DATA CENTER

Metro Cloud Connectivity: Integrated Metro SAN Connectivity in 16 Gbps Switches

BROCADE

CONTENTS

INTRODUCTION

As customers look to Fibre Channel (FC) storage area networks (SANs) for building private storage cloud services for their enterprise data centers, some of the key attributes are:

- Consolidated and highly virtualized pools of compute, storage and network resources
- Secure and efficient use of inter-fabric connectivity
- Lower capital and operational costs, higher asset utilization
- On-demand and efficient provisioning of application resources through automated management

Brocade has developed solutions to address these key attributes leveraging its seventh-generation “Condor3” ASIC features in the 16-Gbps platform, Fabric Operating System (FOS) v7.0, Brocade Network Advisor, Brocade HBA and CNA technology, and are working with transceiver vendors to address these key requirements. In order to enable customers to achieve these goals, Brocade is delivering key technologies (see Figure 1) that would allow customers to:

- Scale up/out based on business growth
- Secure data and optimize inter-data center bandwidth utilization
- Reduce CapEx and OpEx cost with build in diagnostics and SAN management tools
- Optimize both bandwidth and IOPS while being energy efficient

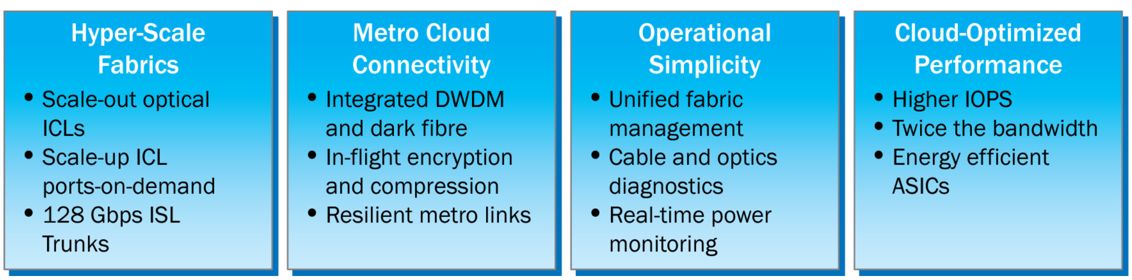

Figure 1. Enabling Private Storage Clouds

OVERVIEW

As the Fibre Channel (FC) network architecture enabled the deployment of the Storage Area Network (SAN) within IT infrastructures beginning in the mid-1990s, it also became the catalyzing point in the evolution of the dynamic data center. As the connection between server and storage changed from a direct connection to a network connection, IT architects began building what are understood today as cloud data center infrastructures. Specifically, with the introduction of FC and SANs, IT architects began building what we know today as the storage cloud.

Almost from the initial deployments of the SAN in the data center, there has been a clear understanding of the need to protect this critical infrastructure. Following the basic high-availability precept of designing a system with no single point of failure, IT architects realized that regardless of how well designed, the local SAN infrastructure would eventually require protection from a total site failure. This realization led to the development and implementation of site-to-site SAN infrastructures, alternately referred to as SAN extension, wide-area SAN, or metro-wide-area SAN, networks.

Today there are two general methods for extending the reach of a SAN from one physical site to another. The first method is by transporting the native FC protocol, without any kind of protocol conversion, over a physical link between sites, through some type of fiber-optic connectivity. The second method transports the FC protocol by encapsulating it within a transport protocol, which is then sent over the physical link between sites. The most common implementation of this method utilizes the Fibre Channel over IP (FCIP) protocol to transport FC frames by encapsulating them within an IP frame, which is then sent over a standard IP network that links both sites. The main difference between these two methods is that the native FC method offers better performance but at shorter distances, while the encapsulation method offers longer distances but at lower performance rates. Both methods have their place in today's wide-area storage cloud infrastructures.

BROCADE SEVENTH-GENERATION SAN—METRO CONNECTIVITY FEATURES

Since delivering the first FC SAN switch in the industry in March 1997, Brocade® has been a leader in developing SAN extension solutions to meet the distance connectivity needs of IT SAN architects, utilizing native as well as encapsulated extension solutions. However, with the introduction of the Brocade seventh-generation 16 Gbps FC SAN switching solutions, Brocade has again raised the bar for SAN distance extension solutions.

Powered by the seventh generation of the Brocade engineered FC, Application-Specific Integrated Circuit (ASIC) internally referred to as Condor3, this new class of Brocade SAN switches support a number of new features that will allow IT architects to build larger, more reliable, wide-area “stretched” SAN storage clouds.

With the new class of Condor3 based switches, Brocade has introduced a number of ASIC integrated extension capabilities that will allow IT architects to design and implement a variety of metro-area SAN extension solutions without requiring additional hardware. The metro area is defined as the distance from site to site of no more than 100 km, (62 miles), and with a round-trip time (RTT) latency of no more than 5 milliseconds. In the event that SAN extension solutions are required over greater distances, with longer RTT times, Brocade provides dedicated SAN extension solutions that will satisfy the most rigorous of these long distance requirements.

The new seventh-generation 16 Gbps Condor3 metro connectivity features include;:

- **16 Gbps Native Fibre Channel Long Distance Support:** In addition to doubling the overall throughput, the new generation of Condor3 based switches will be able to utilize a buffer credit pool of 8,192 buffers, which quadruples the Brocade 8 Gbps Condor2 buffer credit pool of 2,048 buffers. It also supports a variety of distance extension architectures, utilizing native Fibre Channel connectivity.
- **Integrated 10 Gbps Fibre Channel Speed Support:** The Brocade Condor3 based switches support port operating speeds of not only 16, 8, 4, and 2 Gbps, but they also support a port operating speed of 10 Gbps. In addition to native fiber, operating the port at 10 Gbps is also supported over Wave Division Multiplexing (WDM) solutions, such as Dense Wave Division Multiplexing (DWDM) in the initial release and Coarse Wave Division Multiplexing (CWDM) in a future Brocade Fabric OS (FOS) update.

- **Integrated Inter-Switch Link (ISL) Compression:** The Brocade Condor3 based switches provide the capability to compress all data in flight, over an ISL. This requires a Brocade Condor3 based switch on both sides of the ISL, and a maximum of 4 ports per Brocade DCX® 8510 blade, or 2 ports per Brocade 6510 switch can be utilized for this data compression. The compression rate is typically 2:1.
- **Integrated Inter-Switch Link (ISL) Encryption:** The Brocade Condor3 based switches provide the capability to encrypt all data in flight, over an ISL. This requires a Brocade Condor3 based switch on both sides of the ISL, and a maximum of 4 ports per Brocade DCX 8510 blade, or 2 ports per Brocade 6510 switch can be utilized for this data encryption. Both encryption and compression can be enabled on the ISL link simultaneously.
- **Enhanced Diagnostic and Error Recovery Technology:** The Brocade Condor3 based switches incorporate a new class of diagnostic and recovery features that will enable smooth operation of metro-connected SAN clouds. These features include the ability to measure, verify, and saturate the ISL links between switch E_ports, utilizing a brand new Diagnostic Port feature. Additionally, the Condor3 switches are able to detect and recover from buffer credit loss situations, in some cases without traffic disruption. Finally, SAN architects have the ability to enable Forward Error Correction (FEC) on ISL E_Ports in order to improve error detection and correction.

16 GBPS NATIVE FIBRE CHANNEL LONG DISTANCE SUPPORT

Fibre Channel Buffer-to-Buffer Flow Control

Fibre Channel network's flow control mechanism is described as a buffer-to-buffer credit system in which, in order to transmit a frame, the sending port must know beforehand that there is a buffer that is available at the receiving port. This forward flow control model is one of the key technologies that ensure that no frames are ever dropped or lost in a normally operating Fibre Channel network.

The mechanism that allows an FC port to know, ahead of time, how many frames they can send is described as a buffer credit pool. This information is determined upon port initialization, when port pairs exchange information regarding how many buffer credits each port has available and how many buffer credits each port requires.

Once the port knows this information, it can begin sending frames. The key transaction is that when a port sends a frame, it decrements its buffer credit pool by one. When the receiving port receives the frame, it sends an acknowledgement primitive, called Receiver Ready or R_RDY back to the sending port, which then allows it to increment its buffer credit pool by one.

This way, frames can be sent over very complex FC network architectures with guaranteed forward flow control and ensure lossless delivery. Furthermore, in the Brocade FC architecture, if congestion does develop in the network, because of a multiplexed ISL traffic architecture, which Brocade refers to as Virtual Channels (VC), all sending traffic will slow down gradually in response to the congestion. Without an architecture that provides fair access to the wire for all traffic streams, congestion could impose a condition where one traffic stream blocks other streams, generally referred to as Head-of-line blocking (HOL blocking).

Buffer-to-buffer credits are generally allocated and managed transparently within the FC SAN fabric. In most cases, architects do not need to overtly manage them. However, when architecting a SAN design that incorporates a long-distance link (>500 meters), it is important to understand the buffer credit requirements and allocations that are available across these long-distance links.

The challenge with long distance links is that even travelling at the speed of light, frames take time to get from one end of an optical cable to the other. If the cable is long enough and the link speed is fast enough, the result is a situation where there are multiple frames in transit on the wire at any point in time. With each frame that is sent, the sending port is decreasing its buffer credit pool by one and, if this goes on long enough without receiving the R_RDY primitives returning across the long link, its buffer credits will reach zero and it will be forced to wait for its buffer credits to be replenished before being able to begin sending frames again.

In the case of long-distance FC links, the sending port needs to have enough buffer credits available in order to fill the link with as many frames as is required so that this condition does not occur. The SAN architect must ensure that there are enough credits to keep the ISLs “full” at all times.

As a result, SAN architects understand that the relationship between the length of the link, the speed of the link, and the frame size that is being transmitted across the link will determine the correct number of buffer credits that are required for the link. A general formula for determining the correct number of buffer credits that are required for a given port-to-port link in an FC network is as follows:

(link speed in Gbps * distance in kilometers) / frame size in kilobytes = the number of buffer credits that are required

This means that for a 16 Gbps link of 500 km moving frames of 2KB in size would require;

(16 * 500) / 2 = 4,000

or approximately 4,000 buffer credits to be available to the ports on both sides of the extended SAN link in order to operate at full speed.

The problematic parameter in this equation is, of course, the frame size. Fibre Channel defines a variable length frame consisting of 36 bytes of overhead and up to 2,112 bytes of payload for a total maximum size of 2,148 bytes. Devices such as host bus adapters (HBAs) and storage arrays negotiate to 2KB frame sizes for payload, and while this means that a majority of frames are full size (2KB), lower frame sizes are used for various Small Computer Systems Interface (SCSI) commands and FC-class F traffic (such as zoning updates, Registered State Change Notifications [RSCNs], and name server information). In many instances, SAN architects typically assume the maximum 2KB Fibre Channel frame size for these buffer credit calculations. However, if the actual frame size in flight is only, for example, 1KB, the amount of buffer credits that are required is doubled. A more accurate parameter for these formulas would be the average frame size.

It is important to understand that traffic can still flow in an FC network when insufficient buffers are available—they will just operate at slower line rates. This condition is known as buffer credit starvation and will still allow traffic to flow as long as buffer credits are not lost. If, in fact, a buffer credit pool reaches zero, the port can no longer send frames until the credits are replenished or until the link is reset. So, using our earlier example, if only 2,000 buffer credits were available, our 16 Gbps 500 km link with an average frame size of 2KB would be capped at an 8 Gbps line rate.

It is also important to understand that, assuming an average 2KB frame size, if our example ISL of 16 Gbps at 500 km is assigned 4,500 buffer credits (more than the required 4,000), there is no performance improvement. In other words, assigning more credits will not yield better line-rate performance.

In the Brocade FC SAN fabric, in order to enable advanced buffer credit configurations, the ports must be configured in one of several long-distance modes, two of which are enabled via the Extended Fabrics license. They are as follows:

Brocade Fabric Long-Distance Modes

Distance Mode	**Distance**	**License Required**
L0	Local Data Center	No
LE	10 km	No
LD	Auto-Discovery	Yes
LS	Static Assignment	Yes

L0 designates local or “normal” buffer credit allocations. This is the default port configuration and, as previously mentioned, it provides sufficient buffer credits within the normal data center link distances (less than 500 meters). LE is used to support distance up to 10 km and does not require a license. The 10 km limit is not dependent on speed because if the ports negotiate to higher speeds, more credits are automatically assigned to support the higher line speed.

LD (dynamic distance discovery mode) is the most user friendly mode. It automatically probes the link and, via a sophisticated algorithm, calculates the amount of credits that are required, based on the distance and speed set for the link.

LS is a statically configured mode that was added to Brocade FOS 5.1. This is the most flexible mode for the advanced user. It allows complete control of buffer credit assignments, for long distance requirements.

16 Gbps Native Fibre Channel Long-Distance Support

With the introduction of the Brocade seventh-generation ASIC, Condor3, Brocade has significantly expanded the buffer architecture that is available for SAN storage cloud architects.

First and foremost, the Condor3 ASIC provides a massive 8,192 buffers. This is a four-fold increase over the 2,048 buffers that are available with the 8 Gbps Condor2 ASIC. Furthermore, the Condor3 ASIC architecture has the ability to link to the buffer pools of other Condor3 ASICs. This feature will be available in a future Brocade FOS update and will be enabled by configuring an ASIC port as a linking credit (LC) port, which will then be used expressly between ASICs for the purposes of sharing buffers.

This feature will be available on Brocade switch architectures that utilize multiple Condor3 ASICs internally, such as the Brocade DCX 8510 Family. Once enabled, this buffer linking capability will be able to provide a single Condor3 ASIC in a Brocade FC SAN fabric a total pool of 13,000 buffers. This will allow extended SAN architectures to support not only a wide variety of link distances, but also a wide variety of link speeds and average frame sizes.

In order to calculate the maximum amount of buffer credits that are available, it is important to note that the Brocade FC ASIC architecture always reserves buffers out of the total amount available, that are not part of the pool available for long-distance configuration. For example, the Condor3 ASIC will reserve 8 buffers for every front-end Fibre Channel port, 48 buffers for communicating with the switch control processor, and 48 buffers for additional internal functions. Additional buffers are also reserved internally in multi-ASIC switches, such as the Brocade DCX bladed backbone switches, in order to support the internal communications between port blade and core blade. The following table provides the total number of reserved and available buffers, per Condor3 ASIC, within the specific DCX port blade or standalone switch.

Condor3 Switch Type	Total Reserved Buffers	Total Available Buffers
Brocade 6510 Switch	480	7,712
DCX FC16-32 Port Blade	2,784	5,408
DCX FC16-48 Port Blade	3,232	4,960

Using the buffer credit formula that was provided earlier, and assuming an average frame size of 2KB, the specific long-distance link distance support of the Condor3 ASIC based switches are illustrated in the following table.

Link Speed	Brocade 6510	DCX FC16-32 Port Blade		DCX FC16-48 Port Blade	
		Single ASIC	With Credit Linking*	Single ASIC	With Credit Linking*
2 Gbps	7,712 km	5,408 km	10,528 km	4,960 km	10,080 km
4 Gbps	3,856 km	2,704 km	5,264 km	2,480 km	5,040 km
8 Gbps	1,928 km	1,352 km	2,632 km	1,240 km	2,520 km
16 Gbps	964 km	676 km	1,316 km	620 km	1,260 km
* Enabled in a future Brocade FOS update					

The Condor3 long-distance buffer architecture supports distances that are far in excess of what the current Small Form- Factor Pluggable (SFP) optical technology can support, (i.e. light), today. So while these calculations offer dramatic example of the expanded capability of Condor3 extended distance support, the practical benefit of this feature could mean that instead of an increase in distance between any two SANs, IT architects would have the ability to increase the number of SANs that are connected within a metro wide area. This would allow SAN architects who are utilizing Condor3 based switches to design larger, classic “hub and spoke” metro-wide-area SAN storage clouds.

Lastly, because optical network distances must be designed with the appropriate port optics technology (as pointed out earlier), Brocade provides both Short Wave Laser (SWL), as well as Long Wave Laser (LWL), 16 Gbps Small Form-Factor Pluggable Plus (SFP+) optics for the new Condor3 based switches. These LWL SFP+ optics support distances of up to 10 km, and Brocade will extend these offerings in the future by providing 16 Gbps LWL SFP+ optics with longer distance (>10 km) support.

INTEGRATED 10 GBPS FIBRE CHANNEL SPEED SUPPORT

The Evolution of Optical Network Standards

One of the remarkable achievements of the Fibre Channel network standards was that it introduced a network architecture that could evolve through successive generations without imposing significant infrastructure upgrade costs. This is the reason that, once the optical network industry switched from the larger 1 Gbps Gigabit Interface Converter (GBIC) pluggable optics to the smaller SFP optics technology, the FC network upgrade from 2 Gbps to 4 Gbps, and finally to 8 Gbps, was achieved fairly easily and cost-efficiently.

Each successive advance in speed meant that, while distances were shortened for multi-mode fiber-optic cabling (2 Gbps supporting a maximum distance of 500 meters, 4 Gbps at 380 meters, and finally 8 Gbps at 150 meters), as long as the distance limits were sufficient, for the first time, an IT architect could double the speed of their network without changing the wiring plant. Additionally, because the same conversion layer encoding is used (the 8b/10b scheme), the SFP optics technology required only small changes to support the increasing line rates.

This was, in fact, a goal of the ANSI T11.3 subcommittee, which is the standards group responsible for the main body of Fibre Channel network standards. However, changes were introduced into the network conversion layer when the network speed advanced to 10 Gbps, which required changes, primarily to the optics technology that was deployed.

The reasons for this are that, with the advent of high-speed optical networking, the relevant network standards organizations made a concerted effort to harmonize the physical and conversion layer standards, where possible, for different network architectures so that IT architects could benefit from a physical layer that could support multiple network architectures. This meant that the ANSI T11.3 subcommittee and the IEEE 802.3 working group (the standards group that is responsible for the main body of Ethernet standards) settled on the same optical physical and conversion layer standards developed in the T11.3 subcommittee for 1 Gbps FC, specifically standard FC-1. This was mirrored in the IEEE 802.3 Clause 38 Physical Coding Sublayer (PCS) standard. Additionally, because Ethernet did not use the interim 2, 4, and 8 Gbps speeds, the T11.3 subcommittee further developed the physical layer standards for 2, 4, and 8 Gbps FC as extensions to the original FC-1 standard. This enabled the “easy to upgrade” FC network architecture that exists today.

However, when network speeds advanced to 10 Gbps, the 802.3 group was first to define new conversion layer standards, which were different than previous standards, and resulted in the IEEE 802.3 Clause 49 Physical Coding Sublayer (PCS) standard which changed the conversion layer encoding scheme from the 8b/10b scheme to the 64b/66b scheme. This change required new optical technology and standards that, initially resulted in a new, larger, 10 gigabit small form factor pluggable optical module, referred to as an XFP. The ANSI T11.3 subcommittee also defined the 10 Gbps FC conversion layer standards as utilizing the new 64b/66b encoding conversion scheme. The T11.3 subcommittee further developed new optics standards that provided better backwards compatibility with existing SFP technology, which resulted in the development of the SFP+ standards. The resulting SFP+ optics utilizes the same form factor as earlier SFP optics, and draw less power than XFP optical modules. Today, SFP+ is the most popular optical socket technology deployed for 10 Gbps Ethernet as well as for Fibre Channel.

Because of the changes in conversion layer, a noticeable shift occurred in Fibre Channel network architectures with the segregation of 10 Gbps ports and their related technology from the legacy 1, 2, 4, and 8 Gbps FC ports. Some switch vendors begin building dedicated 10 Gbps FC ports, which could typically only be used as ISL ports. However, if the port was not needed, customers still ended up paying for the port capacity, forcing a “stranded capacity” situation. Brocade developed a special 10 Gbps blade, called the FC10-6, for the Brocade 48000 4 Gbps Director, which provided 6-10 Gbps FC ports, for use as ISL ports between similar FC10-6-equipped Brocade 48000 Directors. The FC10-6 blade design incorporated two sets of FC ASICs—the Condor 4 Gbps ASIC and the Egret 10 Gbps ASIC.

10 Gbps Fibre Channel Speed Support

The ANSI T11.3 subcommittee also defined the conversion layer encoding scheme for 16 Gbps and 32 Gbps FC speeds to utilize the 64b/66b encoding as well. Additionally, SFP+ optical technology can be utilized for not only 10 Gbps line rates, but also for higher speed line rates, such as 16 Gbps Fibre Channel and the new Brocade Condor3 based switches.

However, Brocade developed a further integrated enhancement, which can provide the SAN architect with new capabilities in designing a metro-wide-area SAN. The Condor3 ASIC, in addition to supporting FC line rates of 2, 4, 8, and 16 Gbps, will also support FC line rates of 10 Gbps. More specifically, it can do this without specialized hardware and without forcing “stranded capacity.”

In comparison to long-distance fiber-optic links between Brocade Condor3 based switches, which can run natively at 16 Gbps, the ability to run ports at 10 Gbps might not seem like a benefit. However, in the event that the physical link between SANs is provided through alternate service providers, this capability can allow SAN architects the required flexibility in designing a metro-area SAN architecture by providing compatibility with other wide-area network technology.

Today, IT architects can link SANs in a metro area, for native FC protocol transmission, either by directly utilizing a fiber-optic cable between sites or by creating multiple channels on an optical cable between sites, utilizing WDM technology. WDM is a technique for providing multiple channels across a single strand of fiber-optic cable. This is done by sending the optical signals at different wavelengths of light, also called lambda circuits. The two most common WDM technologies are DWDM and CWDM. The main benefits of deploying WDM technology is that you can increase the amount of traffic over a single optical cable as well as increase the types of traffic, going over a single optical cable.

Additionally, both types of metro-area SAN connectivity links, either direct cable, or WDM, can be deployed directly, or can be purchased as a service, from a service provider. Even in the case where an IT architect either outright owns, or leases, the entire fiber optic cable that links two data center sites together, competing interests within the organization might require dividing the cables into multiple channels with WDM technology.

The value in being able to drive port speeds on the Brocade Condor3 based switches at the 10 Gbps rate is because most WDM technology does not currently support 16 Gbps rates. Rather than having to throttle down to either 8 Gbps or 4 Gbps line rates, and waste additional lambda circuits to support required bandwidth, the new Brocade Condor3 switches can drive a given lambda circuit at a 10 Gbps line rate, optimizing the link. Brocade has successfully tested this configuration with DWDM solutions from Adva, in the form of the Adva Optical FSP 3000, and Ciena, in the form of the Ciena ActivSpan 4200. Brocade will continue to test additional DWDM solutions in the future, in order to ensure compatibility with a wide variety of DWDM technology providers. Brocade will also test CWDM solutions in the future with this configuration, so that SAN architects will be able to utilize either DWDM or CWDM solutions in their metro-area SAN architectures.

The actual configuration of the 10 Gbps FC line rate on the Condor3-based switches is done by configuring the speed for an 8-port group, called an octet. These are the octet speed combination options that are available on the Brocade Condor3 based switches:

Speed Mode	Port Speeds Supported
1	16Gbps, 8Gbps, 4Gbps, 2Gbps
2	10Gbps, 8Gbps, 4Gbps, 2Gbps
3	16Gbps, 10Gbps

The default speed mode is 1, which means any port in the 8-port group octet can operate at either 16, 8, or 4 Gbps, utilizing 16 Gbps SFP+ optics, or at 8, 4, or 2 Gbps, utilizing 8 Gbps SFP+ optics. Speed combination modes 2 and 3 enable any port in the octet to operate at a 10 Gbps line rate, but also specifically requires 10 Gbps SFP+ optics. These are also available in SWL as well as LWL models.

Note that the changing of the octet speed mode is a disruptive event and will be initially supported only on the first 8 ports on any blade in the Brocade DCX 8510 (4-slot and 8-slot) and the first 8 ports on the Brocade 6510 switch. The maximum configuration supported will provide 64 ports of 10 Gbps across all 8 port blades on the Brocade DCX 8510-8, 32 ports of 10 Gbps across all 4 port blades on the Brocade DCX 8510-4, or 8 ports of 10 Gbps on the Brocade 6510 switch. A future FOS update will remove the limitation of only being able to select the first 8 ports (octet) on the Brocade DCX 8510 blade or the Brocade 6510 switch for 10 Gbps speed.

Implementing the 10 Gbps FC line-rate feature does not require any additional hardware, but does require that an Extended Fabrics license be enabled on both switches. Additionally, the 10 Gbps FC line-rate capability of the Brocade Condor3 based switches is not compatible with the prior Brocade FC10-6 10 Gbps FC port blade. This means that in order to establish a 10 Gbps ISL between sites, both sites must be connected utilizing the Brocade Condor3 based switches.

Integrated Inter-Switch Link (ISL) Compression

The Brocade Condor3 based FC switches introduce a new capability for metro-area SAN architects: ASIC-integrated, Inter-Switch Link, (ISL), in-flight, data compression. Each Condor3 ASIC can provide up to 32 Gbps of compression, via a maximum of 2 - 16 Gbps FC ports, which can be combined and load-balanced, utilizing Brocade ISL Trunking.

Because the Brocade DCX 32-port and 48-port 16 Gbps port blades are equipped with two Condor3 ASICs, a single port blade in the Brocade DCX 8510 can provide up to 64 Gbps of ISL data compression, utilizing four ports. The maximum DCX configuration supported will provide 512 Gbps of compression across all 8 port blades in the Brocade DCX 8510-8, or 256 Gbps of compression across all 4 port blades in the Brocade DCX 8510-4. The Brocade 6510 switch is limited to providing up to 32 Gbps of compression, on up to two 16 Gbps FC ports. Future enhancements will include support for compression over 10 Gbps FC links.

This compression technology is described as “in-flight” because this ASIC feature is enabled only between E_Ports, allowing ISL links to have the data compressed as it is sent from the Condor3 based switch on one side of an ISL and then decompressed as it is received by the Condor3 based switch that is connected to the other side of the ISL. As mentioned earlier, in-flight ISL data compression is supported across trunked ISLs, as well as multiple ISLs and long-distance ISLs. Brocade Fabric QoS parameters are also honored across these ISL configurations.

The compression technology utilized is a Brocade developed implementation that utilizes a Lempel-Ziv-Oberhumer (LZO) lossless data compression algorithm. The compression algorithm provides an average compression ratio of 2:1, and all Fibre Channel Protocol (FCP), as well as non-FCP, frames that transit the ISL are compressed. with the exception of Basic Link Services (BLS) as defined in the ANSI T11.3 FC-FS standard, and Extended Link Services (ELS) as defined in the ANSI T11.3 FC-LS standard, frames.

When enabling the in-flight ISL data compression capability, the ISL port must be configured with additional buffers, requiring the switch port to be configured in an LD mode. Enabling in-flight ISL data compression also increases the time it takes for the Condor3 ASIC to move the frame. This is described as latency and should be understood by SAN architects. Normally the transit time for a 2KB frame to move from one port to another port on a single Condor3 ASIC is approximately 700 nanoseconds, a nanosecond representing one-billionth (10^{-9}) of a second. Adding in-flight data compression increases the overall latency by approximately 5.5 microseconds, a microsecond representing one-millionth (10^{-6}) of a second.

This means there is an approximate latency time of 6.2 microseconds for a 2KB frame to move from a source port, be compressed, and then move to the destination port on a single Condor3 ASIC. Of course, calculating the total latency across an ISL link means including the latency calculations for both ends. For example, compressing a 2KB frame and sending it from one Condor3 switch to another would result in a total latency of 12.4 microseconds, (6.2 * 2), not counting the link transit time.

One of the use cases for utilizing Condor3 integrated ISL data compression is when a metro-area SAN infrastructure includes an ISL for which there are either bandwidth caps or bandwidth usage charges. Finally, implementing the Condor3 ISL compression capability requires no additional hardware and no additional licensing.

INTEGRATED INTER-SWITCH LINK (ISL) ENCRYPTION

The Brocade Condor3 based FC switches, in addition to ISL data compression, also introduce the new capability of implementing ASIC integrated Inter-Switch Link, (ISL), in-flight, data encryption. Each Condor3 ASIC can provide up to 32 Gbps of encryption, via a maximum of 2 - 16 Gbps FC ports, which can, again, be combined and load-balanced, utilizing Brocade ISL Trunking.

The two Condor3 ASICs on both the Brocade DCX 32-port and 48-port 16 Gbps port blades enable a single port blade in the Brocade DCX 8510 to provide up to 64 Gbps of ISL data encryption, utilizing four ports. The maximum DCX configuration supported will provide 512 Gbps of encryption across all 8 port blades in the Brocade DCX 8510-8, or 256 Gbps of encryption across all 4 port blades in the Brocade DCX 8510-4. The Brocade 6510 switch is limited to providing up to 32 Gbps of encryption, on up to two 16 Gbps FC ports.

As with Condor3 integrated compression, the integrated encryption is supported in-flight, exclusively for ISLs, linking Condor3 based switches. Enabling ISL encryption results in all data being encrypted as it is sent from the Condor3 based switch on one side of an ISL and then decrypted as it is received by the Condor3 based switch connected to the other side of the ISL. As with integrated ISL compression, this integrated ISL encryption capability is supported across trunked ISLs, as well as multiple ISLs and long-distance ISLs. Brocade Fabric QoS parameters are also honored across these ISL configurations. It is important to note that, when implementing ISL encryption, using multiple ISLs between the same switch pair requires that all ISLs be configured for encryption or none at all.

While the Condor3 based switches support a Federal Information Processing Standard (FIPS) mode for providing FIPS 140 Level 2 compliance, upon release the integrated ISL encryption will only work with the FIPS mode disabled. A future Brocade FOS update will allow enabling integrated ISL encryption with FIPS mode enabled. Additionally, in order to implement ISL encryption, some room is necessary within the FC frame payload area. As mentioned previously, the maximum payload size of an FC frame is 2,112 bytes, which is typically the maximum size used by most drivers. In order to support this requirement for integrated ISL encryption, Brocade will change the default frame payload size setting for the Brocade HBA driver to 2,048 bytes. In all other cases, in order to enable ISL data encryption, all other drivers will need to be manually configured to utilize a maximum payload size of 2,048 bytes.

Both compression and encryption can be enabled, utilizing the integrated features of the Brocade Condor3 based switches. As is the case with integrated data compression, enabling integrated data encryption adds approximately 5.5 microseconds to the overall latency. This means an approximate latency time of 6.2 microseconds for a 2KB frame to move from a source port, be encrypted, and then move to the destination port on a single Condor3 ASIC. Also, calculating the total latency across an ISL link means including the ASIC latency calculations for both ends. Encrypting a 2KB frame and sending it from one Condor3 switch to another would result in a total latency of 12.4 microseconds (6.2 * 2), not counting the link transit time. If both encryption and compression are enabled, those latency times are not cumulative. For example, compressing and then encrypting a 2KB frame incurs approximately 6.2 microseconds of latency on the sending Condor3 based switch and incurs approximately 6.2 microseconds of latency at the receiving Condor3 based switch in order to decrypt and uncompress the frame. This would result in a total latency time of 12.4 microseconds, again, not counting the link transit time.

The encryption method utilized for the Condor3 integrated ISL encryption is the Advanced Encryption Standard (AES) AES-256 algorithm using 256 bit keys, and uses the Galois Counter Mode (GCM) of operation. AES-GCM was developed to support high-throughput message authentication codes (MAC) for high data rate applications such as high-speed networking. In AES-GCM, the MACs are produced using special structures called Galois field multipliers, which are multipliers that use Galois field operations to produce their results. The key is that they are scalable and can be selected to match the throughput requirement of the data.

As with integrated ISL data compression, when enabling integrated ISL encryption, all FCP and non-FCP frames that transit the ISL are encrypted, with the exception of BLS and ELS frames. In order to enable integrated Condor3 ISL encryption, port-level authentication is required and Diffie-Hellman Challenge Handshake Authentication Protocol (DH-CHAP) must be enabled. The Internet Key Exchange (IKE) protocol is used for key generation and exchange. The key size is 256 bits, the Initialization Vector (IV) size is 64 bits, and the Salt size is 32 bits. Unlike traditional encryption systems that require a key management system for creating and managing the encryption keys, the integrated Condor3 ISL encryption capability is implemented with a simpler design, utilizing a non-expiring set of keys that are reused. While this represents a security concern because the keys are non-expiring and reused, it also allows this integrated ISL encryption to be implemented with very little management impact.

One use case for utilizing Condor3 integrated ISL encryption is to enable a further layer of security for a metro-area SAN infrastructure. Implementing the Condor3 ISL encryption capability also requires no additional hardware and no additional licensing.

ENHANCED DIAGNOSTIC AND ERROR RECOVERY TECHNOLOGY

The Brocade seventh-generation Condor3 based FC switches include several categories of management, diagnostic, and error recovery technologies, which are designed to enable the IT SAN architect to be able to scale, manage, diagnose, and troubleshoot larger and more complex SAN storage clouds. These new features are integrated into the Condor3 ASIC, and require no additional hardware or licenses. Some of these technologies have increased relevance when planning a metro- or wide-area SAN network infrastructure.

Forward Error Correction, (FEC)

The Brocade Condor3 ASIC includes integrated Forward Error Correction, (FEC), technology, which can be enabled only on E_Ports connecting ISLs between switches. FEC is a system of error control for data transmissions, whereby the sender adds systematically generated error-correcting code (ECC) to its transmission. This allows the receiver to detect and correct errors without the need to ask the sender for additional data.

The Brocade Condor3 implementation of FEC enables the ASIC to recover bit errors in both 16 Gbps and 10 Gbps data streams. The Condor3 FEC implementation can enable corrections of up to 11 error bits in every 2,112-bit transmission. This effectively enhances the reliability of data transmissions and is enabled by default on Condor3 E_ports.

Enabling FEC does increase the latency of FC frame transmission by approximately 400 nanoseconds, which means that the time it takes for a frame to move from a source port to a destination port on a single Condor3 ASIC with FEC enabled is approximately .7 to 1.2 microseconds. SAN administrators also have the option of disabling FEC on E-Ports.

Brocade Diagnostic Port, (D-Port)

The Brocade Condor3 based switches deliver enhanced diagnostic capabilities in the form of a new port type, called the Diagnostic Port (D_Port). The D_Port is designed to diagnose optics and cables before they are put into production. Initially supported only on E_Ports, the D_Port will be able to perform electrical as well as optical loopback tests, and will also be able to perform link-distance measurement and link saturation testing.

The D_Port diagnostic capability provides an opportunity to measure and thoroughly test ISL links before they are put into production. D_Ports can also be used to test active ISL links. However, the link must first be taken down in order to enable the D_Port configuration and tests.

Brocade 16 Gbps SFP+ optics support all D_Port tests, including loopback and link tests. The accuracy of the 16 Gbps SFP+ link measurement is within 5 meters. 10 Gbps SFP+ optics do not currently support the loopback tests, but do support the link measurement as well as link saturation tests, and provide link measurement accuracy to within 50 meters.

Buffer Credit Loss Detection and Recovery

The final category of Brocade Condor3 diagnostic and error recovery technologies is in the area of buffer credit loss detection and recovery. It should be evident by now that the management of buffer credits in wide-area SAN architectures is critically important. Furthermore, many issues can arise in the SAN network whenever there are instances of either buffer credit starvation or buffer credit loss.

Conditions where a particular link may be starved of buffer credits could include either incorrect long-distance buffer credit allocations or links where buffer credits are being lost. Lost buffer credits can be attributed to error conditions such as a faulty physical layer component or misbehaving end node devices. If this condition persists untreated, it can result in a “stuck” link condition whereby the link is left without buffer credits for an extended time period, (e.g. 600 milliseconds), stopping all communications across the link.

These problem conditions are only exacerbated when they exist in wide-area SAN architectures. The Brocade Condor3 ASIC includes a number of new features that are designed to diagnose, troubleshoot, and recover from these types of conditions.

As previously mentioned, the Brocade Fibre Channel network implements a multiplexed ISL architecture called Virtual Channels (VCs), which enables efficient utilization of E_Port to E_Port ISL links. However, in terms of being able to diagnose and troubleshoot buffer credit issues, being able to do so at the VC granularity is very important.

While the Brocade Condor2 ASIC and FOS provide the ability to detect buffer credit loss and recover buffer credits at the port level, the Brocade Condor3 ASIC diagnostic and error recovery feature set includes the following features:

- The ability to detect and recover from buffer credit loss at the VC level
- The ability to detect and recover "stuck" links at the VC level

The Brocade Condor3 based switches can actually detect buffer credit loss at the VC level of granularity and, if the ASICs detect only a single buffer credit lost, can restore the buffer credit without interrupting the ISL data flow. If the ASICs detect more than one buffer credit lost or if they detect a "stuck" VC, they can recover from the condition by resetting the link, which would require retransmission of frames that were in transit across the link at the time of the link reset.

CONCLUSION

The Brocade seventh-generation of Condor3 based FC SAN switches raises the bar on delivering the technologies and tools to build today's SAN storage cloud. The Brocade Condor3 line of FC SAN switches also provide integrated technologies that enable building metro-area SAN storage clouds more efficiently and more cost-effectively than ever before.

Brocade developed these solutions based upon the partnership that exists with our customers. Brocade developed the FC network solutions that enabled the building of the SAN storage cloud and, by implementing these solutions, our customers have shown the way forward, indicating what capabilities will be required to support the SAN storage clouds of tomorrow.

Being able to protect SAN storage clouds from single points of failure, including site failures, is a requirement that continues to evolve. As Brocade works to integrate more metro- and wide-area SAN technologies into the Brocade FC SAN fabric technology, it is our expectation that our customers will continue to guide us forward with their requirements as their SAN storage clouds evolve.