

Storage Fabrics using the Intel[®] Ethernet Switch Family

Reduces Data Center Cost by Converging Fabric Resources

White Paper

June, 2008



Legal

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Intel may make changes to specifications and product descriptions at any time, without notice.

Intel Corporation may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

The Controller may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel and Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2011. Intel Corporation. All Rights Reserved.

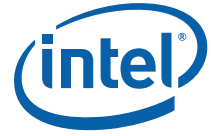


Table of Contents

Introduction	4
Traditional Data Center Storage	4
iSCSI in the Data Center.....	5
Fibre Channel over IP (FCIP)	6
Fibre Channel over Ethernet (FCoE)	6
Data Center Storage using the Intel® Ethernet Switch Family	6
Intel® Ethernet Switch Family Latency and Congestion Management	7
Scalable Fabric	8
Fabric Ports	8
Quality of Service	8
High Availability	9
Management	9
Conclusion	9

Introduction

Enterprise data centers have traditionally used Fibre Channel as a dedicated storage fabric. With the recent appearance of 10G Ethernet switch silicon and iSCSI NICs, system designers are starting to develop converged Ethernet fabrics that can include network, storage and HPC traffic. To make this possible, the Ethernet switch fabric must provide features such as quality of service and congestion management. This paper will describe how the Intel® Ethernet Switch Family 10G Ethernet switch family can address the needs of storage area networks.

Traditional Data Center Storage

There are many data center configurations in use today. Figure 1 attempts to show some of the traditional components used. Most data centers will use a subset of these components, but they are shown here for discussion purposes. Redundancy is typically used for HA which is not shown in Figure 1. Also, this paper does not discuss direct attached storage (DAS).

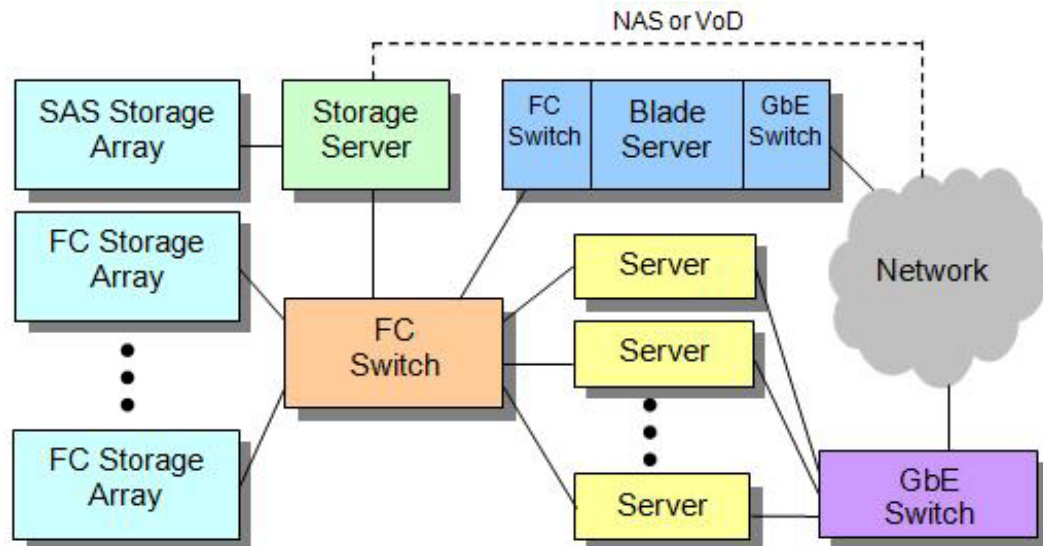
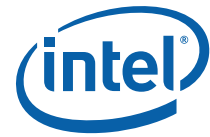


Figure 1. A Traditional Data Center Storage Network Block Diagram

There are several types of storage arrays on the market today utilizing different types of disk drives. 3.5 inch Fibre Channel (FC) drives have been traditionally used in high performance enterprise class arrays. Because of their higher rotational speeds and faster seek times, they must use wider track spacing which limits their density compared to consumer drives like SATA. The lower cost and higher density of SATA drives has opened new applications such as high-density arrays, which are replacing tape drives in near-line applications.



SAS drives are starting to replace FC drives in enterprise class arrays because they have the same performance, reliability and density as FC drives at a lower cost. For even higher performance applications, 2.5 inch SAS drives are entering the market as they provide faster rotational speeds and seek times, but at a higher cost per byte compared to 3.5 inch drives. Solid-state flash based drives are also entering the market at cost points only slightly higher than 2.5 inch drives. Although the flash drive IOPS performance may only be slightly better than SAS disk drives, raw bandwidth is much higher which is ideal for applications like video on demand (VoD).

Each storage array contains dual redundant controller cards, which connect to the array of drives on one side using some sort of switch, and provide an interface to the storage network on the other. These cards may translate from FC drives to FC network interfaces, SAS to FC, SATA to FC, SAS to SAS or SATA to SAS. Arrays for smaller enterprise applications may contain a mixture of drives (SATA, SAS, solid-state) providing a range from high density to high performance. The controller cards may also contain RAID processors to provide improved performance and reliability.

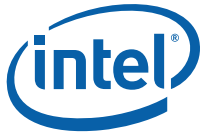
Figure 1 shows a storage server directly connected to one array, although this could also be connected to multiple arrays. A storage server typically provides specialized storage services through software. Applications can include storage virtualization, classification for tiered storage, data compression, data encryption, file to block level translation for network attached storage (NAS), virtual tape library or applications such as VoD.

Data centers traditionally use at least two separate fabrics, FC for storage and Ethernet for the data network. Storage arrays connect to the FC switch through FC controller cards. Servers use HBAs and blade servers use FC switch modules to connect to the FC network. Servers traditionally connect to the network using NICs while blade servers have an embedded GbE switch module. If High Performance Computing (HPC) is required in the data center, a third InfiniBand fabric may be used to interconnect the servers, which is not shown in **Figure 1**.

iSCSI in the Data Center

Today there is a relatively small market for iSCSI storage arrays, but the market is growing at a very high rate. One major advantage of iSCSI is the ability to create a storage network using traditional Ethernet infrastructure. This helps reduce system costs as well as cost of ownership. In addition, by using the IP layer, virtual storage networks can be spread across different geographical locations.

Factors that are critical to storage networks include bandwidth, latency and packet loss. Bandwidth must be comparable to current FC solutions and this is one reason why iSCSI has not had strong market



penetration until the recent availability of 10GbE silicon solutions. Data center performance can be reduced due to storage access latency, and packet loss or large retransmit latency cannot be tolerated. iSCSI requires a low latency 10GbE Ethernet switch with congestion management features.

Fibre Channel over IP (FCIP)

Fibre Channel over IP was developed to connect separate Fibre Channel (FC) networks across existing IP infrastructure. An example of this is the connection of two FC networks located at two different physical locations to create one logic FC network. The driving force behind this is the large installed base of FC systems and IP infrastructure.

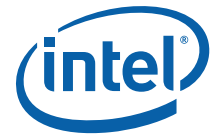
Fibre Channel over Ethernet (FCoE)

Fibre Channel over Ethernet, is being developed to replace physical FC networks in the data center with existing Ethernet fabrics. This is part of an industry initiative called Converged Enhanced Ethernet (CEE) which has a goal to route network traffic, storage traffic and HPC traffic across a single fabric such as 10G Ethernet. FCoE is effectively SCSI over FC over Ethernet.

Data Center Storage using the Intel® Ethernet Switch Family

The Intel® Ethernet Switch Family has many features that make it an attractive solution for replacing both the storage fabric and the network fabric in data center environments as shown in the example below. Here multiple iSCSI storage arrays, FCoE/FCIP storage arrays and servers connect directly to the Intel® Ethernet Switch Family switch. Arrays used for longer-term backup could be connected through the network at a distant location. Due to the Intel® Ethernet Switch Family's low latency and congestion management features, data centers requiring HPC could also replace their InfiniBand fabric with the Intel® Ethernet Switch Family switch, unifying all three fabrics into one. This is a significant savings in both up-front costs and cost of ownership and allows flexible data center deployment.

Due to the low latency of the Intel® Ethernet Switch Family, the storage server functions in [Figure 2](#) could be moved into a standard server and even be run on a virtual server. This allows flexible utilization of server resources as the data center demands change. The Intel® Ethernet Switch Family can also be implemented as a switch module in the blade server utilizing a link aggregation group (LAG) between the blade server and the rest of the storage network. This will provide low latency



storage access for the server blades. Some data center installations will have legacy FC networks. In order for these data centers to migrate into an Ethernet environment, servers or gateways may be used.

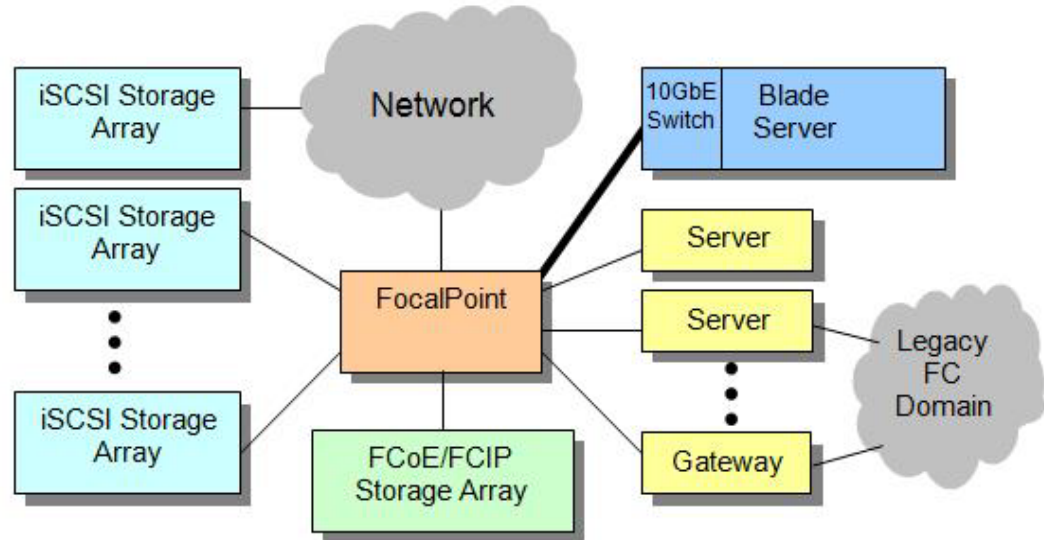


Figure 2. A Data Center Storage Network using the Intel® Ethernet Switch Family

Intel® Ethernet Switch Family Latency and Congestion Management

Both latency and congestion management are important in storage networks. In applications that require a high rate of random storage access such as database processing, storage network latency can have a significant impact on overall system performance. In traditional Ethernet switches, congestion can lead to dropped packets requiring a retransmit at the IP layer. SCSI and FC class-2, which use acknowledge time-out methods, cannot tolerate these latency spikes. The Intel® Ethernet Switch Family, with its industry leading latency and congestion management is a perfect solution for these storage environments.

With the Intel® Ethernet Switch Family, you could separate the storage, network and HPC traffic into separate virtual fabrics using shared memory partitions. By using strict priority and minimum bandwidth guarantees, maximum latency for storage traffic can be bounded. Storage data security can be enhanced using ACLs and storage server functions such as VoD could be provided with features such as minimum bandwidth or traffic shaping. Intel® Ethernet Switch Family congestion management opens the door to new services and features for storage system vendors.



Scalable Fabric

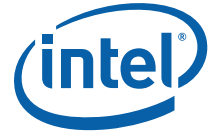
Storage fabrics must be scalable in order to meet the growing demands of an enterprise. In addition, data center costs can be dramatically reduced by converging data, storage and IPC traffic into one fabric. Ethernet is a natural choice for this converged fabric as it allows data center expansion using existing network infrastructure. Also, Ethernet's VLAN capability along with the Intel® Ethernet Switch Family's shared memory partitions, congestion management and quality of service mechanisms allow isolation of storage traffic onto a separate virtual fabric. The Intel® Ethernet Switch Family provides features such as ISL trunking to form fat trees, which along with load balancing, allows scaling to large multi-stage fabric topologies. This, combined with extremely low latency, provide the performance needed in storage protocols such as iSCSI, FCIP and FCoE.

Fabric Ports

Fibre Channel has been the choice for large data centers for many years. Due to cost reasons, IT managers are starting to move to converged Ethernet fabrics, but they must maintain the ports speeds of legacy FC products which range from 1Gb/s to 8Gb/s. Up until recently, Ethernet could only satisfy the 1Gb/s storage requirements. Now, the Intel® Ethernet Switch Family opens the full range of data rates to storage system designers from 1Gb/s to 10Gb/s per port. In addition, the Intel® Ethernet Switch Family can mix port rates on a single chip allowing the development of low-cost storage aggregation functions. The Intel® Ethernet Switch Family can also directly connect to industry standard physical interface modules, which provide hot plug mechanisms and a variety of low cost cabling options including copper and fiber.

Quality of Service

Quality of service and congestion management are key features for converged Ethernet storage fabrics. Both iSCSI and FC class-2 traffic require acknowledgement frames, which are time sensitive. These services cannot tolerate the long re-transmission times when traditional Ethernet fabrics drop packets due to periods of high congestion. The Intel® Ethernet Switch Family can guarantee lossless fabric operation with the use of class-based pause frames for link-level flow control. In addition, the Intel® Ethernet Switch Family contains up to 8 class of service queues at each output port, which can provide strict priority or minimum bandwidth guarantees along with traffic shaping. By providing minimum bandwidth guarantees to storage traffic along with the use of rate based scheduling, maximum latency can be assured.



High Availability

High availability has become a requirement for storage fabrics. This can be accomplished with the use of RAID controllers within a storage array, with the use of redundant storage arrays along with data replication and with the use of redundant paths in the storage fabric. The Intel® Ethernet Switch Family can be configured for data replication using various multicast methods. In addition, path fail-over can be implemented to route around a failed switch in the system or to bypass a failed link in a link aggregation group for N+1 protection.

Management

Storage fabric management capabilities can affect metrics such as downtime and cost of ownership. The Intel® Ethernet Switch Family provides several features which are useful for managing storage fabrics. Fulcrum in-band management frames can be used to read and write any register in the Intel® Ethernet Switch Family fabric as well as receive interrupts using only a single active (and protective) processor connected to the fabric. These frames can also be used to reset any device in the fabric. The Intel® Ethernet Switch Family supports storage zoning through the use of up to 4096 VLAN IDs and can authenticate transactions through the use of access control lists (ACLs).

Performance of the fabric can be monitored through the use of frame counters. The Intel® Ethernet Switch Family devices maintain numerous per-port and switch-wide frame counters that provide management with statistical information about the state of the switch and of the network in general. The majority of these counters are provided for compatibility with network monitoring-related IETF RFCs: 2819 (RMON), 3273 (High-Capacity RMON), 2613 (SMON), and 4502 (RMON2).

Conclusion

There is a trend in the storage industry to reduce cost by converging to a single data center fabric. 10Gb Ethernet switch silicon provides sufficient bandwidth for this converged fabric, but must provide features such as scalability, congestion management, bandwidth allocation, high availability and robust management features in order to satisfy the functional needs of storage networking. The Intel® Ethernet Switch Family can provide these features at a much lower cost point than competing fabric technologies.





NOTE: *This page intentionally left blank.*