data center, the operations team must have an up-to-date configuration inventory across all types of devices, irrespective of vendors. IMC provides the capability to manage the configuration of over 2700 devices from over 30 different manufacturers.
The time required to roll out network changes and the likelihood of configuration errors are both greatly reduced with IMC’s powerful Bulk Configuration functionality. Configuration baselining ensures that changes to the stable network configuration are flagged promptly.

The powerful configuration comparison feature provides the rapid identification of configuration differences, enabling the System Administrator to either accept the new configuration or roll back to the original stable configuration. Additional functionality includes Bulk Backup and Restore, an extremely flexible agent management function that enables the System Administrator to have total control of the upgrade process.

The ability for change detection and the system inventory also act as informational resources for the fault management process. Understanding how a device is supposed to be configured enables the database operations staff to recognize and correct faults. Change recommendations can only be made after the current configuration status is completely known, understood, and stable.

Configuration management is both reactive (changes can be identified over time) and proactive (control can be established to prevent configuration changes). Small configuration changes can go unnoticed on a LAN or WAN, and they can completely collapse an entire nationwide network. Arguably, configuration management is the most important component of the FCAPS model.

**Traffic Analysis and Capacity Planning**

As the enterprise network infrastructure expands to support different types of traffic and users, traffic management becomes critical, and complete visibility into a network’s behavior becomes more important and more challenging. What is or is not happening throughout the network grid—including application performance, bandwidth utilization, network congestion, and appropriate prioritization of user and application traffic—are questions that often go unanswered.

In today’s connected business environment, straightforward and effective traffic management from the network core to the network edge is essential. Enterprises need a network infrastructure that scales to meet new business needs and manages added complexity in a cost-effective manner. In addition, the data center operations team is expected to control the network in such a way that it is transparent to users. Essential information assets need to be instantly available around the clock. However, this is impossible to achieve without the right tools to make smart, informed decisions.
Most network administrators do not have simple, affordable tools that can quickly answer the following questions, regardless of the size of the network:

- Is network performance slowing down or becoming congested?
- Is a Network Interface Card (NIC) chattering, effectively clogging the network?
- What is the current network usage, and what has it been in the past hour?
- Which network routers are most active or over-utilized?
- Why is a server slow or inaccessible?
- Which users and applications are driving network traffic?
- Which users and applications are starving for bandwidth?
- How much bandwidth do I need for new applications?

HP’s IMC Network Traffic Analyzer (NTA) is a graphical network monitoring tool that utilizes industry-supported flow standards to provide real-time information about the top users and applications consuming network bandwidth.

Figure 19: Applications consuming network bandwidth
A reliable solution for enterprise and campus network traffic analysis, IMC NTA statistics help network administrators better understand how network bandwidth and resources are being used, as well as which source hosts carry the heaviest traffic. This information is invaluable in network planning, monitoring, optimization, and troubleshooting – IMC NTA identifies network bottlenecks and applies corrective measures to ensure optimal throughput.

IMC NTA utilizes instruments embedded in switches and routers to collect its data. Support for most standard traffic monitoring protocols – sFlow, NetFlow and NetStream – eliminates the need for expensive probes, and reduces the cost of network monitoring.

**Out-of-Band Management**

Out-of-band management (OOBM) operates on a “management plane”, which is separate from the “data plane” used by data traffic on the switch and by in-band management traffic. That separation means that out-of-band management can continue to function even during periods of traffic congestion, equipment malfunction, or attacks on the network. In addition, it can provide improved switch security: A properly configured switch can limit management access to the management port only, which prevents malicious attempts to gain access via the data ports.

Network OOBM typically occurs on a management network that connects multiple switches. It has the added advantage that it can be carried out from a central location and does not require an individual physical cable from the management station to each switch’s console port.

In a typical data center installation, Top of Rack switches connect servers to the data network, while the management ports of those switches connect to a physically and logically separate management network. This allows network administrators to manage the switches even if operation on the data network is disrupted.

In the illustration below (Figure 21), the switches face the hot aisle of the data center, allowing easy connection to the network ports on the backs of the servers.
For even more control, the serial console ports of the switches can be connected to the management network through a serial console server (essentially, a networked serial switch), allowing the network administrators to view the console activity of each switch at boot time and to control the switches through the console ports (as well as through the management ports).

**Port Mirroring**

Port mirroring, also known as SPAN (Switched Port Analyzer), is a method of monitoring network traffic. In port mirroring, the packets passing through a port/VLAN (called a mirroring port/VLAN) are copied to another port (called the monitor port) connected to a monitoring device for packet analysis. The administrator can select to port mirror inbound, outbound, or bidirectional traffic on a port/VLAN as needed. A network administrator uses port mirroring as a diagnostic tool or debugging feature, especially when fending off an attack. Port mirroring can be managed locally or remotely.

In local mirroring, the mirroring ports/VLANs and the monitor port are located on the same device, while in remote mirroring traffic on the mirroring ports or ports in the mirroring VLANs is mirrored to the monitor port located on another device.

Network traffic monitoring is needed for packet analysis or IPS deployment. However, monitoring all the traffic in a large switching network is difficult, so port mirroring can be configured to copy selected traffic of a port or ports to a specific port for monitoring.
Converged Network Infrastructure: Unifying Data and Storage Networks

Convergence is a technical term historically used to express the combining of voice and data onto the same network fabric. Now expressed as a “converged network infrastructure”, it encompasses the sharing of network resources between data and storage networks. This trend constitutes a move towards a unification of data and storage networks.

Network technologies like Fibre Channel, used to connect storage resources to computers, differ substantially from network technologies used to connect computer networks. Although high performance, these network types create two dissimilar data center networks (LAN/WAN and storage), which increases the number of cables and management.

Technologies such as blade servers have addressed this challenge by drastically reducing the number of interconnections. Blade servers have simplified the network by reducing cables and ethernet ports by over 75%. Converged network infrastructure can reduce data center complexity by an additional 50%, using technologies like Fibre Channel over Ethernet (FCoE), and more efficient technologies like Data Center Bridging (DCB), also known as Converged Enhance Ethernet (CCE). See Figure 22 (“Converged infrastructure”).

The Data Center Bridging (DCB) standard will also have a far-reaching impact on how data center networks are designed and deployed. This trend will include similar class and congestion management controls currently found in converged voice and data networks.

Fibre Channel over Ethernet (FCoE) is currently used in many enterprise data center networks. It has provided acceptable performance, while allowing data storage access over ethernet. However, the theoretical bandwidth and performance limit has been reached with this technology. The most common end-to-end bandwidth and congestion management tool is TCP window sizing. With window sizing, TCP dynamically determines the number of frames to send at once without an acknowledgement. It continuously ramps this number up dynamically if the pipe is empty, as long as acknowledgements are being received. But, if a packet is dropped due to congestion and an acknowledgement is not received, TCP halves the window size and starts the process over. This provides a mechanism in which the maximum available throughput can be achieved dynamically.
The TCP sliding windows algorithm does not provide the required low-latency/uninterrupted network performance necessary for the long term goals of FCoE.

In the DCB standard, Quantized Congestion Notification (QCN) is the new form of end-to-end congestion management defined in IEEE 802.1.Qau. The purpose of end-to-end congestion management is to ensure that congestion is controlled from the sending device to the receiving device in a dynamic fashion. Enhanced Transmission Selection (ETS) provides the bandwidth guarantee which deal with network bottlenecks without limiting performance. Priority Flow Control (PFC) will provide lossless packet delivery.

All of these emerging network technologies have an effect on how data centers are being planned in the future, but is also important to understand how these technologies evolved.

The remainder of this section is focused on identifying what is currently used in data center network deployments, as well as identifying HP’s vision of converged network infrastructure.

**Networked Attached Storage**

The requirement for users to access shared data on the network has grown immensely over the past decade. Fibre Channel was developed to meet the high speed and reliability needs of enterprise customers, and therefore was very expensive. A more cost-effective solution was to use the existing IP infrastructure and attach storage devices to it, today known as Network Attached Storage (NAS).

NAS devices can scale to multi-terabyte capacity, but typically don’t meet the performance and reliability requirements of today’s large scale data centers. Protocols like NFS, CIFS (Samba), iSCSI, etc. are the foundation of NAS.

NAS is a dedicated storage device that operates in a client/server mode. NAS is connected to the file server via a LAN. The protocols supported for NAS are Network File System (NFS), and Common Internet File System (CIFS) and use cases are shown below. See Figure 23 (“Typical NAS configuration on a LAN”).

- NFS and CIFS are single threaded and have lower performance in comparison with Storage Area Networks (SANs). NAS is connected over the LAN using TCP/IP.
- Network File System (NFS) – UNIX®/Linux®.
- Common Internet File System (CIFS) – Windows remote file system (drives) mounted on the local system (drives). This evolved from Microsoft NetBIOS, NetBIOS over TCP/IP (NBT), and Server Message Block (SMB).
- Samba – SMB on Linux (making Linux a Windows file server).

**Advantages**

- No distance limitations.
- Bridges the SCSI and Serial Advanced Technology Attachment (SATA) logical unit capacity limitations.

**Disadvantages**

- Performance (although recent advances have closed or eliminated this gap).
- Latency.

**Weakness**

- NFS and CIFS are inherently insecure protocols. Encapsulation of NFS and CIFS traffic into tunnels only adds to the speed and latency issues.
In storage systems, the read/write transactions span from the magnetic domains on the hard disk to the operating system kernel. The OS treats the storage as locally connected, though the data might actually be traversing multiple hops. A storage network mimics the physical path normally traversed by data, from the systems PCI bus, and then to the processing engine and operating system OS. The speed of the storage I/O operations is critical, and delays can result in operating systems hanging, or even crashing. Also, it is critical to ensure that networks utilized to transport storage traffic are reliable and resilient to common errors. See Figure 24 ("Generic network transport architecture"). Note that data requests can be supported from a variety of resources.
Storage Area Networks (Non-IP)

Storage Area Networks (SAN) are specialized, dedicated high-speed networks joining servers and storage. It can include disks, disk arrays, tapes, etc. Storage (data store) is separated from the servers, and the I/O is offloaded from the server’s processors. SAN technology provides high capacity, high availability, high scalability, ease of configuration, and ease of reconfiguration. Fibre Channel is the de-facto standard SAN architecture, although other network standards can be used. SANs differ from NAS in that the Fibre Channel network does not carry any other traffic except data storage traffic. See Figure 25 (“Typical SAN networks”).

Supported SAN Topologies:

**Point-to-Point**

- The simplest topology for very limited connectivity needs.
- Guarantees “in order” delivery and full bandwidth access.
- Can handle any multipath connectivity to a set of disks, since there are no other elements in the communication path.

**Arbitrated Loop**

- Designed to scale to a limited number of nodes (up to 127).
- Low cost (no interconnecting switch devices required).
- Arbitration protocol is designed to manage media sharing across nodes. This may be disruptive when a node gets added/removed from the loop and the loop initialization protocol starts.
- An arbitrating hub can be used instead of a distributed protocol.

**Switched Fabric**

- A switching element is added to the nodes to allow interconnections via point-to-point links.
- Extends number of devices (potentially thousands), and greater distances can be achieved.
- A scalable, robust and reliable architecture, but the cost of the interconnection devices can add up.

---

**Figure 25: Typical SAN networks**
Storage Area Networks (IP & Ethernet-Enabled)

SANs deliver a substantial number of advantages to the data center. However, the biggest disadvantages have been the need for additional Fibre Channel interface cards, cables, additional devices to manage, and distance limitations. IP SAN technologies are now ported onto the IP layer. However, the SAN must have a reliable transport to enable the communication of SCSI commands to and from SCSI controllers and peripherals. There are two IP-enabled SAN standards in use today.

Both Internet SCSI (iSCSI) and IP-enabled Fibre Channel (FCIP) provide the following benefits in common with the legacy Fibre Channel standard:

- Enable storage consolidation.
- Data sharing.
- Non-disruptive scalability for growth.
- Provide SAN extension over MAN/WAN networks.
- Improved backup and recovery.
- Good performance (relative to the network structure).
- Data integrity.
- Disaster tolerance.
- Ease of data migration.
- Lower cost of ownership.

A substantial number of IT managers deployed iSCSI and FCIP with early acceptance of the performance without changing any of their network structure. As the adoption of the technology into their data centers became more prevalent, additional IP SAN systems were deployed and network performance dropped. Network managers have had to adapt their network designs to compensate for the convergence of storage traffic onto their networks, just as they had to do with voice traffic.

The issue now becomes the network that storage traffic rides on and the transport protocol. A data center or enterprise IT architect has to identify whether they have enough bandwidth, switch packet forwarding capacity, and QoS configured to support acceptable storage performance.

Network Architects have to identify and plan for the following in support of IP SANs:

- The need for true non-blocking ethernet switches.
- Higher performance routers.
- Oversubscription ratios on routers and switches (WAN & LAN).
- Physical or logical ethernet separation (will this affect the ROI?).
- QoS requirements (now a true need for classes of traffic).
- VLAN separation.
- Flow control.
- Redundant switches and paths (impact of STP?).
- Jumbo frames need to be supported end-to-end.

Storage traffic on the LAN has provided the following:

Benefits
- Ubiquitous technology.
- Operates on a known platforms (IP and Ethernet).
- Standards-based solutions.
- Commodity pricing driven by large vendor community and customer demand.
- Operational efficiency.
- Ethernet is universally deployed.
- Technical staff that know how to manage IP and ethernet.
- Reliable transport.
- Wide Area Networking.
- Removal of distance limitations.
- Flexible deployment designs.
- Aids in disaster recovery initiatives and data replication.
- Strong long-term viability as ethernet speeds increase to 10G, 40G, and 100G.

**Negatives**

- TCP slow start impacts I/O latency and throughput.
- The sliding windows algorithm can cause substantial re-transmissions during heavy congestion periods.
- Data re-transmissions can affect applications dependent on timely execution of data requests.
- Difficult, expensive, and inefficient integration with other types of IP SANs and legacy Fibre Channel systems.

**iSCSI**

One of the reasons that the Fibre Channel protocol became prevalent is that it is a very lightweight, high-performance protocol that maps locally to the SCSI initiators and targets (Figure 37, far left). But Fibre Channel protocols are not routable and can be used only within a relatively limited area, such as within a single data center.

The iSCSI protocol sought to improve that limitation by moving the SCSI packets along a typical ethernet network using TCP/IP. The iSCSI protocol serves the same purpose as the Fibre Channel in building SANs, but iSCSI runs over the existing ethernet infrastructure, and thus avoids the cost, additional complexity, and compatibility issues associated with Fibre Channel SANs.

An iSCSI SAN is typically comprised of:

- Software or hardware initiators on the host server connected to an ethernet network.
- Storage resources (targets).
- The iSCSI stack at both ends of the path is used to encapsulate SCSI block commands into ethernet packets for transmission over IP networks (Figure 26). Initiators include both software- and hardware-based initiators incorporated on host bus adapters (HBAs) and NICs.
It is possible to create a lossless ethernet network using older 802.3x mechanisms. However, if the network is carrying multiple traffic classes, which many data centers do, the existing mechanisms can cause QoS issues, which limit the ability to scale a network and impact performance.

The newer generations of iSCSI technology solve these issues:

- High-performance adapters are now available which fully offload the protocol management to a hardware-based iSCSI HBA or NIC. This is called full iSCSI offload.
- Using 10Gb networks and 10Gb NICs with iSCSI SANs generates performance comparable to a Fibre Channel SAN operating at 8Gb.
- New centralized iSCSI boot management tools provide mechanisms for greater scalability when deploying large numbers of servers.

With 10Gb-based iSCSI products, iSCSI becomes a more viable solution for converged networks for small-medium businesses as well as enterprise data centers. For customers with iSCSI-based storage targets and initiators, iSCSI can be incorporated today with their existing ethernet infrastructure. Unlike FCoE, iSCSI does not require the new Data Center Bridging (DCB) infrastructure. But if present, iSCSI will benefit from the QoS and bandwidth management offered by DCB. Because iSCSI administration requires the same skill set as a TCP/IP network, using an iSCSI network minimizes the SAN administration skills required, making iSCSI a good choice for greenfield deployments or when there are limited IT teams and budgets. iSCSI can be enhanced even more with HP’s IRF technology, which is covered elsewhere in this document. iSCSI will continue to be an integral part in many converged network infrastructure designs with the enhancements that HP brings to ethernet switching technology.
Fibre Channel over Ethernet (FCoE)

FCoE takes advantage of 10Gb Ethernet’s performance while maintaining compatibility with existing Fibre Channel protocols. Typical legacy ethernet networks allow ethernet frames to be dropped, typically under congestion situations, and then rely on upper layer protocols such as TCP to provide end-to-end data recovery. It is possible to create a lossless ethernet network using existing 802.3x mechanisms. However, if the network is carrying multiple traffic classes, the existing mechanisms can cause QoS issues, limit the ability to scale a network, and impact performance.

Because FCoE is a lightweight encapsulation protocol and lacks the reliable data transport of the TCP layer, it must operate on Data Center Bridging (DCB) enhanced ethernet to eliminate frame loss under congestion conditions.

Because FCoE was designed with minimal changes to the Fibre Channel protocol, FCoE is an L2 (non-routable) protocol just like Fibre Channel, and can only be used for short-haul communication within a data center. FCoE encapsulates Fibre Channel frames inside of ethernet frames (Figure 27).

Figure 27: FCoE packet

The traditional data center model uses multiple HBAs and NICs in each server to communicate with various networks. In a converged network, the converged network adapters (CNAs) can be deployed in servers to handle both FC and DCB traffic, replacing a significant amount of the NIC, HBA, and cable infrastructure (Figure 28).

FCoE uses a gateway device (an ethernet switch with DCB, legacy ethernet, and legacy FC ports) to pass the encapsulated Fibre Channel frames between the server’s converged network adapter and the Fibre Channel-attached storage targets.

Advantages to FCoE

- FCoE uses existing OS device drivers.
- FCoE uses the existing Fibre Channel security and management model with minor extensions for the FCoE gateway and ethernet attributes used by FCoE.
- Storage targets provisioned and managed on a native FC SAN can be accessed transparently through an FCoE gateway.

Challenges with FCoE

- Must be deployed using a DCB network.
• Requires converged network adapters and new ethernet hardware between the servers and storage targets (to accommodate DCB).
• Is a non-routable protocol and can only be used within the data center.
• Requires an FCoE gateway device to connect the DCB network to the legacy Fibre Channel SANs and storage.
• Requires validation of a new fabric infrastructure that includes both the ethernet and Fibre Channel.

Figure 28: Host bus adapters compared to converged network adapters (CNAs)

Data Center Bridging (DCB) Ethernet

Data Center Bridging Ethernet is a suite of proposed standards from the Data Center Bridging (DCB) task group within the IEEE 802.1 Working Group. The IEEE is defining the DCB standards so that they can apply to any IEEE 802 MAC layer network type, not just ethernet. Data center architects can consider the term “Converged Enhanced Ethernet” (CEE) as being the application of the DCB draft standards to the ethernet protocol. In many cases DCB and CEE are used interchangeably; however, recent trends are to utilize the term DCB, which we will use throughout the remainder of this paper.

There are four new technologies defined in the DCB draft standards:

1. **Priority-based Flow Control (PFC), 802.1Qbb** – An Ethernet flow control that discriminates between traffic types (such as LAN and SAN traffic) and allows the network to selectively pause different traffic classes.

2. **Enhanced Transmission Selection (ETS), 802.1Qaz** – Formalizes the scheduling behavior of multiple traffic classes, including strict priority and minimum guaranteed bandwidth capabilities.

3. **Quantized Congestion Notification (QCN), 802.1Qau** – Supports end-to-end flow control in a switched LAN infrastructure and helps eliminate sustained, heavy congestion in an ethernet fabric. QCN must be implemented in all components in the DCB data path (i.e., converged network adapters and DCB switches), before the network can use QCN. QCN provides granular control.
over flow rate limiting, and unless implemented, a true control of a converged network cannot be guaranteed. In addition, it must be used in conjunction with PFC to completely avoid dropping packets and guarantee a lossless environment.

4. **Data Center Bridging Exchange Protocol (DCBX), 802.1Qaz** – Supports discovery and configuration of network devices that support the technologies described above (PFC, ETS, and QCN).

While two out of the three standards that define the DCB protocols are still nearing completion, the QCN standard has been ratified for quite some time, yet it is the most complex and least understood, and requires the most hardware support. The QCN standard will be critical in allowing data center architects the ability to converge networks “deeper” into the distribution and core network layers. Also, it is important to note that QCN will require a hardware upgrade of almost all existing DCB components before it can be implemented in data centers.

HP advises that the transition to FCoE can be a graceful implementation with little disruption to existing network infrastructures if it is deployed first at the server-to-network edge and migrated further into the network over time. The lack of finalized DCB standards is another reason HP views the server-to-network edge as the best opportunity for taking advantage of “converged network infrastructure” benefits.

Data center architects could also start by implementing FCoE only with those servers requiring access to FC SAN targets. Not all servers need access to FC SANs. In general, more of a data center’s assets use only LAN-attach than use both LAN and SAN. CNAs should be used only with the servers that actually benefit from it, rather than changing the entire infrastructure to FCoE, which can be costly and time consuming. If administrators transition the server-to-network edge first to accommodate FCoE/DCB, this will maintain the existing architecture structure and management roles while reducing future capital outlay for edge connectivity and cabling. This will also allow administrators to maximize their existing investment in their core SAN and LAN architectures. Updating the server-to-network edge offers the greatest benefit and simplification without disrupting the data center’s architectural paradigm.

**The Future of Storage Area Networks**

New technologies and standardization have made it possible to converge Fibre Channel traffic and network traffic at the server edge. It will take time and a multi-step process to converge further into the core of data center networks, as shown in Figure 29. This figure unfortunately reflects the current early stage of deployment for new storage area network technologies. This highlights the need for planning for the future, to enhance the networks supporting the new demands. In fact, iSCSI is greatly improving, especially with the increase in network connectivity speeds at 10Gb Ethernet. However, most storage designs in data centers are based on Fibre Channel technology, and, as a result, FCoE and DCB are critical for converging large scale data centers.

The complexity of guaranteeing low latency/high performance delivery for each type of data traffic, and specifically storage traffic, continues to be the critical component of data center network design. For example, applications using NAS can require network performance that is measured in real-time.
The Server to Network Edge

By implementing convergence first at the server edge, data center architects will be able to take advantage of the standards that are in place, while avoiding the risk of implementing technology that will have to be replaced or modified as standards solidify.

The server-to-network edge is also becoming increasingly complex due to the sprawl of virtual machines, which is covered in greater detail in another section of this document. However, the management of virtual machines and their demand for I/O resources drive the need for DCB.

The HP approach to technologies at the server-to-network edge is based on using industry standards. This ensures that new technologies will work within existing customer environments and organizational roles, yet will preserve customer choice. The HP goal for customers is to enable a simple migration to advanced technologies at the server-to-network edge without requiring an entire overhaul strategy for the data center.

The server-to-network edge refers to the connection points between servers and the first layer of both local area network (LAN) and storage area network (SAN) switches. The most common networks used in enterprise data centers are ethernet for LANs and Fibre Channel for SANs.

Different topologies are one of the reasons that HP is focusing on the server-to-network edge. Administrators should be allowed to maintain similar processes, procedures, data center structures, and organizational governance roles while reaping the benefits of reduced infrastructure at the server-to-network edge.
The server edge is the most physically complex part of data center networks; the server edge has the most connections, switches, and ports (especially in environments using only rack-based servers). For most enterprise data centers using Fibre Channel SANs and ethernet LANs, each network type has its own requirements:

- Unique network adapters for servers.
- Unique switches.
- Network management systems designed for each network.
- Different organizations to manage the networks.
- Unique security requirements.
- Different topologies.
- Different performance requirements.

Each of these differences add complexity and cost to the data center, especially for enterprise data centers that have large installed Fibre Channel and ethernet networks.

Some data center architects use non-Fibre Channel storage. In this case iSCSI solves these challenges without requiring a new ethernet infrastructure. However, iSCSI technology is still evolving and has experienced a slow rate of adoption, especially in enterprise data centers. The question that will be posed later in this section is whether to enhance the converged network infrastructure with iSCSI, or wait for the DCB standards to become ratified and solidified into ethernet switches. Either way, the demand for bandwidth in data centers and their encompassing racks continues to grow with the widespread adoption of server virtualization.

**Virtual Connect – the Best of Both Worlds**

HP Virtual Connect (VC) technology is a hardware-based solution that enables users to partition a 10 Gigabit Ethernet (10GbE) connection and control the bandwidth of each partition in increments of 100Mb.

Administrators can configure a single BladeSystem 10Gb network port to represent multiple physical network interface controllers (NICs), also called FlexNICs, with a total bandwidth of 10Gbps. These FlexNICs appear to the operating system (OS) as discrete NICs, each with its own driver. While the FlexNICs share the same physical port, traffic flow for each one is isolated with its own MAC address and virtual local area network (VLAN) tags between the FlexNIC and VC interconnect module. Using the VC interface, an administrator can set and control the transmit bandwidth available to each FlexNIC.

Data center administrators can use HP Virtual Connect technology to aggregate multiple low bandwidth connections into a single high bandwidth connection. For example, Virtual Connect allows partitioning of a single 10Gb Ethernet connection into up to four lower bandwidth connections required on the blade.
Virtual Connect management tools allow the administrator to assign unique MAC addresses to each of the FlexNICs. All network and storage connections can be defined at deployment and do not need to change if the servers are changed. The MAC address is assigned to a particular bay enclosure in the blade chassis. Once the server blade powers on, the connection profile for that bay is transferred to the server. If a physical server is changed, the MAC assignment for the bay remains unchanged, and a replacement server blade assumes the same connection characteristics as the original one.

HP is also building storage network protocols into its Virtual Connect technology that will combine the existing Flex-10 capabilities with DCB and FCoE technologies. This will allow customers to use a single Virtual Connect server connection to access both storage and server networks. HP FlexFabric uses industry standards to converge storage, server, and network resources. It provides an orchestration and management layer to provide virtualization-enabled networking and rapid provisioning. FlexFabric offers a wire-once data center that can respond to application and workload mobility. Network connections can move with workloads across or even between data centers.

HP Virtual Connect technology offers significant advantages over other 10Gb devices that provide large bandwidth but do not provide segmentation. The ability to adjust transmit bandwidth by partitioning data flow makes 10GbE more cost effective and easier to manage. It is easier to aggregate multiple 1Gb data flows and fully utilize the 10Gb bandwidth. The fact that Virtual Connect is hardware-based means that multiple FlexNICs are added without the additional processor overhead or latency associated with server virtualization (virtual machines). Significant infrastructure savings can also be realized since additional server network mezzanine cards and associated interconnect modules may not be required.

There are several advanced features available when using the Virtual Connect FlexFabric module and adapter technology when enabling FCoE. One of these advanced features is support for Flex-FC. This feature allows for higher-level software tools such as Insight Dynamics and Orchestration to assign multiple World Wide Names (WWNs) to an individual FCoE PF (up to 16) and map all those WWNs to the VC SAN Fabric.

In addition, it allows up to 8 SAN boot targets to each FCoE PF.
Also, iSCSI boot and offload support on the FlexFabric adapters enables TCP/IP and iSCSI processing offload as well as iSCSI boot functions of these adapters when connecting to either Flex-10 or FlexFabric modules. One advanced iSCSI feature is support for up to 128 iSCSI boot targets per iSCSI function. Volume Control Manager (VCM) provides centralized boot configuration management of iSCSI boot parameters, and provides primary and secondary boot path management. Although it requires a separate iSCSI driver, it does not require an NIC driver and will drastically reduce CPU utilization by offloading processing to the adapter.

**FlexFabric helps modify the tasks of provisioning SANs**

**The old way:**

- Connect servers individually to LAN and SAN.
- Require LAN and SAN coordination to add, move, or replace a server.
- Wait for everyone to finish their pieces.

**The new way:**

- Pool server connections to LAN and SAN.
- System Administrator can add, move, or replace servers transparent to LAN and SAN.

**Other Topics to Consider**

Switches with a non-blocking architecture should be used. In most commonly used switches, the backplane capacity is less than the sum of the capacity of all the ports.

- A dedicated high-speed network with either a physical or logical separation is recommended. Flow control in the ethernet protocol is very minimal, and several standards are emerging to tackle the complexities of storage traffic.
• Redundancy is also a critical factor in designing a network for storage traffic. There are several pieces to keep in mind when designing a network with redundant switches:
  o Multiple paths add fault-tolerance and performance since more than one physical path exists between the computer system and storage devices through the buses, controllers, switches, and bridge devices connecting them.
  o An IRF-based network extends the control plane across multiple active switches, enabling interconnected switches to be managed as a single common fabric with one IP address; this increases network resilience, performance, and availability while simultaneously reducing operational complexity.
• The support for jumbo frames is the biggest imposition of iSCSI on ethernet switches due to the characteristic block size of storage volumes, which at 2KB is larger than the nominal ethernet frame size of 1.5KB.

Section Summary
FCoE, DCB, and current-generation iSCSI are standards to achieve the long-unrealized potential of a single unified fabric. With respect to FCoE in particular, there are important benefits in beginning to implement a converged fabric at the edge of the server network. By implementing converged fabrics at the server-to-network edge first, customers gain benefits from reducing the physical number of cables, adapters, and switches at the edge where there is the most traffic congestion and the most associated network equipment requirements and costs. In the future, HP plans to broaden Virtual Connect Flex-10 technology to provide solutions for converging different network protocols with its HP FlexFabric technology. HP plans to deliver the FlexFabric vision by converging the technology, management tools, and partner ecosystems of the HP Networking and Virtual Connect network portfolios into a virtualized fabric for the data center.

HP Data Center Reference Architecture (DCRA)

Specific Data Center Reference Architecture models are described below. The core building blocks and concepts have previously been reviewed. This section introduces HP’s reference architectures for the virtualized and converged network infrastructure. The Data Center Reference Architecture is less of a focus on a specific piece of HP equipment than an overall discussion of the use of current HP technologies to fulfill the vision of virtualization.

All HP Data Center Reference Architecture examples fulfill the following vision:

• **Best of Breed Servers** – Meeting the demands of virtualized and non-virtualized data center deployments. For example, HP’s c-Class BladeSystem meets the demands of the converged network infrastructure through enhancements in processing density, Virtual Connect, reduced cabling, and enhanced ability to function in flatter L2 network designs to meet the demands of server virtualization.

• **Best-of-breed technology partnerships** – No networking vendor has a stronger product and interoperability alliance with VMware. As your virtualization deployments continue to grow, the alliance between VMware and HP will continue to add performance, flexibility, and simplified management of virtual machines and their attached networks.

• **Best-of-breed Layer 2/3 networking components** – To meet the demands of virtualization and converged I/O, all of which are enhanced by HP’s standards-driven architecture and IRF.

• **Utilization of best-of-breed network security products** – This not only secures the network, but is also designed with the security challenges of the virtualized network in mind. There are options to embed security into HP core switches like TippingPoint’s Intrusion Prevention System. There is also the option to operate them on external and virtual appliances using vController and vIPS which extend network security to the virtualized network. Security is also controlled with discovery and management of VMs using the Security Management System (SMS). Network security can be
complex and involve a great deal of management overhead. HP security products specifically address this problem and simplify network security. For example, TippingPoint products work in unison with IMC, to allow a “single pane of glass” management system to manage multiple security controllers throughout the data center.

- **Best-of-breed management tools** – To simplify and enhance the management of data center resources. Orchestration is highlighted in HP’s network and server hardware in the data center through the use of IMC. The drive towards a “single pane of glass” management structure from server hardware through VMs and up through the network components are now becoming a reality.

- **Utilization of best-of-breed storage and storage networking technologies** – From HP’s StorageWorks division, with enhancements to FCoE, iSCSI, and the future with DCB. All of these are enabled through the use of Virtual Connect FlexFabric, DCB-enabled ToR, EoR, and core switches for aggregation.

- **Utilization of best-of-breed power and cooling** – HP Data Center Smart Grid leads the industry in lower power consumption and device monitoring.

### Demystifying Data Center Reference Architecture

When looking at all the different variables that have to be implemented in a data center design, the easiest way to approach the explanation of each model is by using a layered approach. This way, HP technologies and products can be applied to represent how a specific architecture model would be deployed. The diagram below shows how each layer of the architecture supports the ultimate goal of virtualization.

**Figure 32: DCRA example methodology**

#### HP Data Center Reference Architecture Example Methodology

- Designed for Virtualization
- LAN & WAN Safe
- Hardened Security
- Simplifies Networking & Management
- Resilient and Available
- Ecocentric
- Backed by HP and Partners

**Layer 1: Virtualization**

**Layer 2: Security**

**Layer 3: Orchestration**

**Layer 4: Converged Network Infrastructure**

**Layer 5: Economic Data Center**

**Layer 6: Support, Services, and Partners**
Evolving Network Designs

Access Layer Switching

In current data center deployments, there are two access layer switching deployment models: Top of Rack and End of Row.

Top of Rack (ToR)

As the name implies, in a ToR placement, servers within a rack connect to an access layer switch generally placed at the top of the rack, as in Figure 33 ("Top of Rack placement").

Figure 33: Top of Rack placement

In a ToR design, servers connect to an access switch via copper or fibre within the rack, while the access switch connects to other consolidation or backbone switches within the data center via Multi-mode fibre. Typically, the access switches are managed by the L2 or L2/3 switches. Often, the data center backbone consists of high-capacity aggregation or distribution switches with L2/3 capabilities. Ethernet copper is generally restricted to relatively short runs inside the rack, while connections outside the rack can be made with smaller form factor multi-mode fibre. This reduces the overhead weight, as well as having the ability to be connected to different capacity interfaces to support the bandwidth requirements of the rack.

Shorter copper ethernet runs in the rack allow for multiple choices on cable types (TwinAx, CAT5, CAT5e, and/or CAT6) which can support various required speeds dictated by the systems in the rack. In many cases, server to switch connections can be up to 10G connections with support for integrated I/O.
Advantages

Issue isolation – Each rack, or group of racks using IRF, with a ToR configuration is treated as an independent module. Any issues or outages that occur with an access layer switch typically affect only the servers within that rack which are connected to the access switch.

Traffic isolation – Since each access switch is a home run back to a backbone aggregation switch, traffic can be monitored and isolated to an identified switch port within a specific rack.

Software and firmware upgrades – Although there are more switches to upgrade, a switch within a rack can be identified for the upgrade, and any issues occurring during the upgrade can be solved or backed out. Again, the exposure to the rest of the network is virtually eliminated.

Capital outlay – This design also takes into consideration that every rack is not populated with access switches when the data center is built. A general rule of thumb is to only install equipment in a rack when there is a demand for it. The ToR placement allows the data center manager to populate the rack as required with little planning except for available port capacity on the aggregation switch.

Disadvantages

Number of switches – Each rack adds to the number of switches that need to be managed as independent entities.

Data center aggregation switch port capacity – Data center planners must carefully plan the port capacity on their network backbone. The aggregation switches generally require greater port density, which can affect data center backbone scalability in the future.

End of Row (EoR)

In an EoR placement, a rack containing the switching equipment is typically placed at either end of a row of cabinets or racks. See Figure 34 (“End of Row”). Bundles of copper cabling (Cat 6 or 6A) provide the connectivity from each server rack to the switching equipment rack. Servers are usually connected to a patch panel inside each server rack. The copper bundles are home run to another patch panel in the rack containing the access switches. The switches are then connected to the patch panel via patch cables. EoR switches are typically connected back to the core with a series of fibre patch cables.

The term “End of Row” does not imply that the network racks have to be placed at the end of the row. There are designs where network switch racks are placed together in the middle of cabinet/rack rows. Placing the switch rack in the middle of a row limits the length of copper cables required to connect the furthest server racks to the nearest network rack.

Unlike the ToR model, where each rack is treated as an independent module, in the EoR placement model, each row is treated as an independent module.

Advantages

Managed entities – There are fewer switches to manage as separate entities.

Network pre-planning – A data center can be pre-planned and cabled without deploying a substantial number of switches into individual racks.

Capital outlay – Fewer switches means less capital outlay.

Disadvantages

Physical disasters – A potential physical disaster affecting cabling or the End of Row aggregation switch can have adverse effects on not just one rack of servers, but an entire row.

Software and firmware upgrades – The End of Row switch is connected to all the servers in that row. Issues with the switch or upgrade corruption can possibly affect the entire row of servers.
Copper cabling – The magnitude of copper ethernet home runs back to the EoR switch makes for substantial cabling bundles.

Figure 34: End of Row

Blade Servers in ToR and EoR Placements
Blade servers may contain integrated copper and FC switches. In effect, the ToR switch has been moved inside the blade server chassis. In an EoR placement, the integrated switches will still need to be connected to the rack patch panel for home run back to the EoR switch. Integrated network switches reduce cabling, but add to the number of different network entities that have to be managed in the data center.

HP’s Virtual Connect mitigates this issue by providing a controlled transparent solution within the enclosure that is an extension of the network interface cards of the servers and can be considered LAN/SAN Safe. Since Virtual Connect is not an active entity in the network fabric, it can be deployed without the concern of adding another active device into the network that needs to be managed.

Cabling Trends - Fibre vs. Copper
Early client/server-based data center networks primarily addressed static applications and email. Data centers today support new applications and traffic patterns that are dynamic and demand new approaches for real-time application access, as well as guaranteed performance, low latency, and any-to-any communication patterns. The ever-increasing demand for bandwidth and throughput, as well as the promise of converged network infrastructure, justifies the need for 10G or higher speeds.

10G Ethernet no longer mandates Fibre cabling. Recent IEEE standards enable 10G Ethernet over copper using TwinAx, Cat 6 or 7 cabling. The consideration for fibre vs. copper is on the basis of cost and future expandability. At this point, 10G over copper is still less expensive. However, with 40G and 100G standards being ratified in the near future, data center architects must look closely at the cost benefit of fibre. This is due to the fact that 40G and 100G Ethernet are only supported on multi-mode fibre.
**Converging the Network Infrastructure Today**

Data center administrators can address complexity at the server-to-network edge by consolidating server I/O in various ways:

- Combining multiple lower-bandwidth connections into a single, higher-bandwidth connection.
- Converging (combining) different network types (LAN and SAN) to reduce the amount of physical infrastructure required at the server edge.

Administrators can already use HP Virtual Connect Flex-10 technology to aggregate multiple independent server traffic streams into a single 10Gb Ethernet (10GbE) connection.

For instance, administrators can partition a 10GbE connection and replace multiple lower bandwidth physical NIC ports with a single Flex-10 port.

**Advantages**

- Reduces management requirements.
- Reduces the amount of physical infrastructure (number of NICs and number of interconnect modules).
- Simplifies the amount of cross-connect in the network.
- Reduces power and operational costs.

InfiniBand, iSCSI, and other protocols have been introduced with the promise of providing efficient transport on a single fabric for multiple traffic classes and data streams.

The goals of a converged network are:

- To simplify and flatten the typical ethernet topology.
- To improve quality of service (QoS).
- To reduce host/server adapters.
- To reduce cabling requirements.
- To reduce switch ports and costs.
- To enable simpler, more centralized management.

**Blade Server 1-Tier Design**

This type of network deployment signifies the current pinnacle of network virtualization. Blade servers allow for substantial compute density per rack, row, and data center. HP has optimized the c7000 blade server portfolio to support the vision and reality of virtualization. The network design optimizes the reality of high performance networking with simplicity. It allows flexibility in virtual machine networking and converged I/O options. This approach is at the forefront of network design for virtualization, since it utilizes both the IRF framework and FlexConnect fabrics. It operates well while allowing for seamless management and troubleshooting of virtual machines.

**Objectives**

When determining if this deployment type is right for you, keep in mind the overall objectives:

- **L2 Flexibility** – Since the overall goal is to support virtual machines, a flattened L2 network allows virtual machines to be moved without the need for IP address changes. Keep in mind that a flattened L2 design supports long range vMotion using VPLS.
- **Reduced management complexity** – Flattening the network with large aggregation switches reduces the number of devices that have to be provisioned and monitored when setting up VLANs for virtual machines.
- **Zero need for STP/RSTP** – Coupled with IRF, this network design is loop-free, thus the need for STP is eliminated.
• Frame forwarding and packet forwarding performance are optimized in this flat IRF framework. Virtual machines communicate between each other, with their converged I/O resources served by an optimal L2 network, VLANs, and LACP. The requirement for reliance on TCP flow control is virtually eliminated.

• VLAN management – A flatter L2 network is ideal for virtual machines. It provides support for network extension across the data center and between sister data centers, the framework for virtual machine mobility.

• Centralized security – L2 networks allow the IP-based IPS devices to be integrated into the core switches, where VLANs are aggregated. This means less network security devices to provision and monitor. It allows for less complexity in security policies, since they require provisioning in fewer devices. Spanning of VLANs across a flattened L2 data center network provides greater flexibility in rack mount server connectivity in disparate rack locations without the need for the creation of complex IP-based ACLs.

**Where Is a Blade Server 1-Tier Network Best Suited?**

Anywhere cost, performance, flexibility, and real estate are at a premium!

**Rack and Logical Diagrams**

---

**Figure 35: Blade Server 1-tier Architecture Rack View**

---
Logical Layers of this Design
Virtualization

The network edge physically starts with the FlexConnect fabric, but really starts internally in the VMs and configured virtual switches. HP A12518 switches are able to communicate natively with VMware virtual switches, allowing updates of ARP tables between physical and virtual switches. The virtual machines and virtual switches can provision VLANs, which in turn interoperate with the IRF fabric, allowing seamless VM movement with vMotion and high performance frame forwarding. All of the switches in IRF frameworks support 4096 VLANs, which provides substantial network flexibility to configure a substantial number of virtual machine groups onto their own virtual network segments.

Combining LACP and IRF in this design provides high-speed link aggregation with re-convergence times at sub-50ms in the event of a link failure. It also allows links to be aggregated and utilized for bandwidth from the converged network adapters across all switches to forward traffic.

This design also supports the dynamic storage of World Wide Names. This feature allows a FlexConnect module to be configured with the required IP and VLAN information once. In the event of a device failure, the replaced FlexConnect device will gracefully provision itself with the original configuration. This feature provides advantages when a specific component can have an impact on a substantial number of virtual machines.

This network design supports long-range vMotion connectivity, enabling clustering or synchronizing virtual machines between disparate locations. The IRF fabric is capable of supporting dark fibre connectivity between switches up to 70 kilometers apart. The switches also allow VPLS connectivity between switches in geographically separated data center locations. Be aware that long-range fibre segments using MPLS can add latency that limits the ability for VMs to use converged I/O resources between data center locations. Long distance WAN networks generally address normal server communications and disaster recovery efforts.
Security

Although not depicted in the diagrams, the TippingPoint integrated IPS can be added to the A12518 switches. Since VLANs can be configured to span all the way to the core switches, all network security can be centralized in a fully redundant N+1 configuration by placing an IPS blade in each of the core switches.

Orchestration

This network configuration supports the full feature complement of HP’s Intelligent Management Console from the core switches all the way down the network to the physical servers themselves. This provides a “single pane of glass” management platform to support the servers, VMs, virtual switches, IRF frameworks, IP routing, and security devices.

Convergence Network Infrastructure

In this configuration, both Fibre Channel and iSCSI technologies are supported. The FlexConnect fabric allows the conversion of FCoE traffic to Fibre Channel, right at the FlexFabric level. The FlexConnect fabric and 12518 switches also support iSCSI. The IRF framework provides a high-speed/low loss network with substantial bandwidth provided by the 10G aggregated links from the converged network adapters all the way up to the iSCSI SAN endpoint.

HP Converged Infrastructure provides a complete end-to-end virtual computing solution encompassing racks, servers, storage, networking technologies, security, and management.

Simplified 2-Tier Design (ToR)

This type of network deployment can be deployed in a two-tier, or simplified three-tier design. HP network technologies support both approaches. Optimized three-tier deployments transition to multiple L2 aggregation points, rather than two tiers of L3 termination. This is mainly due to virtualization, but can also be attributed to the fact that the data center has lower density switches already deployed in the network. It can also be attributed to the amount of ethernet port capacity available in the core switches within the data center.

Objectives

When determining if this deployment type is appropriate, keep in mind the overall objectives:

- **Layer 2 flexibility** – Flattening the network with L2 better suits virtual machines residing on rack mount servers which require the ability to communicate with other virtual machines that may be in the same rack or row, within the same data center facility, or data center campus. Since the overall goal is to support virtual machines, a flattened L2 network allows virtual machines to be moved without the need for IP address changes. Keep in mind that a flattened L2 design supports long-range vMotion using VPLS.

- **Reduced management complexity** – Flattening the network with large aggregation switches reduces the number of devices that have to be provisioned and monitored when setting up VLANs for virtual machines.

- **Less dependence on STP/RSTP** – HP recommends switches that support IRF, but many data centers already have switches that may not support IRF. In this case, using IRF in the core will certainly help optimize and simplify the network. In either case, fewer ethernet switches means less link state management between servers, ToR switches, and aggregation switches within the data center network.

- **VLAN management** – A flatter L2 network is ideal for virtual machines. It provides support for network extension across the data center and between sister data centers, the framework for virtual machine mobility.

- **Centralized security** – L2 networks allow the IP-based IPS devices to be integrated into the core switches, where VLANs are aggregated. This means less network security devices to provision and
monitor. It allows for less complexity in security policies, since they require provisioning of fewer devices. Spanning of VLANs across a flattened L2 data center network provides greater flexibility in rack mount server connectivity in disparate rack locations, since it removes the need for complex IP-based ACLs.

Where Is ToR Best Suited?
A top of rack deployment is generally utilized when the majority of servers are rack mounted. Rack mount servers are still used extensively in data centers, even with the wide adoption of blade servers. You will also find ToR switches in racks originally deployed in the traditional 3-tier networks. This was due to the lack of available port capacity in aggregation switches, or limited adoption of fibre modules in rack mount servers due to cost.

Rack and Logical Diagrams

Figure 37: Simplified 2-tier Architecture Rack view
Logical Layers of this Design
Virtualization

The network edge physically starts at the converged network adapters, but really starts internally in
the VM machines and configured virtual switches. The virtual machines and virtual switches can
provision VLANs, which in turn interoperate with the IRF fabric, allowing seamless VM movement with
vMotion and high performance frame forwarding. All the switches in IRF frameworks support 4096
VLANs, which provides flexibility to configure a substantial number of virtual machine groups onto
their own virtual network segments.

The network design does not require the STP/RSTP protocol. But in the event that existing third party
switches are already deployed in the racks, STP protocols are supported through standards
interoperability. MSTP leverages gratuitous ARP with the IRF IDs created for the links from the IRF
fabric in the A12518 switches.

Combining LACP and IRF in this design provides high-speed link aggregation with re-convergence
times at sub 50ms in the event of a link failure. It also allows links to be used aggregated and utilized
for bandwidth from the converged network adapters across all switches to forward traffic.

This network design supports long range vMotion connectivity, enabling clustering or synchronizing
virtual machines between disparate locations. The IRF fabric is capable of supporting dark fibre
connectivity between switches up to 70 kilometers apart. The switches also allow VPLS connectivity
between switches in geographically separated data center locations. Be aware that long-range fibre
segments using MPLS can add latency that limits the ability for VMs to use converged I/O resources
between data center locations. Long distance WAN networks generally address normal server
communications and disaster recovery efforts.
Security

Although not depicted in the diagrams, the TippingPoint integrated IPS can be added into various layers in the network. Since VLANs can be configured to span all the way to the core switches, all network security could even be centralized in a fully redundant N+1 configuration by placing an IPS blade in each of the core switches.

Orchestration

This network configuration supports the full features of HP’s Intelligent Management Console from the core switches to the physical servers. This provides a “single pane of glass” management platform to support the servers, VMs, vSwitches, IRF frameworks, IP routing, and security devices.

Convergence Network Infrastructure

In this configuration, both FCoE/Fibre Channel and iSCSI technologies are supported. The A5820 receives the FCoE traffic from the configured converged network adapters. It converts the FCoE traffic over to Fibre Channel at the switch, which in turn can be connected to other Fibre Channel directors or Fibre Channel SANs. In support of iSCSI, the IRF framework provides a high-speed/low loss network with substantial bandwidth provided by the 10G aggregated links from the converged network adapters to the core A12S18 switches.

Support for the Eco-centric Data Center

HP servers and switches lead the industry in limiting power consumption without compromise on performance. Every component in this design from the rack up supports a “sea of sensors”, which can be monitored using HP’s System Insight Manager (SIM).

Optimized 3-Tier Design

This type of network deployment can be deployed as a greenfield, or it fits well into data center networks where added bandwidth, 10G port capacity, and simplified management are paramount. It also ensures the interoperability of legacy deployed EoR and ToR. Although the depicted design focuses on HP switches, with IRF third party switches can be inserted at any level and will interoperate using standards-based networking.

HP has optimized its network technologies to support this type of deployment design. Optimized 2- and 3-tier deployments transition to a flatter layer aggregation design, rather than multiple tiers of L2 and three aggregation points. This is mainly due to virtualization, but can also be attributed to the fact that the data center has lower density switches already deployed in the network. It can also be attributed to the amount of ethernet port capacity available in the core switches within the data center.

Objectives

When determining if this deployment type is appropriate, keep in mind the overall objectives:

- L2 flexibility – Since the overall goal is to support virtual machines, a flattened L2 network allows virtual machines to be moved without the need for IP address changes. This design also provides the flexibility to leverage L3 separately where needed to mitigate broadcast domains and control traffic between segments. Also, a flattened L2 design supports long range vMotion using VPLS.

- Optimized management complexity – Leveraging IRF in the various layers of the network allow what was typically a complex management scheme to be implemented in a 3-tier framework. IRF allows each layer to be managed as a single entity. This provides a clear depiction of the devices and traffic in the data center.

- Less dependence on STP/RSTP – Ideally, all switches should support IRF. However, existing switches may not support IRF. In this case, using IRF in the core will certainly help optimize and simplify the network. In either case, less ethernet switches means less link state management between servers, ToR switches, and aggregation switches within the data center network.
- VLAN Management – A flatter L2 network is ideal for virtual machines. It provides support for network extension across the data center and between sister data centers, the framework for virtual machine mobility.
- De-centralized security – The 3-tier model allows security to be distributed throughout the network allowing for a granular view of the individual segments within the data center. With centralized management, one can deploy many smaller IPS systems in the distribution layers and configure different security profiles depending upon the application requirements.

Where Is EoR Best Suited?
EoR deployment is generally being phased out of modern data centers with the advent of higher port capacity switches. They do exist, generally due to the fact that many data centers were built with pre-existing legacy L3 and L2 port capacity models. With the adoption of virtual machine and converged networking, they are being phased out, since they require substantial amounts of hardware with nominal performance gain compared with 2- and 3-tier network deployments. The fact that HP A12518 switches can each support 864 Gigabit and 512 10G Ethernet ports, as well as soon being able to interconnect four of these switches into an IRF framework, aggregates substantial port capacity, security, and management functionality. This advantage alone allows the network to be flattened with fewer tiers. This design allows legacy third party switch deployments to be integrated into the IRF framework, adding substantial port capacity, security, and management functionality.

Rack and Logical Diagrams

Figure 39: Optimized 3-tier Architecture Rack View

Optimized 3-tier Physical View

![Optimized 3-tier Physical View Diagram]
Logical Layers of this Design

Virtualization

The network edge physically starts with the converged network adapters, but really starts internally in the VM machines and configured virtual switches. The virtual machines and virtual switches can provision VLANs, which in turn interoperate with the IRF fabric, allowing seamless VM movement with vMotion and high performance frame forwarding. All the switches in an IRF framework support 4096 VLANs, which provides substantial network flexibility to configure a substantial number of virtual machine groups onto their own virtual network segments.

In comparison with a 3-tier network deployment, the A9505 EoR switches add additional port aggregation and cabling route flexibility. The A9505 switches are also IRF-capable and can be connected into their own IRF framework. The EoR switches function well when the rows within a data center are extremely long, and the data center architects want to transition away from middle of row switch deployments. This approach aggregates port connectivity into high capacity switches and removes multiple points of management.

The network design does not require the STP/RSTP protocol. However, in the event that existing third party switches are already deployed in the racks, STP-family protocols are supported. MSTP leverages gratuitous ARP with the IRF IDs, which are created for the links from the IRF fabric in the A12518 switches.

Combining LACP and IRF in this design provides high-speed link aggregation with re-convergence times at sub-50ms in the event of a link failure. It also allows links to be aggregated and utilized for bandwidth from the converged network adapters across all switches to forward traffic.

This network design supports long-range vMotion connectivity, enabling clustering or synchronizing of virtual machines between disparate locations. The IRF fabric is capable of supporting dark fibre.
connectivity between switches up to 70 kilometers apart. The switches also allow VPLS connectivity between switches in geographically separated data center locations. Be aware that long range segments using MPLS can add latency that limits the ability for VMs to use converged I/O resources between data center locations. Long distance WAN networks generally address normal server communications and disaster recovery efforts.

Security

Although not depicted in the diagrams, the TippingPoint integrated IPS can be added into the various layers in the network. Since VLANs can be configured to span all the way to the core switches, all network security could even be centralized in a fully redundant N+1 configuration by placing an IPS blade in each of the core switches.

Orchestration

This network configuration supports the full feature complement of HP’s Intelligent Management Console, from the core switches to the physical servers. This provides a single pane of glass management platform to support the servers, VMs, vSwitches, IRF frameworks, IP routing, and security devices.

Support for the Eco-centric Data Center

HP servers and switches lead the industry in limiting power consumption without compromise on performance. Every component in this design from the rack up supports the HP “sea of sensors”, which can be monitored using IMC.

HP Converged Infrastructure provides a complete end-to-end virtual computing solution, encompassing racks, servers, storage, networking technologies, security, and management.

Summary

The fundamental nature of data center computing is rapidly changing. Today’s data center networks must evolve to support on-demand, virtualized IT environments. HP delivers the foundation for the data center of the future, today, by providing a unified, virtualization-optimized infrastructure. HP Networking solutions enable the following:

- Breakthrough cost reductions by converging and consolidating server, storage, and network connectivity onto a common fabric with a flatter topology and fewer switches.
- Predictable performance and low latency for bandwidth-intensive server-to-server communications.
- Improved business agility, faster time-to-service, and higher resource utilization by dynamically scaling capacity and provisioning connections to meet virtualized application demands.
- Removal of costly, time-consuming, and error-prone change management processes.
- Modular, scalable, industry standards-based platforms and multi-site, multi-vendor management tools to connect and manage thousands of physical and virtual resources from a “single pane of glass”.

To learn more about how HP Networking can help your business enjoy the benefits of a converged infrastructure, please contact your HP account manager or reseller.
HP Networking Services

HP offers a comprehensive set of network services to help design, deploy, integrate, manage, and support your next-generation connectivity and communications environment. HP Services for Data Center Transformation can help you simplify, optimize and integrate your existing facilities and implement a next-generation data center that is efficient, virtualized, and highly available. Contact your HP account manager or reseller to find out how HP Networking Services can help you implement an automated, high-performance data center network.

Support, Services, and Partners

The core foundation to the solution (HP and its technology partners):

- HP is #14 on the list of Fortune 500 companies.
- HP has a global reach, so your data center solution will be supported internationally.
- HP has industry-leading support and professional services to support your network.
- HP products can be integrated with other vendor solutions through the use of standards-based technologies.
- No other vendor can provide a complete end-to-end virtual computing solution, encompassing racks, servers, storage, networking technologies, security, and management.
Glossary of Acronyms

**DCB:** Data Center Bridging (DCB) is a series of enhancements to the IEEE 802.1 standard to provide extensions to ethernet for support for converged technologies such as FibreChannel over Ethernet (FCoE).

**CEE:** Converged Enhanced Ethernet (CEE) is an enhanced ethernet that enables the convergence of various applications in data centers (LAN, SAN, and HPC) onto a single interconnect technology.

**CIFS:** The Common Internet File System (Microsoft CIFS), an enhanced version of Microsoft Server Message Block (SMB), is the standard way that computer users share files across intranets and the Internet.

**CoS:** Class of Service (COS) is one type of the techniques or methods used to deliver Quality of Service (QoS) in a network.

**DoS:** A Denial of Service (DoS) attack is an occasion in which a legitimate user or a group of users is prevented from accessing the services and information of network resources they would normally receive.

**EoR:** End of Row topologies, which rely on larger switches placed on the end of each row for server connectivity.

**ETS:** Enhanced Transmission Selection (ETS) is to allocate bandwidth based on the different priority settings of the converged traffic.

**EVBS:** Edge Virtual Bridging (EVB) is an IEEE standard that involves the interaction between virtual switching environments in a hypervisor and the first layer of the physical switching infrastructure.

**FCAPS:** An extension of the popular network management conceptual frameworks called Telecommunication Management Network (TMN), FCAPS describes network management in 4 layers. Each TMN layer needs to perform some or all FCAPS functions in certain ways.

**FCIP:** Fibre Channel Over TCP/IP (FCIP) describes mechanisms that allow the interconnection of islands of Fibre Channel storage area networks over IP-based networks to form a unified storage area network in a single Fibre Channel fabric.

**FCoE:** Fibre Channel over Ethernet (FCoE) is an encapsulation of Fibre Channel frames over ethernet networks. This allows Fibre Channel to use 10 Gigabit ethernet networks (or higher speeds) while preserving the Fibre Channel protocol.

**FC:** Fibre Channel, a gigabit-speed network technology primarily used for storage networking.

**HP IMC:** HP Intelligent Management Center (IMC) delivers next-generation, integrated and modular network management capabilities that efficiently meet the end-to-end management needs of advanced, heterogeneous enterprise networks.

**IDS:** An intrusion detection system is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a management station.

**IRF:** Intelligent Resilient Framework (IRF) is a software virtualization technology developed by H3C (3COM). Its core idea is to connect multiple devices through physical IRF ports and perform necessary configurations, and then these devices are virtualized into a distributed device.

**iSCSI:** The Internet Small Computer System Interface (iSCSI) is a TCP/IP-based protocol for establishing and managing connections between IP-based storage devices, hosts, and clients, called the Storage Area Network (SAN).

**Jumbo Frames:** Jumbo frames often mean 9,216 bytes for Gigabit ethernet, but can refer to anything over 1,500 bytes.
**MSTP:** The Multiple Spanning Tree (MST) protocol carries the concept of the IEEE 802.1w Rapid Spanning Tree Protocol (RSTP) a leap forward by allowing the user to group and associate VLANs to multiple spanning tree instances (forwarding paths) over Link Aggregation Groups (LAGs).

**NAT-PT:** Network Address Translation with Port Translation (NAT-PT) is a service which can be used to translate data sent between IP-heterogeneous nodes.

**NIC:** Network Interface Cards (NIC) are adaptors attached to a computer (or other network device such as a printer) to provide the connection between the computer and the network.

**NMS:** The Network Management System (NMS) is a combination of hardware and software used to monitor and administer a network.

**PFC:** Priority Flow Control (PFC) is defined by a one-byte bitmap. Each bit position stands for a user priority. If a bit is set, the flow control is enabled in both directions (Rx and Tx).

**Port mirroring:** Port mirroring is used on a network switch to send a copy of network packets seen on one switch port (or an entire VLAN) to a network monitoring connection on another switch port.

**PXE:** Pre-Execution Environment (PXE) is an environment to boot computers using a network interface independently of data storage devices (like hard disks) or installed operating systems.

**QCN:** Quantized Congestion Notification (QCN) is a form of end-to-end congestion management defined in IEEE 802.1.Qau. The purpose of end-to-end congestion management is to ensure that congestion is controlled from the sending device to the receiving device in a dynamic fashion that can deal with changing bottlenecks.

**RAID:** Redundant Array of Inexpensive Disks (RAID) is a technology that provides increased storage functions and reliability through redundancy. Combining multiple disk drive components into a logical unit, where data is distributed across the drives in one of several ways called "RAID levels".

**RSTP:** The Rapid Spanning Tree Protocol (RSTP IEEE 802.1w) can be seen as an evolution of the 802.1d standard more than as a revolution. 802.1w is also capable of reverting back to 802.1d in order to interoperate with legacy bridges (thus dropping the benefits it introduces) on a per-port basis.

**SAN:** A Storage Area Network (SAN) is a high-speed special-purpose network (or subnetwork) that interconnects different kinds of data storage devices with associated data servers on behalf of a larger network of users.

**SCSI:** The Small Computer System Interface (SCSI), an ANSI standard, is a parallel interface standard used by Apple Macintosh computers, PCs, and many UNIX systems for attaching peripheral devices to computers.

**SVF:** A Secure Virtualization Framework is an HP TippingPoint technology designed specifically for implementing best-of-breed threat protection for the virtualized infrastructure.

**SIIT:** Stateless IP/ICMP Translation (SIIT) Algorithm mechanisms translate IPv4 datagram headers into IPv6 datagram headers or vice versa.

**SMB:** The Server Message Block (SMB) protocol is an IBM protocol for sharing files, printers, serial ports, etc. between computers.

**SNMP:** The Simple Network Management Protocol (SNMP) is used by network management systems to communicate with network elements.

**SPB:** Shortest Path Bridging (SPB) provides logical ethernet networks on native Ethernet infrastructures using a link state protocol to advertise both topology and logical network membership.

**STP:** The Spanning Tree Protocol (STP) is an L2 protocol designed to run on bridges and switches. The main purpose of the spanning tree is to prevent loops from forming in a bridged network.

**TCP windows sizing:** With window sizing, TCP dynamically determines the number of frames to send at once without an acknowledgement.
**TOE:** TCP offload engine (TOE) technology aims to take the server CPU out of I/O processing by shifting TCP/IP processing tasks to the network adapter or storage device. This leaves the CPU free to run its applications, so users get their data faster.

**ToR:** Top of Rack utilizes a switch at the top of each rack (or close to it).

**VEPA:** Is a standard being led by HP for providing consistent network control and monitoring for virtual machines (of any type).

**VIP or VIPA:** A virtual IP address is an IP address not connected to a specific computer or network interface card (NIC) on a computer. Incoming packets are sent to the VIP address, but they are redirected to physical network interfaces.

**VM:** A virtual machine is a system that enables multiple operating systems to concurrently run on a single physical server, providing much more effective utilization of the underlying hardware.

**VLANs:** Virtual LANs (VLANs) provide the capability to overlay the physical network with multiple virtual networks. VLANs allow you to isolate network traffic between virtual networks and reduce the size of administrative and broadcast domains.

**WWN:** A World Wide Name or World Wide Identifier (WWID) is a unique identifier which identifies a particular Fibre Channel, Advanced Technology Attachment (ATA) or Serial Attached SCSI (SAS) target.
For More Information

**HP Converged Infrastructure**
HP Converged Infrastructure white papers and videos
http://www.hp.com/go/ci
HP Converged Infrastructure Reference Architecture Guide

**HP FlexFabric**
HP FlexFabric white papers and videos
http://www.hp.com/go/flexfabric

**HP Intelligent Resilient Framework**
HP IRF White Paper—reducing network complexity, boosting performance with HP IRF technology

**HP Virtual Connect**
HP Virtual Connect data sheets and videos
http://www.hp.com/go/virtualconnect

**HP Intelligent Management Center**
HP IMC data sheets and product details

**HP TippingPoint Security**
HP TippingPoint data sheets and product details

**HP Networking Services**
HP Networking Services brochures and videos

© Copyright 2011 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

UNIX is a registered trademark of The Open Group. Linux is a U.S. registered trademark of Linus Torvalds. Microsoft is a registered trademark of Microsoft Corporation in the United States and other countries.