



FocalPoint Switches in the Datacenter

Driving Consolidation, Virtualization
and Convergence

White Paper

November, 2007

Introduction

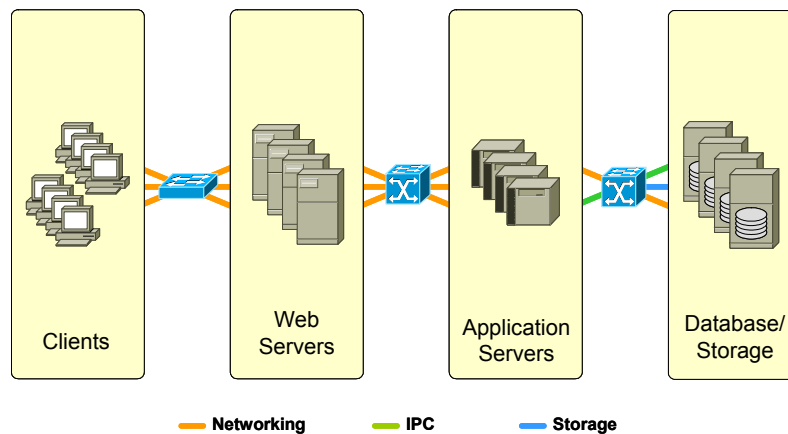
In the complex evolution of the datacenter, many technology trends have been noted but there are three that recently have emerged as key issues for the IT professionals in charge of datacenter installations. These three are consolidation, virtualization and convergence. Advances in these three areas help datacenter designers cope with the problems of scalability, space, power and cost as they evolve to meet increasing demands for applications, networking, storage and computation. As will become evident in this paper, there is a synergistic relationship between the three initiatives, where advances in one area enhance the value of the others. Volumes have already been written on these three topics, but this white paper will attempt to take a different tack from the rest. It will examine these trends while focusing on the features of Fulcrum’s FocalPoint 10 Gigabit Ethernet switches that have been designed to expedite advances in the trends.

Consolidation

The first trend to be considered is consolidation. Datacenters that are distributed and composed of heterogeneous, complex technologies tend to be costly and difficult to manage. In addition, decentralized datacenters can result in slow service provisioning, high maintenance costs, and inconsistent IT policies. Because of this, enterprises are tending to consolidate geographically dispersed datacenter resources into fewer, centralized locations. Moving such resources as application and computation servers, storage devices and networking equipment to fewer, larger physical locations creates an economy of scale and a potential money saving uniformity in resource management. There is also the potential for real estate cost reduction as the resource density is optimized when consolidation efforts include the deployment of fewer servers that are larger and more powerful. This effort also brings with it a beneficial reduction in the number of instances of the operating systems that are running in the datacenter.

As a result of consolidating datacenter operations into fewer, large physical locations, the interconnecting network becomes a critical component of the infrastructure. The performance of the interconnect can be severely challenged as traffic loads become highly concentrated and varied in type. Figure 1 shows a simplified datacenter highlighting the varied nature of the interconnects.

Figure 1: Simplified diagram of a datacenter architecture

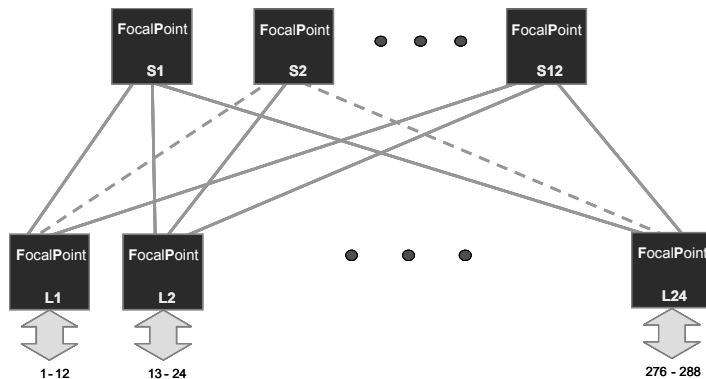


The most obvious requirement is for a switched interconnect solution that is not only high bandwidth, but that can also scale efficiently. With a greater number of applications running on more powerful servers that are connected to more storage resources, fat pipes become indispensable. To accommodate growth, new servers and more storage are inevitably added, requiring that the interconnect be able to scale efficiently. Efficient scaling implies not only scaling in terms of bandwidth, but scaling that is cost effective, easily implemented, and that allows the deployment of added bandwidth with an appropriate level of granularity.

FocalPoint switches provide this scalability with up to 24 ports of 10 Gigabit Ethernet connectivity – the most highly integrated 10 Gigabit Ethernet switch available. If even higher bandwidth connections are needed, Ethernet’s native Link Aggregation Protocol provides the ability to aggregate multiple 10 Gigabit ports into a single, higher bandwidth pipe. With FocalPoint, link aggregation can be done even across multiple switches in a multi-switch topology such as a fat tree or stack.

And when nothing but a vast amount of raw, non-blocking switch bandwidth will solve the scaling problem, FocalPoint switches provide the means for this also. With them, it is possible to build a multi-tiered, non-blocking (Clos) architecture known as a Fat Tree. The switches were designed with a set of features optimized for the effective implementation of this architecture, where a 2-tier system provides up to 288 user ports (as shown in Figure 2) and a 3-tier system up to 3,456 ports. With the use of the highly integrated FocalPoint switches deployed in this architecture and using standard Ethernet switching throughout the fabric, scalability is achieved that is more cost-, space- and power-effective than the standard “Big Iron” approach.

Figure 2: A two-tier Clos architecture with 24-port FocalPoint switches.



Related to the highly concentrated nature of consolidated datacenters, the switches in them (along with the other elements of the system) must exhibit low power consumption. With a great deal of power being dissipated in a relatively small footprint, heat removal has become a major issue for datacenters. FocalPoint switches dissipate only 1 Watt per port under aggressive use conditions, another factor allowing the efficient scaling of the switch solution along with the rest of the datacenter.

Part of the difficulty in scaling switch topologies lies in scaling the control plane along with it. Generally, each switch in a multi-switch topology is managed by its own control processor, with all the cost, board space and programming complexity that goes along with that. Through the Inter-Switch Link (ISL) tagging mechanism FocalPoint switches provide a method for in-band management, where switch management commands are

encapsulated within Ethernet frames. In this scenario, a management CPU communicates directly with a single switch. Other switches receive their management commands through the 10G links connecting them with that single switch.

Finally, the dense array of storage, compute and application servers will generate different types of traffic with different delivery priorities and tolerances for packet jitter and packet loss. What is required of the interconnect here is a combination of traffic classification, traffic management, priority based queue management and congestion management. With these mechanisms FocalPoint switches deal with the different traffic requirements effectively, providing lossless transmission for storage traffic, low latency to compute or transaction oriented traffic where required and best effort transmission for the less sensitive data traffic.

Virtualization

The virtualization of datacenters has already become a well established practice. The term “virtualization” is getting a lot of press these days, but looking just a bit deeper, there are actually at least two distinct types of virtualization taking place in the datacenter.

The first type is server virtualization, where a software abstraction layer is used to divide a single server into multiple independent environments or partitions. In this way the server resources – the processors and operating systems - are masked from the applications and users. As a consequence, virtualization allows server resources to be allocated more efficiently. Instead of a single server dedicated to each application, multiple instances of an operating system allow multiple applications to share the same physical server. The result here is more efficient use of server resources, a reduction in server sprawl and improved manageability.

Storage virtualization is the second type. With it, a multitude of storage resources are pooled into what appears to be a single, centrally managed storage device. The virtualized storage system presents to the user a logical space for data storage and then maps it to the actual physical storage location. The results here are similar to those for server virtualization: improved utilization of resources, improved manageability, and the ability to migrate data, where data may be moved or replicated to another physical location with the abstraction layer providing the isolation required to avoid disruptions of running processes.

Both of these types of virtualization entail the use of physical resources by multiple users and applications. The increase in utilization of servers and storage elements means greater demand on system uptime and makes resiliency a more pressing issue. The resiliency and fast failover characteristics of Fat Tree and stacking topologies that can be built with FocalPoint devices will play an important role in facilitating highly resilient virtualized datacenters.

Virtualized server and storage resources require some degree of network partitioning to maintain separation between groups of users. In a sense, virtual networks need to be created, allowing user groups to maintain independent communications and shared security policies, while communications between such virtual networks occurs in a highly controlled manner. FocalPoint switches enable effective network partitioning through extensive 802.1Q Virtual LAN (VLAN) and stacked VLAN (“Q-in-Q”) support. This IEEE standard protocol enables the partitioning of a single physical switch into two or more virtual switches, and conversely to create logical workgroups that span two or more physical switches. FocalPoint switches support the creation of up to 4096 simultaneous VLAN’s.

Network partitioning can also be implemented through the segmentation of memory buffer space, essentially creating two smaller switch domains with a single switch. Communications between VLAN's occur via Layer 3 routing, and FocalPoint FM4000 devices provide up to 16,000 IPv4 routes and up to 4,000 IPv6 routes for a robust, highly scalable routing domain.

To create the security policies within virtual networks, 802.1x Network Access Control is fully supported by all FocalPoint switches. Policies for communications between virtual networks can be effectively enforced using Access Control Lists (ACL's) based on layer 2, 3 or 4 packet header information.

Convergence

The last of the three datacenter trends to be discussed is fabric convergence. With local networking (LAN), storage networking (SAN) and computation or inter-processor (IPC) traffic circulating in the datacenter, several different fabric protocols generally co-exist. The system administrator may have to support three different protocols - standard IP/Ethernet for network traffic, Fibre Channel for storage traffic and Infiniband or other proprietary protocol for IPC traffic. Supporting three heterogeneous fabrics causes management complexities, increased costs and replication of equipment.

It would be highly advantageous to only support one fabric, and the most desirable of the three would be IP/Ethernet due to its ubiquity, low cost, proven scalability and massive ecosystem. Ethernet, however, was conceived as a local area networking protocol and has lacked the performance and a number of features that made development of alternative protocols necessary in the first place. For example, Ethernet's historically high latency has made it the choice of only the lower performance IPC clustering applications whereas Infiniband and Myrinet, with end-to-end latencies of only a few microseconds, are most often chosen for high performance applications. In the storage area, Ethernet and IP, being best effort protocols and lacking sufficient congestion management, could not support the lossless transmission of data required for effective file and block transfers.

This situation has not escaped the attention of various standards bodies, industry trade associations and individual equipment and component vendors. A number of standards initiatives and key component performance enhancements are happening now that are rapidly paving the way for Ethernet to be the convergence fabric of choice. For example, the development of iSCSI and FCoE provide parallel paths for the use of Ethernet in storage networks, as does the IEEE's efforts to standardize on congestion management mechanisms. NIC vendors are reducing the latency and CPU overhead that has until now crippled the performance of high speed Ethernet endpoints. Phy vendors now offer a complete suite of optical and copper physical interconnects, giving systems designers a broad array of connectivity options.

FocalPoint switches make an important contribution to the advancement of Ethernet as the converged fabric. As the central nervous system of the datacenter, the switch/router network must effectively deal with all three types of traffic at once, and be able to distinguish between them, allocating its resources intelligently among them. The switches provide eight priority-based classes per port, on which are based traffic management and queue management algorithms, all enabling the administrator to finely tune the transmission characteristics of each type of traffic. Advanced scheduling, shaping and policing algorithms allow the administrator to monitor and enforce bandwidth allocation among different classes of users.

The switches also offer advanced flow control and congestion notification mechanisms, providing the means for creating lossless data transmission for those classes of traffic,

White Paper: Datacenter Trends

such as storage, where it is required. Ethernet's standard PAUSE based flow control pauses all incoming traffic, regardless of traffic class. The result is often congestion spreading rather than congestion relief, and the possible pausing of high priority traffic due to a high volume of low priority traffic. For Ethernet to be a serious candidate for the single datacenter fabric, that situation must be rectified. FocalPoint's class-based PAUSE mechanisms provide the answer. With class-based PAUSE, the only the offending priority of traffic is paused, allowing other classes to proceed unaffected.

Finally the switches exhibit an extremely low latency – only 200 nanoseconds for layer 2 switching and 300 nanoseconds for layer 3 routing. These latencies place Ethernet in the same class as the best proprietary switch fabrics, namely Infiniband and Myrinet.

Conclusion

The triad of significant trends in the datacenter – consolidation, virtualization and convergence – is helping IT administrators cope with the rapidly escalating demand on server, storage and networking resources. As geographically dispersed datacenters are consolidated into centralized locations, resource density increases and overall flexibility and manageability is enhanced while costs are reduced. The vast numbers of collocated resources in such consolidated datacenters form fertile grounds for virtualization. With virtualization, resource usage is optimized as they are formed into a pool of shared, load-balanced hardware or a single resource such as a server is shared among several applications. With compute, storage, application and networking traffic all coexisting in a highly concentrated datacenter, the need for interconnect convergence becomes especially acute. Ethernet is by far the most preferable candidate for fabric convergence and Fulcrum's FocalPoint switches were designed with the features and performance to catalyze this convergence, as well as to enhance the benefits of virtualization and consolidation.