



DATA CENTER

Industry Trends and Vision: Evolution toward Data Center Virtualization and Private Clouds

Discusses current and future applications of virtualization, cloud computing, and convergence in the data center and presents the Brocade vision for building the next-generation data center with an emphasis on network and business requirements.

BROCADE

CONTENTS

Introduction	3
Executive Summary	3
High-level Architecture Model	4
Virtualization	4
Cloud Computing	5
Physical Infrastructure	6
Overview of Building Blocks	6
Network Convergence in the Data Center	7
SAN and LAN Architecture Review	8
Network Convergence Options.....	8
Virtualization Architecture with Separate Networks	9
Virtualization Architecture with Converged Networks.....	11
Network Convergence Considerations.....	12
Cost	12
Security	12
High Availability	12
Change Management.....	13
The Brocade Vision: Revolution Through Evolution	13
Start with the Target Design.....	13
Add Network Convergence Where Needed	14
Remove Network Tiers.....	14
Define a Pathway and Milestones	16
Start Making Strategic Decisions.....	17
Appendix A: Virtual Server Technology and Network Building Blocks	18
Server and Application Virtualization Software	18
Command and Control Software (Orchestration).....	19
High-Performance, Intelligent Client/Server Network (Client-side LAN)	19
High-Performance Server/Server Network (Server-side LAN)	20
Shared Block Storage Network (SAN).....	21
High-Performance Storage	21
Geographically Scalable Layer 2 Network (WAN).....	21
Coordination with Application Delivery Controllers for Application Provisioning (Layer 4-7)	22

INTRODUCTION

This paper focuses on the architectural requirements for data center networks that support virtualization and cloud computing. The intended audience is architects and designers planning large-scale data center virtualization and a phased rollout of private cloud computing. Such a plan starts with products and features that are production ready today but also have the potential for additional capabilities as the data center evolves.

The content is organized as follows:

- An introduction to the concepts and design considerations associated with data center virtualization and cloud computing
- High-level architecture models for deploying data center virtualization and private cloud computing
- A discussion of network convergence in the data center, including the impact of convergence on the total cost of data center infrastructure
- Recommended questions to ask and steps to take as you work toward your future data center
- A review of important technologies used for deploying server virtualization and private cloud computing in the data center (Appendix A)

EXECUTIVE SUMMARY

How can virtualization and cloud computing be used to lower total cost without sacrificing performance, availability, security, and operational efficiency in the data center? What are the design considerations surrounding network convergence, which is often positioned as the network architecture for server virtualization?

Cloud computing and server virtualization strategies abound. Cloud computing can be private, public, or a hybrid of the two. Although this paper briefly discusses public cloud computing, the focus is on cloud computing in the data center. Virtualization can be applied to servers, storage, networks, and desktops. Mainframe and UNIX server virtualization technologies are mature and widely deployed. But x86 server virtualization is emerging rapidly and this paper discusses the impact of x86 server virtualization on the data center network.

Network convergence was originally applied to the convergence of voice, video, and data on the same telco network. But when network convergence is applied to the data center, it means transporting IP network and block-level storage traffic on the same physical network. Although this is a separate topic from virtualization and cloud computing, some have positioned it as essential for virtualization to succeed. Two high-level architectures are presented, one with network convergence and one without. Physical network convergence can introduce complications for availability, security, and operations if it is applied indiscriminately and without careful attention to design considerations. This paper discusses the key considerations for engineering a network for converged traffic, so that the network designer can make an informed decision about how best to meet the goal of lower total cost without compromising business requirements for performance, availability, security, and operational efficiency.

Much discussion about server virtualization and cloud computing highlights immediate access to unlimited amounts of computing, storage, and network bandwidth. But, of course, virtualization and cloud computing cannot exist without data centers and the physical hardware they house. To support virtualization, the data center architect has to harden the network against failures and make it adaptable and flexible without disrupting traffic—and to do this while continuing to support existing data center assets. A key conclusion is this: physical assets needed to support virtualization largely exist today. They will need to be extended to support server virtualization software where it makes sense and when it makes sense. *Doing so does not require upgrading the entire data center network from the access layer to the core.*

Finally, this paper shows how to take advantage of the Brocade® networking strategy to deploy server virtualization and cloud computing in the data center. It is based on a “revolution through evolution” model for controlled change and calls for the separation of virtualization software from the network foundation—using open standards. Open standards enable flexibility about what x86 virtualization software is deployed and what applications it supports and ensures that it can leverage a common network foundation. An open architecture recognizes that virtualization software is evolving rapidly. There will be winners and losers as Darwinian evolution progresses in these markets. Therefore, a proprietary architecture that tightly couples software with hardware is risky. Tight coupling can mean changes in virtualization software will ripple to the network and server foundation and necessarily disrupt the business.

HIGH-LEVEL ARCHITECTURE MODEL

IT needs to do more with less—the one thing everyone agrees on. Therefore, any high-level data center architecture focuses on the need to first reduce operating cost. It may be tempting to trade performance and reliability for lower cost, but a sound architecture delivers greater reliability over time, since more applications are becoming mission critical, and improves performance so that response time doesn’t suffer as more data becomes digital and more users access more data.

The virtualized data center, which leverages server virtualization and private cloud computing, promises to do more with less. Total cost is reduced by improving utilization and reducing time to deploy new applications, while offering 24x7 application availability when hardware has to be serviced or replaced via application mobility. To validate this value proposition, the underlying network architecture has to increase availability, performance, and security while simplifying data center procedures.

Network convergence, in which IP and storage traffic flow on the same server Ethernet network, is often linked with server virtualization. The goal of lower total cost via virtualization can be achieved with either network convergence or without, but due diligence is important when considering network convergence.

Before considering two high-level virtualization architectures, one with network convergence and one without, it may be helpful to review x86 server virtualization and cloud computing concepts.

Virtualization

While the technology for virtualization is not new, the scale at which it is being deployed is unprecedented. The concept of a virtualized data center means that every aspect of every piece of hardware is abstracted from every other piece of hardware. Moving data from one array to another, moving applications from one server to another, moving network services from one switch to another—essentially all add, move, or change operations—could be conducted without applications or users knowing anything about it. It is possible to build an abstraction layer that covers the entire data center itself. Solutions exist today to support nearly seamless application mobility across inter-state distances without disrupting applications or users being aware of this movement.

Virtualization means that the services provided by a hardware device are abstracted from the physical hardware. For example, if a disk provides the service of “storing data” for a server, the virtualization abstraction layer separates the server’s access to the “storing data” service from the physical disk on which the data is stored. This allows the disk to be serviced or upgraded to increase capacity without disrupting the application while it is storing data. The “storing data” service the storage performs persists even when the hardware supporting the service changes.

A sound infrastructure for the next-generation data center should be robust enough to support virtualization of every component and the data center itself. Not all applications require or even benefit from every type of virtualization. Therefore, the infrastructure must be flexible enough to support both applications on dedicated hardware and applications on virtualized hardware. This implies that any network port can be configured for performance (bandwidth), security policies, and oversubscription ratios dynamically via management software.

A network with these kinds of intelligent capabilities will be able to meet a wide range of application service levels dynamically. Achieving this degree of flexibility is the long-term goal of virtualization and directly supports the business objectives behind cloud computing. Brocade believes that it will be achieved incrementally, “one rack at a time,” in the same way as prior technology transitions have been implemented in the data center.

Cloud Computing

In cloud computing, applications, computing and storage resources live somewhere in the network, or cloud. User's don't worry about the location and can rapidly access as much or as little of the computing, storage and networking capacity as they wish—paying for it by how much they use—just as they would with water or electricity services provided by utility companies.

Some features that apply to cloud computing are:

- **Virtual infrastructure to provide resources.** The data center itself becomes a dynamic pool of resources, enabled by virtualization technology. Applications are not constrained to a specific physical server and data is not captive to a single storage device. Operations focus on ensuring that adequate resources are available; the function of service provisioning handles what resources are allocated to an application or user.
- **Service provisioning.** Services must be provisioned with little or no effort on the part of the IT group responsible for maintaining the resource pools. Self-service portals that users can access let them request computing, storage, and network connectivity—all provided within minutes. This is a significant departure from the earlier IT model of project-based application deployment on dedicated hardware.
- **Payment at time of use.** Cloud computing supports quite a few innovative financial models, such as pay-as-you-go based on the resources used, and even a no-cost model in the case of some public cloud applications in which advertising pays for the infrastructure.

Cloud computing over the Internet is commonly called “public cloud computing.” When used in the data center, it is commonly called “private cloud computing.” The difference lies in who maintains control and responsibility for servers, storage, and networking infrastructure and ensures that application service levels are met. In public cloud computing, some or all aspects of operations and management are handled by a third party “as a service.” Users can access an application or computing and storage using the Internet and the HTTP address of the service. Google Apps is a well-known example of public cloud computing, in which virtualization resides between the Internet connection and the data centers delivering the Google Apps service.

Clearly, public cloud computing is at an early stage in its evolution. However, all of the companies offering public cloud computing services have data centers, in fact, they are building some of the largest data centers in the world. They all have network architectures that demand flexibility, scalability, low operating cost, and high availability. They are built on top of products and technologies supplied by Brocade and others network vendors. These public cloud companies are building business on data center designs that virtualize computing, storage, and network equipment—which is the foundation of their IT investment.

And they are building those data centers using proven products available today. Their data center architecture keeps the virtualization and cloud computing software stacks (which in many cases are proprietary) independent from the foundation networking, storage, and computing hardware. Clearly, these companies are betting the business on an open hardware architecture that decouples the software stack from the hardware. *This is the architecture Brocade supports and the reason we have significant footprint with public cloud service companies worldwide.*

PHYSICAL INFRASTRUCTURE

Cloud computing and server virtualization complement each other in the following ways.

- Cloud computing services can be implemented on top of virtual data centers. Virtualization can support a cloud architecture.
- Cloud computing software can be used to orchestrate virtual server deployments. Cloud management software can be used for command and control of virtualization services.
- Cloud computing adds another virtualization layer between end users and the entire IT environment, implementing a pay-per-use model.
- Both demand robust physical infrastructure. They both rely heavily on the network and are demanding changes in traditional assumptions about network architecture and design.

This last point is important. As more server virtualization and cloud computing projects are deployed in the data center, the existing network design has to gracefully adapt, one rack at a time, while continuing to operate with non-virtualized applications.

Overview of Building Blocks

The high-level data center architecture consists of building blocks that abstract the details about the actual equipment being used. It identifies key design elements and allows the architect to decide where those elements are deployed and how they are implemented. The data center architecture of the future can be deconstructed into a set of building blocks, illustrated in Figure 1 and discussed in greater detail below. A robust architecture includes all of these building blocks and considers how and where they should be implemented.

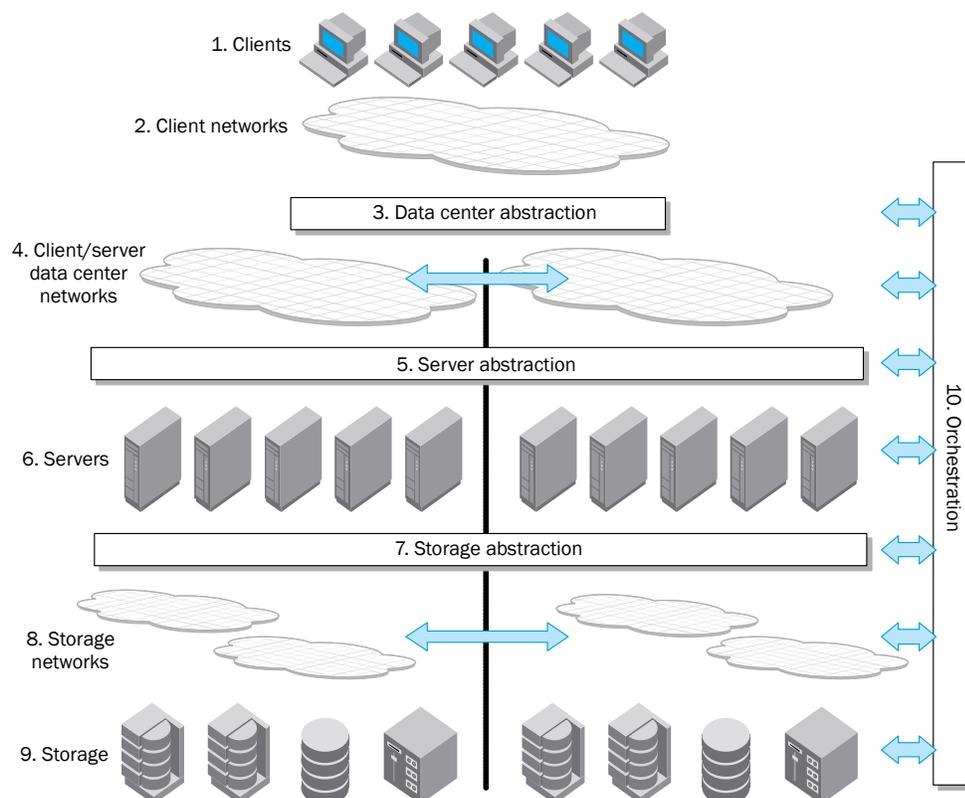


Figure 1. Next-generation data center, non-converged

Here are the building blocks:

1. **Clients.** All data centers have the mission of serving applications to users. These might be thin clients, smart phones, tablet computers, virtual desktops, or traditional PCs and laptops.
2. **Client networks.** Regardless of the type of client, clients will need to connect to the data center via a network. This is generally optimized for cost rather than performance or features.
3. **Data center abstraction.** If you want to support a multi-data center environment for Disaster Recovery/Business Continuance (DR/BC) purposes or use a blend of internal/external cloud, a layer of abstraction needs to be implemented between clients and data centers. Application data controllers with advanced application resource provisioning and management features, such as the Brocade ADX Series, provide several advantages and are discussed in Appendix A.
4. **Client/server data center networks.** Each data center requires a high-performance, high-reliability LAN. To support application mobility between data centers or between internal and external clouds, you will need high-speed connectivity between the data centers, Layer 2 extension between data centers and “virtualization-aware” intelligence all the way to the network edge.
5. **Server abstraction.** This is where most of the management of the physical resources and cloud servers takes place: whatever will be delivered “as a service” needs a management layer that sits above the physical platforms.
6. **Servers.** At some point, virtual servers must reside on physical platforms. These will need to be very robust, as the value proposition of server consolidation doesn’t work well if you can only deploy a few Virtual Machines (VMs) per platform. Also, be sure to avoid internal performance bottlenecks, such as oversubscribed blades.
7. **Storage abstraction.** If you want to support non-stop operations, you can’t take down VMs whenever you need to upgrade a RAID array.
8. **Storage networks.** Every server must have shared storage to support VM mobility. Any type of HA requires at least two physically separate SANs per data center. This is a long-standing best practice, and there are very good reasons why SANs do not follow the LAN model of a single, fully connected network.
9. **Storage.** Historically, SAN administrators used rule-of-thumb ratios of initiators to targets and similar strategies that relied on knowledge about the behavior of the network endpoints. Because endpoints can move dynamically, these strategies break down, so storage systems and SAN paths must be fast and reliable enough to support any application.
10. **Orchestration.** Orchestration requires coordinating LAN and SAN behaviors—whether or not LAN and SAN traffic is converged.

Network Convergence in the Data Center

There are many different ways to design a data center to support virtualization and cloud computing services and there are a number of key decisions to make. An important decision that impacts the top-level network architecture is network convergence. You can deploy network convergence or not and still deliver the cost reduction driving virtualization and private cloud computing.

Note that network convergence has several meanings. In the past, it meant converging voice, video and data for unified communications on a single telco or cable company network connection. In the data center, network convergence means merging LAN traffic (IP network) and SAN traffic (block-level storage network) on the same physical network infrastructure and is the definition used in this discussion. But before making a decision about converging networks, it’s important to understand the various iterations and what is taking place.

SAN and LAN Architecture Review

The overwhelming majority of SANs in data centers employ two or more physically separated Fibre Channel (FC) fabrics dedicated to moving storage traffic between servers and storage subsystems and are built to ensure guaranteed delivery of storage traffic. Compare this with a typical LAN used for IP traffic. Virtually all data center and client-side LANs use some form of Ethernet deployed in a single, fully connected topology with multiple switching stages and network layers. They provide “best effort” delivery of IP traffic, meaning that frames are discarded when congestion occurs. There may be resilient network elements, but the network operates as a single entity. The IP network and the storage network also have quite different operational requirements and so they are managed independently, most often by different administrators with different skill sets and independent operational procedures.

Network Convergence Options

Network convergence of LAN and SAN traffic has broad implications beyond the simple notion of merging traffic on the same wire to reduce capital costs. Fortunately, convergence is not an all-or-nothing proposition. Network architectures can make a number of choices about the scope and extent of network convergence as outlined below:

- **No convergence.** Retaining the classic architecture is a valid choice. However, integrating LAN and SAN management functions with virtual server orchestration software can automate changes when VMs are moved across physical servers as described below.
- **Management convergence.** Orchestration requires coordinating LAN and SAN behaviors. Single-pane-of-glass management requires that tools talk to LAN and SAN switches at the same time. It does not require convergence of IP and Fibre Channel traffic on the same physical network.
- **Layer 2 technology convergence.** You can retain physically separate networks, but use the same type of Layer 2 (data link) infrastructure for IP and Fibre Channel traffic. Physical separation implies separate switches for LAN vs. SAN (which have no network path between them), not merely the use of VLANs. This can reduce cost and simplify training. Other storage protocols such as iSCSI, NFS, and CIFS can continue to be used and best practices for LAN and SAN design do not change.
- **Access layer convergence.** Physically converging IP and Fibre Channel traffic inside a server, external network adapter, and top-of-rack switch can reduce the number of cables and switches required. A number of converged switch products are available that separate IP and Fibre Channel traffic at the top of the rack, forwarding each to the appropriate layer (the aggregation switch for IP traffic and the core switch for Fibre Channel traffic).
- **Entire network convergence.** Full physically converged IP and storage networks require a fully connected network that can handle IP and Fibre Channel traffic on any port.

The total cost of each of these options should be considered beyond initial costs, as they impact performance, availability, security, and operations as well. Physical network convergence is one option that will grow in adoption as technology develops, but it is not an inevitability for all customers and all applications. When and how far to converge the IP and Fibre Channel traffic is a decision that should be made in the context of *all of the requirements that have to be met*.

The strategy of separate storage networks was driven by application requirements. High-performance applications that are mission critical do not use file systems to access storage. Instead they use block-level storage access. Networked file systems were not efficient enough; they did not provide enough bandwidth and were subject to the availability limitations of a “best effort” Ethernet network.

There are other requirements for block storage traffic that kept it on a separate physical network from the IP traffic. These include a fixed amount of network latency, independent change control processes, regulatory compliance, and security. These stringent requirements reflect the difference in how applications and servers react to disruptions when accessing storage versus disruptions in client access to applications. Client/server applications are designed to handle interruptions in the client connection, but servers are designed to assume that storage connections are always available. These are critical differences and strongly influence the design of IP and storage networks. For these reasons, physically separating storage and IP networks is common practice and dual physically separate storage networks is best practice.

Over time, alternatives to Fibre Channel SANs have been developed. For example, the iSCSI protocol has had success. It gained market share in smaller storage environments but did not displace Fibre Channel SANs in enterprise data centers. iSCSI is a good solution for shared storage in smaller environments and Fibre Channel is the right choice in large, enterprise environments. However, although iSCSI uses TCP/IP over an Ethernet LAN, best practice restricts its traffic to a physically separate Ethernet network for the same reasons Fibre Channel storage and IP network traffic have been separated. Clearly, the network protocols used are not the reason IP networks and storage networks are physically separated.

Brocade's data center network architecture provides choice, including protocol choice and choice about the extent of network convergence. Each data center has its own requirements and one architecture does not fit all. The following section discusses two architectures, one without network convergence and one with, and suggests what you should consider when using them with server virtualization and cloud computing.

Virtualization Architecture with Separate Networks

Figure 1 illustrates a flexible architecture showing key building blocks. Where and how those building blocks are implemented are important decisions the architect has to make. For instance, although the orchestration software block is shown, an architecture that omits it could also be valid.

This architecture is shown at a high level because there are many ways you can implement each of the building blocks. Note that the client/server data center networks and storage networks are physically separate LANs and SANs. As discussed earlier, SAN best practice is to implement at least two physically separate SANs in each data center, generally referred to as "Fabric A and Fabric B," for highest levels of availability. (See *Principles of SAN Design* by Josh Judd for a discussion of SAN best practices.) Server and storage abstraction layers are shown as they virtualize the data center: server abstraction separates the applications from the physical server and storage abstraction separates the storage from the disk array.

Often, additional networks could be positioned between servers, as shown in Figure 2. Use of dedicated networks for hypervisor-to-hypervisor traffic such as VMware VMotion, server management, and metadata such as cluster heartbeats is best practice as this traffic is critical and physical isolation ensures that it is not disrupted. There could be one or more tape backup networks as well. For these reasons, it is common for physical servers running server virtualization software to have eight to ten Ethernet NICs in addition to two Fibre Channel HBAs.

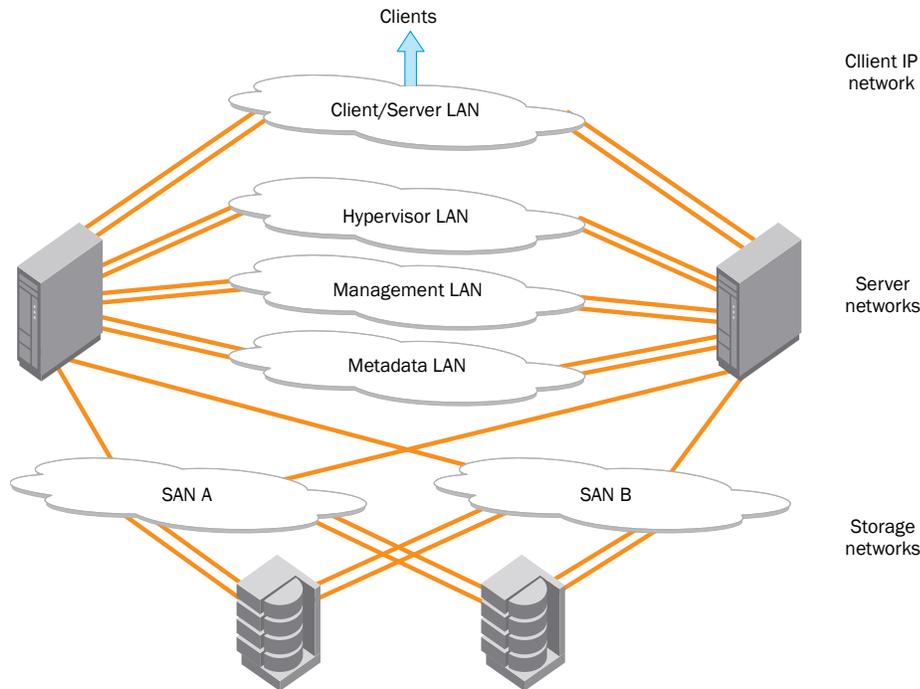


Figure 2. Server networks

All of these networks could be converged immediately if the only barriers were the selection of Layer 2 technology and bandwidth, however there are additional considerations. Other networks may be kept physically separated for reasons unrelated to protocol. You could use the same switches for Internet DMZ and cluster heartbeats, but you wouldn't because of security concerns.

Even the Voice over IP (VoIP) analogy is a poor comparison for physical convergence in the data center. Certainly, voice and IP communications are often unified for employee telephones, but the analogy between desktop phones and enterprise storage is not a good one as they are not equally as critical and don't have the same technical requirements. Humans are capable of understanding speech even when large numbers of packets are dropped, but storage cannot tolerate even one lost packet. If all VoIP phones in a company simultaneously stopped working for one second and then came back on, it's unlikely that anybody would open a help desk ticket. At any given moment, most people aren't using their phones, and wouldn't even notice the outage. Most others would simply redial. But the same incident in a storage network could severely impact applications in the data center causing delays or application disruptions. This is why factors such as Quality of Service (QoS), management, performance, and security must be considered for each application environment in the data center (tier 1, tier 2, tier 3, and so on).

Whether additional networks exist or not, there are at least three types of networks as shown in Figure 2: the client IP network, the server network, and the backend storage network. The client network may be a campus or enterprise network with access to the Internet. With the possible exception of distance storage extension tunneled over the WAN, there is no physical convergence between the LAN and SAN traffic. It is important to be clear on this point: *neither virtualization nor cloud computing requires any change to this model* any more than VoIP required physically converging carrier and emergency service phone networks. You can choose to implement cloud computing and server virtualization using an architecture that physically separates IP and storage networks and there are often important reasons to do this.

In a physically separate LAN and SAN architecture, the management model you select can monitor and control both the SAN and the LAN, in addition to the server and storage nodes. This is what the orchestration building block on the right in Figure 1 would be responsible for.

Also note that this model can be used even if you chose to use network Layer 2 (data link) Ethernet technologies with storage traffic. As mentioned previously in reference to iSCSI, it has become best practice to maintain physical separation of the storage network from the IP network. Although these networks all use Ethernet switches, they are deployed, maintained, and managed as physically separate networks. Similarly, you can use emerging technologies such as Fibre Channel over Ethernet (FCoE) via a physically separate lossless (Data Center Bridging, or DCB) Ethernet network.

Virtualization Architecture with Converged Networks

There are many opinions about if and how far LAN/SAN network convergence should be pursued at the physical level. Figure 3 illustrates a physically converged architecture.

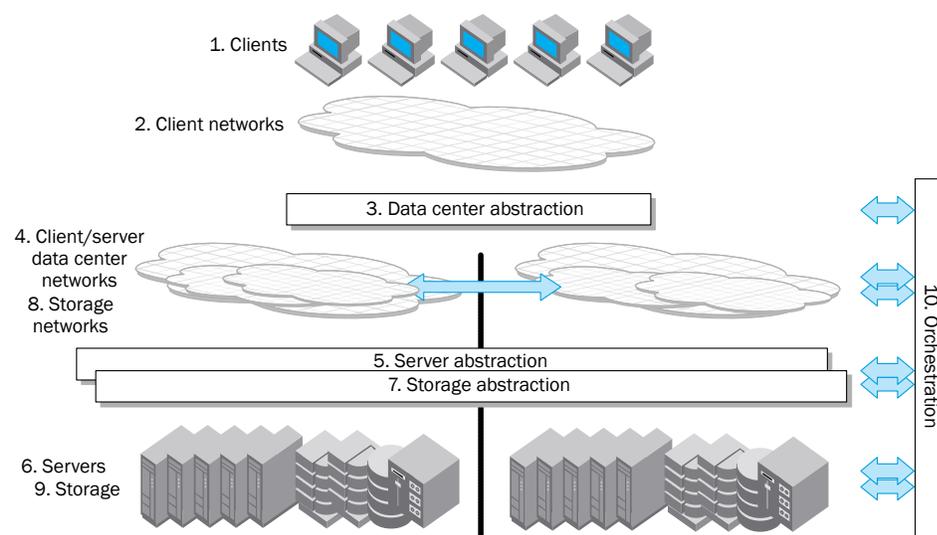


Figure 3. Next-generation data center, fully converged

Note that network convergence does not eliminate any of the building blocks from the model. The physical network for IP and storage traffic and some of the abstraction layers have been merged to illustrate how network convergence would affect these building blocks.

Physically converging IP and storage traffic does not merge their network services. Having server and storage virtualization layers at the same tier of the data center does not mean merging either their software or hardware components. You still need separate interfaces into the orchestration software to manage, monitor and configure the SAN and LAN services in the converged network. IT personnel will still need storage management skills and LAN management skills, as these skills don't readily "converge" in a single person. Similarly, the interfaces for storage and server virtualization will remain independent. These are indicated in both figures by arrows from the orchestration layer.

The converged network model requires removing the separation between SANs and merging what was once two physically independent storage networks into a single network. As discussed later, this changes the fundamental architecture of the SAN and LAN.

Network Convergence Considerations

The following sections discuss some factors to consider when you are thinking about network convergence; additional factors you should consider include performance, compliance, and network management.

Cost

The underlying driver for network convergence in the data center is cost. The idea is that half as many network interfaces and switch ports are required, thus reducing cost. All things being equal, the data center should be less expensive to deploy and manage with half the switches, half the cables, half the power; and half the cooling. But while this is a simple, powerful value proposition, at a deeper level, all things are not equal. When IP and storage traffic are converged over the same physical network, every device along the data path (server interface, device driver, switch, uplink, and so on) must support a superset of LAN and SAN protocols. The network path must be engineered to provide lossless rather than best effort service. Therefore, the cost of equipment upgrades along that path offset the initial cost savings.

Digging a little deeper, the efficiency of a network is strongly influenced by the efficiency of the protocols it uses. In particular, control traffic has to be efficient or the theoretical performance won't be achieved. Control frame size and the number of control frames processed are the key metrics. If one protocol has larger control frames and more of them, either the control processing hardware has to be more powerful (and more expensive) or actual performance will be reduced. Network convergence relies on the FCoE protocol to carry the storage traffic. FCoE has larger frame sizes and uses more control frames than the native Fibre Channel transport protocol. Therefore, FCoE switches have to be more powerful than standard Ethernet switches, which will affect the cost per port. This increase in port cost along the converged network data path has to be offset against the initial cost savings.

Security

Security in a converged network is also different. Without convergence, there is an air gap separating the SAN from the sources of threats in IP networks. A hacker cannot directly target a storage device since they cannot get to it through any IP network port. Hacker tools and techniques designed for IP or Ethernet do not work in Fibre Channel SANs. Data center architects have to meet the security requirements of all types of traffic as they make decisions about the extent of network convergence.

High Availability

The high-availability (HA) model is another critical decision. The Ethernet LAN and Fibre Channel SAN approach HA differently, and reconciliation is not straightforward. In IP networks there are methods for supporting separate frontend and backend networks and means to provide link and component failover. But the primary client/server network needs to be fully connected for redundancy solutions, such as NIC teaming or Ethernet trunking, to work properly. SANs are architected to provide HA without any connectivity between redundant elements. This is done to prevent network-layer problems from taking down both SAN connections at the same time. In the LAN world, a Denial of Service (DoS) attack, broadcast storm or similar event impacts all LAN interfaces, even if they are on separate switches. This is considered an acceptable risk in a best effort network. In a SAN, an equivalent incident could—and usually would—cause every application on the SAN to crash simultaneously. Thus, the best practice is to have two air-gap isolated networks, dramatically lowering the chance of a simultaneous outage: both air-gaps separated from the LAN and horizontally separated from each other.

Change Management

Finally, there are decisions about change management. If the networks are converged, then any change made to the LAN will also impact the SAN and vice versa. It is not physically possible to segregate these functions. This affects service windows, maintenance procedures, and operational procedures. All of these need to be addressed prior to deploying network convergence architectures.

The data center architect is confronted with not an either/or choice regarding network convergence, but:

- Decisions about how far to extend network convergence
- Choices about product selection based on overall performance requirements
- Choices about the security, HA and change management
- Choices about which applications are suitable for network convergence.

The goal is lower overall cost, but reaching that goal requires the consideration of many factors.

THE BROCADE VISION: REVOLUTION THROUGH EVOLUTION

Brocade has optimized its technology and partner relationships to allow you to implement a virtualization and private cloud computing architecture at a pace that makes sense for your business. As a piece of hardware is depreciated and slated for retirement, new components seamlessly move in to replace it. New platforms are often acquired for growth or to improve performance for critical applications. Those applications can migrate onto the new platforms, and unless their hardware resources were completely obsolete, those resources can be repurposed into the server pool for lower-tier applications. Eventually, by rolling in new platforms and repurposing old ones, the data center will be transformed: it indeed will use a revolutionary approach to IT management, but one step at a time. This section provides an overview of the “revolution through evolution” process Brocade advocates.

Start with the Target Design

Decide how you want your data center to look in the future. Select the extent and scope of LAN/SAN convergence, the number of layers within each network, and the number of switching tiers in each layer. Ask a series of questions about your ideal model to make sure that all of your future investments bring you closer to that goal:

- Will the final network have physical access-layer switches in each rack or will that function reside inside virtual servers?
- What level of oversubscription is acceptable at each stage of the network, given your target server consolidation ratio?

NOTE: Calculate how much bandwidth each VM would get if they all tried to talk at the same time. Trace through ratios such as hypervisor consolidation, blade server-to-blade switch, NIC-to-uplink and uplink-to-switch, and internal bottlenecks inside switches. The result is sometimes surprising, especially for physically converged designs

- Will there be an aggregation layer on the LAN or will large virtual servers connect directly into a high-port-count collapsed access/aggregation layer?
- Do you want to be locked into one orchestration tool and hypervisor vendor or will you select different solutions to meet differing needs for separate applications and departments?
- Will you be able to continue using any of your existing equipment in that design? And if not, why not?

NOTE: If you don't have a clear picture of your current equipment and how it is being used today, perform an audit. Without knowledge of the present, it's hard to build a vision for the future.

At this stage, unless you're building a new data center using all new equipment in the next month or two, you don't need to create detailed wiring diagrams or deployment guides. You simply need the broad strokes.

For example, say that your target design is based loosely on Figure 1 or Figure 2. You intend to converge management layers and potentially Layer 2 technology, but retain physically separated IP and storage networks, dictated by security, application HA, performance, and change management decisions. The business constraint is to keep using as much existing data center investment and procedures as possible (a common business requirement). Perhaps some of the older assets can be deployed and perform different missions or run different applications, but overall current assets will remain.

Next, identify servers and storage that can be repurposed to run with hypervisors and other abstraction layers once their applications are migrated off.

For the IP network, the target top-level structure is the same and requires no change. The two network tiers and full-performance switches deployed in two physically separate iSCSI or Fibre Channel SAN fabrics meet the scalability, availability, and performance requirements for the higher number of applications per physical server requirement for server virtualization.

A new requirement is convergence of management, which Brocade provides for existing environments via Network Infrastructure as Service (NlaaS) management architecture. NlaaS open interfaces enable management integration with leading orchestration software platforms.

Add Network Convergence Where Needed

As an option for some applications, servers will converge IP and storage traffic at Layer 2 using Ethernet and Fibre Channel technology. This requires FCoE support and lossless Ethernet (DCB) switches. The Brocade DCX® Backbone and Brocade 8000 top-of-rack switches support FCoE and Ethernet DCB extensions. These products integrate with existing FC storage while also supporting backend NAS protocols such as NFS or CIFS, iSCSI, and FCoE storage traffic. FCoE could be terminated natively by attaching FCoE storage to those same products or it can be bridged into the existing FC storage pool. The extent of the converged network will be limited to those network segments that need it and this is done without impact on the aggregation or core of the IP network or disruption to the core of the SAN.

Remove Network Tiers

For the IP network, it may be necessary to take a more aggressive approach in order to reduce cost and complexity. With server virtualization, the access layer LAN will need to use full-performance switches, have a flattened topology with at most two switching tiers, be designed with low oversubscription ratios on uplinks, and use intelligent switches with open standards support and open integration with orchestration software. Support for these requirements may require upgrades to network assets.

Many companies are using virtualization projects to replace aging LAN infrastructure. The cost of maintaining these devices is generally higher than replacing them, offering an opportunity to cost-effectively collapse the number of tiers in the IP network. For example, a two-tier architecture with Brocade NetIron® MLX routers in the core providing Layer 3 services and at the edge running Layer 2 (shown in Figure 4), provides significant capital and operating cost savings, while delivering wire-speed performance, seen increasingly as a requirement for highly virtualized server environments. Brocade offers a full line of IP networking products for the data center.

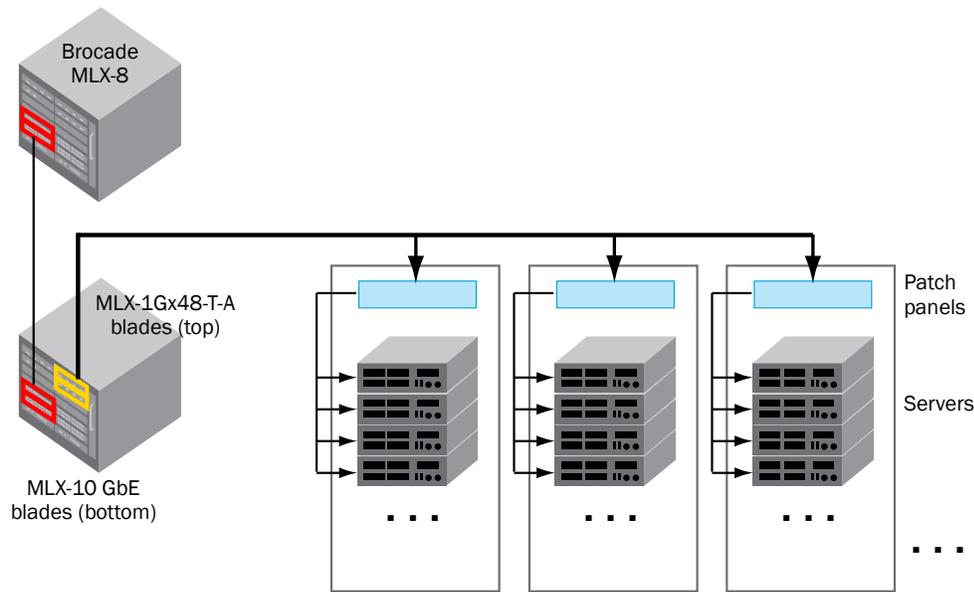


Figure 4. Brocade MLX at the end of the row

Utilizing the MLX-1Gx48-T-A blade, the Brocade MLX can deliver full Gigabit Ethernet (GbE) connectivity at top of rack (no oversubscription) to over 1,000 physical servers hosting thousands of VMs and traffic flows (depending on the chassis model). The blade effectively extends the chassis slots to the top of the rack, eliminating the need for top-of-rack switches and high-speed uplink optics.

For conventional architectures, the MLX platform also delivers high-density, 10 GbE connectivity to Brocade top-of-rack switches. If the target design calls for 40 or 100 GbE links in the future, the MLX is built to support 40 GbE and 100 GbE blades when they become available. Brocade MLX also supports Multiprotocol Label Switching (MPLS) and Virtual Private LAN Service (VPLS), so you can implement multiple networks logically with complete isolation using the same physical L2 network, as well as provide L2 connectivity between physical sites over the WAN.

Similar strategies can be used for physically converged designs (shown in Figure 3). If you want a physically converged network using iSCSI or a NAS protocol, for example, the LAN described above could support your requirements. In fact, Brocade has been supporting converged iSCSI and NAS networks since those protocols were released.

It isn't always possible to remove layers from a data center network due to cable management or geographical constraints. If your data center is split across several rooms or floors in a building or multiple buildings in a campus, you may require added tiers. The rule is to design the network using the fewest number of layers practical, given all the other constraints. Think of each geographical region or room as a pod in the data center. Then, make an effort to localize clusters and I/O patterns within each pod and pursue as collapsed a design as possible inside the pod. MPLS/VPLS solutions can still allow VM mobility across pods.

These rules apply to planning a target design:

1. Ensure that the target meets all scalability, availability, security, manageability, cost and operational requirements.
2. Preserve as much of your existing equipment as possible.
3. Replace LAN switches with full-performance, virtualization-ready technology.
4. Ensure that all network elements can be managed in accordance with the current and target architecture for infrastructure management.

Define a Pathway and Milestones

The key to achieving “revolution through evolution” in data center networking is to move toward a target design along a well-planned path and to use small steps to control risk. You should create separate timelines for each of the functional areas defined in a previous section, “Overview of Building Blocks.”

Using the rules outlined above, consider these actions when making decisions about the steps and the scope of change in each one:

- On the server side, the first step is to classify applications and select the correct virtualization strategy for each. Some may benefit from hypervisors, while others would use application-specific solutions to scale horizontally. Some application groups might be Microsoft centric and fit optimally into Hyper-V servers, whereas others might work best with Xen or VMware ESX Server.
- Next, classify the existing server platforms that are modern and robust enough to be re-used in the target design. To identify candidates for re-use, make sure that they appear on the list of supported servers from the hypervisor vendor and that they have sufficient RAM and CPU power and appropriate SAN interfaces. If a server is otherwise sufficient but lacks FC SAN connectivity, consider purchasing a dual-port HBA: adding a card is often cheaper than replacing a server.
- If enough of your servers can be leveraged in the new design, perform a rolling migration, purchase a few new servers loaded with virtualization software, and migrate existing applications onto the new servers. This frees up the old hardware, which can be reloaded with virtualization software and then used as the migration target for the next set of applications. Eventually, all applications will be migrated to a virtualization platform, with most applications residing on hardware that you already own today.

Similar strategies can be applied in the LAN. For instance, assume that the target design uses the two-tier Brocade NetIron MLX and virtual top-of-rack design discussed above. Many aggregation-layer switches have oversubscribed blades and many access switches do not have interfaces to orchestration software. Replace these aggregation- and access-layer switches with the MLX chassis and connect into the existing core as your needs dictate. Router cores could be replaced with MLX chassis, also one at a time, over a timeline that makes sense.

Initially, the new MLX switches could also be used to connect to the existing top-of-rack access switches, in a classic access-aggregation topology. An MPLS/VPLS configuration could also be utilized as appropriate. As virtual servers are deployed, they would be connected directly to the MLX via NI-MLX-1Gx48-T-A blades or a Layer 2 solution (appropriate for their performance characteristics), thus flattening out the network topology. At this point, their Layer 2 functions could run over MPLS/VPLS, which would enable solutions such as long-distance VM mobility.

The access layer would move inside the servers using virtualization software, so there would be no need to upgrade access-layer switches. This often provides an opportunity to justify the added amortized capital expense of higher-performance, green switching products against the lower OpEx costs (power, cooling, and maintenance contract costs) of current technology. If you eliminate an entire tier of switches, you can save enough on operating costs to pay for a significant portion of the aggregation layer upgrade.

Eventually, once all of the hardware replacements and reconfiguration are made, data center servers will be fully virtualized. The SAN and LAN will both be managed by open interfaces supporting multiple orchestration software applications. You could select one orchestrator for the entire data center or different solutions for each application. Over time, you could connect this network to external cloud providers, for example, to provide capacity for peak loads or to more cost-effectively meet disaster recovery requirements.

In general, technology upgrades should follow this order:

1. SAN and storage upgrades (if required)
2. LAN and Layer 4-7 switches
3. Servers and hypervisors
4. Orchestration
5. Global- or data center-level abstraction.

Start Making Strategic Decisions

You can easily do all of the planning defined in the previous sections without spending money on software or hardware. This isn't surprising, since a major goal of Brocade's "revolution through evolution" strategy is to minimize cost by investing in interoperability when bringing new products to the market.

You will eventually need to deploy new equipment regardless of your long-term strategy. Even if you do not pursue cloud or virtualization, you will still purchase new servers and storage as equipment becomes obsolete and application demand exceeds capabilities. You will still need to replace obsolete LAN equipment. Even if you stay with your current infrastructure, you have to keep paying the OpEx cost of maintaining the old equipment—often spending more than you would if you purchased new, more efficient products.

Start optimizing your acquisitions now, with every purchase taking you toward your future architecture. Network products that support a variety of virtualization and orchestration vendors will serve you regardless of how these emerging technologies develop. As a leader in both SAN and LAN and as the most partner-friendly company in the business, Brocade is the optimal choice to provide the flexible, scalable, high-performance networks you will need, whatever direction you take in the data center.

APPENDIX A: VIRTUAL SERVER TECHNOLOGY AND NETWORK BUILDING BLOCKS

This appendix reviews the technologies used for deploying server virtualization and private cloud computing in the data center. In selecting these, the following design considerations should be kept in mind.

1. **Compatibility.** Can each software and hardware element talk to the others?
2. **Supportability.** Specifically, do vendors commit to making the configuration work?
3. **Availability.** Is the application available to the users?
4. **Adaptability.** Can the network adapt when applications move around constantly?
5. **Performance.** Are applications meeting user expectations for response time?
6. **Reliability.** Component failures bring operational cost even when applications stay up.
7. **Serviceability.** When something does fail, how hard is it to fix?
8. **Manageability.** How deeply can you see into the cloud for troubleshooting?
9. **OpEx cost.** Are you going to realize lower operations costs for power, cooling, and so on?
10. **CapEx cost.** Initial acquisition cost is the small ticket item, but still worth knowing.

The brief descriptions above are simplified. Manageability, for example, means a lot more than troubleshooting. It also encompasses day-to-day operations (adds, moves, and changes), capacity planning, performance monitoring, and so on.

In general, these are listed in order of priority for enterprise-class customers. For example, compatibility is first because if devices can't talk to each other, the solution won't work at all. Supportability is second because if vendors won't commit to the configuration, there won't be any way to fix it if it breaks. While CapEx is often cited as a #1 requirement for IT organizations, they are probably assuming all of the others have been considered whether or not they are explicitly called out.

Server and Application Virtualization Software

There are quite a few ways to approach virtualization of the server layer. Each approach has advantages in particular situations. Perhaps you want to build a single database application that can scale out beyond the confines of a single server. In this case, a parallel database package might provide the best fit. Although this doesn't virtualize the operating system in the same way as VMware ESX Server or Microsoft Hyper-V. It makes a single application span multiple operating systems. There are performance advantages to this approach, and if you are building a parallel database you're almost certainly doing it to increase performance.

If you're trying to virtualize the operating system rather than the application, there are still several options. Perhaps you have a department that relies heavily on applications such as Microsoft SQL or Microsoft Exchange Server. Using the native Microsoft hypervisor to virtualize and control those applications might be best. Another department might use all Linux applications and prefer to virtualize them using Xen. Yet another small group of applications might require continual shuffling between servers for resource optimization, in which case a couple of platforms running VMware VMotion might be better.

The idea is that different applications in different environments may be best served by different virtualization approaches. They may use different hypervisors or no hypervisors at all and instead rely on application-specific technologies. The approach you use for infrastructure and the upper-level command and control software must not preclude the use of different server-tier virtualization solutions. This is an area in which vendor lock-in techniques are often applied. Brocade has gone to great lengths to remain hypervisor neutral, so the software and hardware used in Brocade infrastructure can handle any single server-tier approach or combination. However, other infrastructure providers may not have been so careful in this respect, so "buyer beware" applies.

Command and Control Software (Orchestration)

From a technology perspective, private cloud computing is largely the decision about command and control software, which coordinates configuration of the software layers in the solution. You might use this software to enable a pay-as-you-go business model (as in utility computing) or you might use it to facilitate control over infrastructure you already own and have in place today. You might use one package throughout the organization or deploy different cloud environments by operating system or application tier. Either way, the features and characteristics of the cloud management software are highly business relevant. This is the tool that directly allows you to achieve the benefits of private cloud computing.

But it also means that it could cause the most harm if it didn't work properly. The ideal case of orchestration is an elegant ballet, with user load and compute resources rising and falling, in perfect synchronization. Which orchestration functions should be automatic and which require formal review as part of the change control process are important considerations. Some actions require human validation rather than automatic execution.

If you are pursuing a private cloud model be wary of becoming too dependent on any particular orchestrator, at least until the vendor has proven technology. You could have multiple orchestrators with independent orchestration domains instead of one solution controlling their entire data center. To accommodate these models, Brocade is taking an open approach: any major orchestration software vendor can monitor and control Brocade LAN and SAN behaviors via our open management platforms or via APIs. From an architectural perspective, this is critical to the selection of infrastructure. If a vendor delivers a fully vertically integrated stack—all hardware and software in a single bundle—it may seem easier to deploy and manage, but in reality it makes you dependant on that vendor. Unless your LAN/SAN provider has a market-neutral stance to orchestration packages, you will eventually confront dependencies that can prove expensive.

High-Performance, Intelligent Client/Server Network (Client-side LAN)

Server virtualization has implications for network performance and feature requirements. Historically, data center LANs followed a three-tier design. Lower-cost Ethernet switches at the top of each rack formed the *access layer*. Oversubscribed uplinks connected the access layer to *aggregation* switches, which were generally higher-performance modular chassis. These would be connected to the *core*, consisting of full-performance, modular switches or routers. This design worked well when each server at the access layer ran just one application, and high oversubscription at the access switch was acceptable. If the applications didn't all talk at once, then many, many servers could share the uplinks.

Server virtualization and the use of bladed servers have changed the game, because they add layers to the client-side LAN. By design, there will be one or more network layers inside every virtual server, a software switch or associated hardware. Either way, latency and layer(s) of oversubscription are added. If blade servers are used, there will be at least one more layer inside the server. This has dramatic implications for the design of the external client-side LAN.

For previous performance assumptions to remain valid, that is, a three-stage network with acceptable performance where one or more of those stages now resides inside the server, then some of the external LAN stages need to be eliminated. This isn't really a criticism of the virtualization. Having internal oversubscription is valid in virtual server environments—in fact, it's inevitable if you want to achieve the consolidation benefit that hypervisors can deliver. The point is that network architects need to recognize that oversubscription ratios may be very high as traffic consolidates towards the aggregation layer.

That being the case, the external LAN needs to behave more like a collapsed core: perhaps non-blocking or at minimum composed of non-blocking switches. Consider eliminating access-layer switches entirely and using core-class switches in the access layer. Alternately, use high-performance, fixed-configuration edge switches with low uplink oversubscription ratios and connect them directly to a large port-count collapsed core.

Intelligence at the edge is another concern. A key characteristic of a next-generation data center is that it is dynamic: server images and applications can move about at will, possibly even between different data center locations. To support this, the edge of the network will need to become increasingly virtualization aware. For now, simply understand that the use of commodity switches at the network edge is not suitable for deploying virtual server solutions. The LAN edge should be either an intelligent chassis—core or aggregation class—or an intelligent, high-performance, fixed-configuration switch.

High-Performance Server/Server Network (Server-side LAN)

Historically, most IP traffic in a data center was destined for client and storage traffic was on a different network. Virtualization solutions have increased the amount of server-to-server traffic on the LAN. Dynamic VM migration software can produce multi-gigabit sustained loads from any server to any other at any time.

To prevent this from impacting client/server communications, this traffic is usually not placed onto the client/server LAN. Physically separate Ethernet NICs are connected to physically separate Ethernet switch ports, often on physically separate Ethernet switches. There are several approaches that will work for this traffic in the data center of the future:

- **Server-to-server traffic can remain on a separate network.** The problem is that this separate network is ultimately going to be nearly the same size as the client/server LAN. That isn't the case today because only a fraction of servers are on virtualized platforms, but by 2012 half of all servers will be virtualized. This is still a valid approach and should be considered if you are pursuing the architecture shown in Figure 1.
- **It can share a backend network with storage.** Because SANs must currently be able to handle sustained high data rates, combining server/server traffic and storage traffic may be simpler than moving it to the LAN. The server/SAN concept has been around since the 1990s. If you are pursuing the model illustrated in Figure 1, you should also consider this approach. It keeps all intra-data center traffic on an isolated backend network, while still eliminating an entire network through convergence.
- **It can share the front-end network with client/server traffic.** Making this work requires even higher performance on the LAN side and advanced flow control and QoS features not present in most Ethernet switches.

Whichever approach you take, the best practices to consider are:

- Do not allow client/server traffic to interact with server/server traffic.
- Use full-performance LAN equipment, rather than oversubscribed switches.

Shared Block Storage Network (SAN)

Some current estimates have 70% of applications still running on physical hard drives inside servers. Cloud and virtualization are rapidly changing that, since shared storage is essential to both models. If you want to move a server image anywhere at will, the destination location must already have high-throughput, low-latency access to its data. When a VM moves, it has to maintain connection with its data.

There are quite a few ways to achieve the goal of ubiquitous data access. Depending on the application and its performance and availability requirements, you could use NAS for file system access to storage. However, as a practical matter, a scalable, production-grade solution does need a block-oriented SAN, whether it's Fibre Channel, FCoE, or iSCSI underneath the NAS file system.

The reason IT architects need to worry about this relates to the structure of SCSI. Anything that causes delay in creating and processing SCSI command frames produces a large impact on application performance. If the interface spends ten units of time creating headers for each unit of time it spends servicing payload, there will be increased delay. During this delay the application sits in an "I/O wait state," not doing any work.

There is a decision you need to make on a protocol for the SAN and whether or not to converge it physically. If applications are mission critical, it may be best to use physically separate "A/B" redundant SANs for HA. If applications require high performance in terms of bandwidth or delay in application response time, then you must use a block-oriented approach. If you have a mixture of applications and want to support a cloud model in which any network port might take on any category of application, then every server will need physically separate block-oriented SAN access.

High-Performance Storage

The same capabilities that prompt the use of higher-performing SAN infrastructure also call for higher-performing storage. If the Fibre Channel SAN delivers a control frame to a storage array in a few hundred nanoseconds, which then takes tens of milliseconds to respond, the application wait state condition will be similar to using FCoE, iSCSI, or even NAS. The bottom line is that delay anywhere in the storage pipeline translates into idle CPU cycles and slow application response times. That may translate into fewer VMs per physical platform and eventually into lower ROI on the virtualization or cloud model you are rolling out.

Geographically Scalable Layer 2 Network (WAN)

To support seamless application mobility between internal and external clouds or multi-site business continuity solutions, it must be possible to present the same IP subnet at multiple locations. While there are bleeding-edge proprietary solutions to this problem, there are also production-ready, standards-based solutions already on the market.

The solution with the largest production installed base today is MPLS. With MPLS, you can tunnel an Ethernet subnet across a Layer 3 WAN. Several solutions map onto MPLS to provide this solution, for instance, VPLS can create a Layer 2 pseudo-wire, which connects two LANs over distance. Both LANs use the same IP subnet range, so a VM can migrate from one side of the WAN to the other without changing IP addresses. This technology is well established and production hardened today. You can purchase MPLS/VPLS solutions from network equipment vendors such as Brocade and also from service providers.

Coordination with Application Delivery Controllers for Application Provisioning (Layer 4-7)

For applications that span multiple physical platforms and potentially multiple geographical locations, client requests must be routed to appropriate servers by application-aware network switches. In the 1990s, these were primarily used to distribute load across a single, flat tier of servers running a single application. Today, many applications are balanced across many sites (intelligent global load balancing) and provide more advanced application delivery functions.

For an orchestration solution to properly allocate and balance load in a cloud, it must be integrated with the application-layer switches in the network. If a new VM is brought online to handle increased client requests, application switches must be reconfigured to send new requests to the new VM. In fact, for the orchestration software to even know that a new VM is required, performance data from the application switch must be gathered and analyzed.

The application switch is in a unique position in the network. Because it sits in the data path between client and server and tracks application-level response time, bandwidth usage, and so on, it can provide much more concrete insight into the “user experience” of the application than tracking packets per second on a Layer 2 switch for example. If a Brocade application switch determines that application response time has exceeded a configured threshold, it can trigger the orchestration software to activate additional instances of the application to handle the excess workload.

This is done using software called the Brocade Application Resource Broker (ARB), which enables on-demand application resources in IT data centers. Brocade ARB, working as an extension to Brocade ServerIron® ADX Application Delivery Controllers, can dynamically add and remove application resources based on application traffic or response time demands. This is achieved by the real-time correlation of application performance intelligence from the network and infrastructure capacity information from virtualized server infrastructure.

Whether or not your future network contains ADX switches today, make sure that the solution used for application balancing can work with multiple orchestration vendors and provides the required hooks to allow intelligent automated provisioning of VMs.

© 2010 Brocade Communications Systems, Inc. All Rights Reserved. 04/10 GA-TB-277-00

Brocade, the B-wing symbol, BigIron, DCX, Fabric OS, FastIron, IronView, NetIron, SAN Health, ServerIron, and Turbolron are registered trademarks, and Brocade Assurance, DCFM, Extraordinary Networks, and Brocade NET Health are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. Other brands, products, or service names mentioned are or may be trademarks or service marks of their respective owners.

Notice: This document is for informational purposes only and does not set forth any warranty, expressed or implied, concerning any equipment, equipment feature, or service offered or to be offered by Brocade. Brocade reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact a Brocade sales office for information on feature and product availability. Export of technical data contained in this document may require an export license from the United States government.