Why switched High Performance FICON for z Systems (zHPF)?

Contents

Executive summary

With the many enhancements and improvements in mainframe I/O technology, the question "Do I need zHPF switching technology, or should I go direct-attached?" is frequently asked. With up to 320 IBM® FICON® Express16S channels supported on an IBM z13™, why not just direct-attach the control units? The short answer is that with all of the z Systems™ I/O improvements, switching technology is needed now more than ever. This paper explores both the technical and business reasons for implementing a switched zHPF architecture instead of a direct-attached storage zHPF architecture.

Technology background

IBM z Systems™ delivered zHPF in 2008 on the IBM z10™ EC platform and has been strategically investing in this technology ever since the initial delivery. The raw bandwidth of a FICON Express16S channel using zHPF is 4.2 times greater than a FICON Express16S channel using original FICON capabilities. The raw I/O per second (IOPS) capacity of the FICON Express16S using zHPF is 98,000 IOPS which is 4.3 times greater than a FICON Express16S channel using default FICON capabilities. To get the full benefit of the z System's I/O subsystem, clients need to transition to zHPF.

FICON Express16S channels using default FICON on z13 processors can have up to 64 concurrent I/Os (open exchanges) to different devices. FICON Express16S channels running zHPF can have up to 750 concurrent I/Os on the z13 processor family. Only when a director or switch is used between the host and storage device can the true performance potential inherent in these channel bandwidth and I/O processing gains be fully realized.

Technical reasons for a switched zHPF architecture

Why is it a best practice to implement switched zHPF rather than use direct-attaching zHPF for connecting storage control units? There are five key technical reasons:

- To overcome buffer credit limitations on FICON Express16S channels
- To build fan-in, fan-out architecture designs for maximizing resource use
- To localize failures for improved availability
- To increase scalability and enable flexible connectivity for continued growth
- To use zHPF technologies

FICON channel buffer credits

When IBM introduced the availability of FICON Express8 and FICON Express8S channels, one important change was the number of buffer credits available on each port of the channel card was reduced to 40. Organizations familiar with buffer credits will recall that the number of buffer credits required for a given distance varies directly in a linear relationship with link speed. In other words, doubling the link speed would double the number of buffer credits required to achieve the same performance at the same distance. For the FICON Express16S channel, the number of buffer credits has been increased to 90.

The 40 buffer credits per port on a FICON Express8 or FICON Express8S channel card and the 90 buffer credits on the FICON Express16S channel card can support up to 10 km of distance for full-frame size I/Os (2 KB frames). A switched architecture allows organizations to overcome the buffer credit limitations on the FICON channel card for organizations with less than full-size frames. Depending upon the specific model, FICON directors and switches can have more than 1300 buffer credits available per port for long-distance connectivity.

Fan-in, fan-out architecture designs

In the late 1990s, the open systems world began to implement Fibre Channel storage area networks (SANs) to overcome the low use of resources inherent in a direct-attached architecture. SANs addressed this issue through the use of fan-in and fan-out storage network designs. These same principles apply to a zHPF storage network.

As a general rule, FICON Express16S channels offer better performance, in terms of IOPS or bandwidth, than the storage host adapter ports to which they are connected. Therefore, a direct-attached zHPF storage architecture will typically see very low channel use rates. To overcome this issue, fan-in and fan-out storage network designs are used.

A switched zHPF architecture allows a single channel to fan-out to multiple storage devices through switching, improving overall resource use. This can be especially valuable if an organization’s environment has newer zHPF channels, such as FICON Express16S, but older tape drive technology. Figure 1 illustrates how a single FICON channel can concurrently keep several tape drives running at full-rated speeds. The actual fan-out numbers based on tape drives will depend on the specific tape drive and control unit. It is not unusual, however, to see a FICON Express16S channel fan-out from a switch to five to six tape drives (a 1:5 or 1:6 fan-out ratio). The same principles apply for fan-out to direct-access storage device DASD arrays. The exact fan-out ratio is dependent on the DASD array model and host adapter capabilities for IOPS and/or bandwidth.

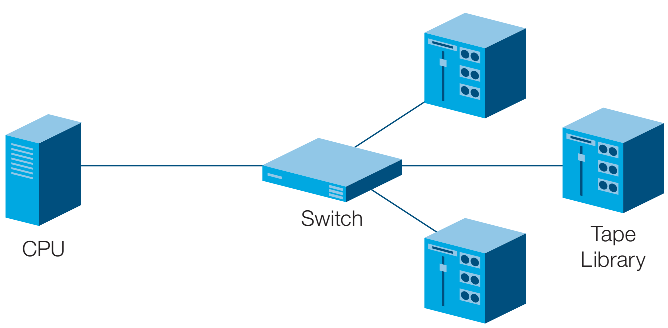

Figure 1. Switched FICON allows one channel to keep multiple tape drives fully used

Keeping failures localized

In a direct-attached architecture, a failure anywhere in the path renders both the channel interface and the control unit port inoperable. This could be the failure of an entire FICON channel card, a port on the channel card, a failure of the cable, a failure of the entire storage host adapter card, or a failure of an individual port on the storage host adapter card. In other words, a failure on any of these components will affect both the mainframe connection and the storage connection. The worst possible reliability, availability, and serviceability for zHPF-attached storage are provided by a direct-attached architecture.

With a switched architecture, failures are localized to only the affected zHPF channel interface or control unit interface. Not both. The non-failing side remains available, and if the storage side has not failed, other zHPF channels can still access that host adapter port through the switch or director (Figure 2). This failure isolation, combined with fan-in and fan-out architectures, allows the most robust storage architectures, minimizing downtime and maximizing availability.

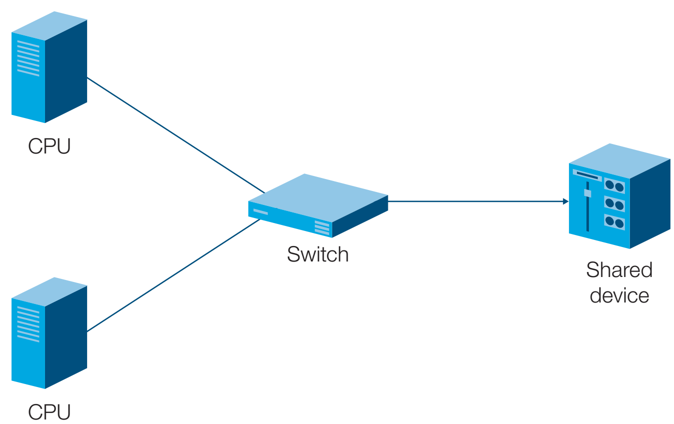

Figure 2. A FICON director isolates faults, and improves availability

Scale and flexible connectivity

Direct-attached zHPF does not easily allow for dynamic growth and scalability, since one zHPF channel card port is tied to one storage host adapter port. In such an architecture, there is a 1:1 relationship (no fan-in or fan-out), and since there is a finite number of zHPF channels available (dependent on the mainframe model or machine type (see Figure 3), growth in a mainframe storage environment can pose a problem. What happens if an organization needs more zHPF connectivity, but has run out of zHPF channels? Use of switching and proper use of fan-in and fan-out in the storage architecture design will go a long way toward improving scalability.

In addition, best-practice storage architecture designs include room for growth. With a switched zHPF architecture, adding devices such as tape is much easier. Simply connect the new control unit to the switch. This eliminates the need to open the channel cage in the mainframe to add new channel interfaces, reducing both capital and operational costs. This also gives managers more flexible planning options when upgrades are necessary, since the urgency of upgrades is lessened.

The next generation of channels

The bandwidth capabilities of channels are growing at a much faster rate than those of storage devices. As channel speeds increase, switches will allow data center managers to take advantage of new technology as it becomes available, while protecting investments and minimizing costs.

It is an IBM best-practice recommendation to use single-mode long-wave connections for zHPF channels. Storage vendors, however, often recommend using multi-mode short-wave connections on their host adapter ports—usually due to cost.

Some organizations' existing storage devices have a mix of single-mode and multi-mode connections. Since they cannot directly connect a single-mode zHPF channel to a multi-mode storage host adapter, this could pose a problem. With a zHPF director or switch in the middle, however, organizations do not need to change the storage host adapter ports to comply with the single-mode best-practice recommendation for the zHPF channels. The zHPF switching device can have both types of connectivity:

- Single-mode long-wave ports for attaching the zHPF channels
- Multi-mode short-wave ports for attaching the storage.

Mainframe	Max. number of FICON channels
z800	32 FICON Express
z900	96 FICON Express
z890	40 FICON Express/80 FICON Express2
z990	120 FICON Express/240 FICON Express2
z9BC	112 FICON Express4
z9EC	336 FICON Express4
z10EC	336 FICON Express8
z114	64 FICON Express8
z196	336 FICON Express8S
zBC12	128 FICON Express8S
zEC12	320 FICON Express8S
z13	320 FICON Express8S/320 FICON Express16S

Figure 3. Direct-attached architectures are limited by the number of channels available.

Furthermore, zHPF switching elements at two different locations can be interconnected by fiber at distances up to 100 km or more—typically used in disaster recovery and business continuance architectures. As previously discussed, zHPF switching allows resources to be shared. With cascaded zHPF switching, those resources can be shared with no performance loss between geographically separated locations, allowing data to be replicated or tape backups to be completed at the alternate site from the primary site. Often, workloads will be distributed such that both the local and remote sites are primary production sites, and each site uses the other as its backup.

While the fiber itself is relatively inexpensive, laying new fiber may require a construction project. Dense Wave Division Multiplexing (DWDM) helps get more out of the fiber, but with zHPF switching vendors now offering Inter-Switch Links (ISLs) with up to 16 Gbps of bandwidth, the expense of DWDM can be reduced, or even eliminated. zHPF switches maximize use of this valuable inter-site fiber by allowing multiple environments to share the same fiber. In addition, zHPF switching devices offer unique storage network management features—such as ISL trunking and preferred pathing—that are not available with DWDM equipment.

zHPF switches allow data center managers to further exploit inter-site fiber sharing by enabling them to intermix zHPF and original Fibre Channel Protocol (FCP) traffic—known as Protocol Intermix Mode (PIM). Even in data centers where there is enough fiber to separate zHPF and open systems traffic, preferred pathing features on a zHPF switch can be a great cost saver. With preferred paths established, certain cross-site fiber can be allocated for the mainframe environment, while other fiber can be allocated for open systems. The ISLs can be configured such that in the event of a failure—and only in the event of an ISL failure—the links would be shared by both open systems and mainframe traffic.

Using new technologies

Over the past several years, IBM has announced a series of technology enhancements that require the use of switched zHPF. These include:

- N_Port ID virtualization (NPIV) support for Linux on z Systems
- Dynamic channel path management (DCM)
- z/OS FICON Discovery and Auto-Configuration (zDAC)

IBM announced support for NPIV on Linux for z Systems in 2005. Today, NPIV is supported on the System z9®, z10™, z196, z114, zEC12, zBC12, and z13. Until NPIV was supported on z Systems, adoption of Linux on the mainframe had been relatively slow. NPIV allows for full support of LUN masking and zoning by virtualizing the Fibre Channel identifiers. This in turn allows each Linux on z Systems operating system image to appear as if it has its own individual Host Bus Adapter (HBA)—when those images are in fact sharing Fibre Channel protocol (FCP) channels. Since IBM began supporting NPIV on z Systems, adoption of Linux on z Systems has grown significantly—to the point where IBM believes a significant percent of MIPS shipping on new z Systems machines are for Linux on z Systems implementations. Implementation of NPIV on z Systems requires a switched architecture. With z13, IBM has announced support of up to 8000 virtual machines in one system, as well as increasing the support of the number of virtual images per FCP channel from 32 to 64 virtual images.

Dynamic Channel Path Management (DCM) is another feature that requires a switched zHPF architecture. DCM provides the ability to have z Systems automatically manage zHPF I/O paths connected to DASD subsystems in response to changing workload demands. Use of DCM helps simplify I/O configuration planning and definition, reduces the complexity of managing I/O, dynamically balances I/O channel resources, and enhances availability. DCM can best be summarized as a feature that allows for more flexible channel configurations— by designating channels as “managed”—and proactive performance management. DCM requires a switched zHPF architecture because topology information is communicated through the switch or director. The zHPF switch must have a Control Unit Port (CUP) license, and be configured or defined as a control unit in the Hardware Configuration Definition (HCD) Sysgen.

IBM z/OS® FICON Discovery and Auto-Configuration (zDAC) is the latest technology enhancement for zHPF. IBM introduced zDAC as a follow-on to an earlier enhancement in which the zHPF channels log into the Fibre Channel name server on a zHPF director. zDAC enables the automatic discovery and configuration of zHPF-attached DASD and tape devices. Essentially, zDAC automates a portion of the HCD Sysgen process. zDAC uses intelligent analysis to help validate the z Systems and storage definitions’ compatibility, and uses built-in best practices to help configure for high availability and avoid single points of failure. zDAC is transparent to existing configurations and settings. It is invoked and integrated with the z/OS HCD and z/OS Hardware Configuration Manager (HCM). zDAC also requires a switched zHPF architecture.

IBM introduced support for zHPF in October 2008, and has made several enhancements to zHPF, most recently with the 2015 announcement and GA of zHPF Extended Distance II. While not required for zHPF, a switched architecture is recommended.

Organizations that implement NPIV with a switched zHPF architecture can realize massive consolidation benefits in their Linux on z Systems implementation. They can realize even greater cost savings by implementing a PIM SAN.

Business reasons for a switched zHPF/ FICON architecture

In addition to the technical reasons described earlier, the following business reasons support implementing a switched zHPF architecture:

- Enable consolidation in order to reduce capital and operating expenses
- Improve application performance at long distances
- Support growth and enable effective resource sharing

Massive consolidation

With NPIV support on z Systems, server and I/O consolidation is compelling. IBM undertook a well-publicized project at its internal data centers called Project Big Green.[1] Project Big Green consolidated 3900 open systems servers onto 30 z Systems mainframes running Linux. IBM’s total cost of ownership (TCO) savings—taking into account footprint reductions, power and cooling, and management simplification costs—was nearly 80 percent for a five-year period. These types of TCO savings are why nearly 30 percent of new IBM mainframe processor shipments are now being used for Linux.

Implementation of NPIV requires connectivity from the FICON (FCP) channel to a switching device (director or smaller port-count switch) that supports NPIV. A special microcode load is installed on the FICON channel to enable it to function as an FCP channel. NPIV allows the consolidation of up to 255 Linux on z Systems images ("servers") behind each FCP channel, using one port on a channel card and one port on the attached switching device for connecting these virtual servers. This enables massive consolidation of many HBAs, each attached to its own switch port in the SAN. See Figure 4.

IBM currently supports up to 64 Linux images per FCP channel, and 8000 total images per host. Although this level of I/O consolidation was possible prior to NPIV support on z Systems, implementing LUN masking and zoning in the same manner as with open systems servers or SAN or storage was not possible prior to the support for NPIV with Linux on z Systems.

NPIV implementation on z Systems has also been driving consolidation and adoption of PIM for distributed and open systems and mainframes (FICON). While IBM has supported PIM in z Systems environments since 2003, adoption rates were low until NPIV implementations for Linux on z Systems picked up with the introduction of System z10 in 2008.Enhanced segregation and security beyond simple zoning was now possible through switch partitioning or virtual fabrics or logical switches. With a large number of new mainframes being shipped for use with Linux on z Systems, it is safe to say that a significant number of mainframe environments are now running a shared PIM environment.

Using enhancements in switching technology, performance, and management, PIM users can now fully populate the latest high-density directors with minimal or no oversubscription. They can use management capabilities such as virtual fabrics or logical switches to fully isolate open systems ports and FICON ports in the same physical director chassis. Rather than having more partially populated switching platforms that are dedicated to either open systems or mainframe/FICON, PIM allows for consolidation onto fewer switching devices, reducing management complexity and improving resource use. This, in turn, leads to reduced operating costs, and a lower TCO for the storage network. It also allows for a consolidated, simplified cabling infrastructure.

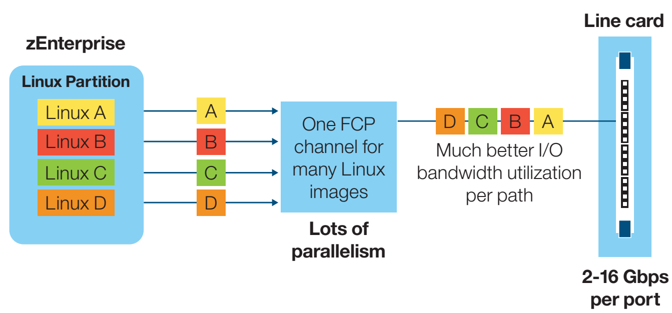

Figure 4. Organizations implement NPIV to consolidate I/O in Linux on z Systems environments

Application performance over distance

As previously discussed, the number of buffer credits per port on a FICON Express16S channel is 90, supporting up to 10 km without performance degradation. If an organization needs to go beyond 10 km for a direct-attached storage configuration they will likely see performance degradation due to insufficient buffer credits. Insufficient quantities of buffer credits do not keep the "pipe" full with streaming frames of data.

Switched zHPF avoids this problem. zHPF directors and switches have sufficient buffer credits available on ports to allow them to stream frames at full-line performance rates with no bandwidth degradation. IT organizations that implement a cascaded zHPF configuration between sites can, with the latest zHPF director platforms, stream frames at 16 Gbps rates with no performance degradation for sites that are 100 km apart. This data traffic can also be compressed—and even encrypted—while traversing the network between sites, allowing IT to securely move more data, faster. See Figure 5.

Switched zHPF technology also allows organizations to take advantage of hardware-based FICON protocol acceleration or emulation techniques for tape (reads and writes), as well as with zGM (z/OS Global Mirror, formerly known as XRC, or Extended Remote Copy). This emulation technology—available on standalone extension switches or as a blade in zHPF directors—allows the channel programs to be acknowledged locally at each site and avoids the back-and-forth protocol handshakes that normally travel between remote sites. It also reduces the impact of latency on application performance and delivers local-like performance over unlimited distances. In addition, this acceleration and emulation technology optimizes bandwidth use.

Bandwidth efficiency is important because it is typically the most expensive budget component in an organization's multisite disaster recovery or business continuity architecture. Anything that can be done to improve the use or reduce the bandwidth requirements between sites would likely lead to significant TCO savings.

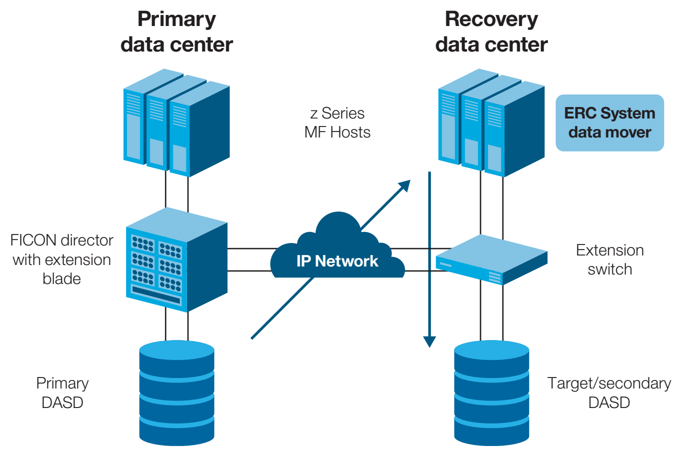

Figure 5. Switched FICON with emulation allows optimized performance and bandwidth use over extended distance

Enabling growth and resource sharing

Direct-attached storage forces a 1:1 relationship between host connectivity and storage connectivity. Each storage port on a DASD array host adapter requires its own physical port connection on a FICON Express16S channel card. These channel cards are typically very expensive on a per-port basis—typically four to six times the cost of a zHPF director port.

Also, there is a finite number of FICON Express16S channels available on a z13 (a maximum of 320), as well as a finite number of host adapter ports on a DASD array. If an organization has a large configuration and a direct-attached zHPF storage architecture, how does it plan to scale its environment? What happens if an organization acquires a company and needs additional channel ports? A switched zHPF infrastructure allows cost-effective, seamless expansion to meet growth requirements.

Direct-attached zHPF storage also typically results in underused channel card ports and host adapter ports on DASD arrays. FICON Express16S channels can comfortably perform at high-channel use rates, and direct-attached storage architecture typically sees channel use rates of 10 percent or less. As illustrated in Figure 6, using zHPF directors or switches allows organizations to maximize channel use.

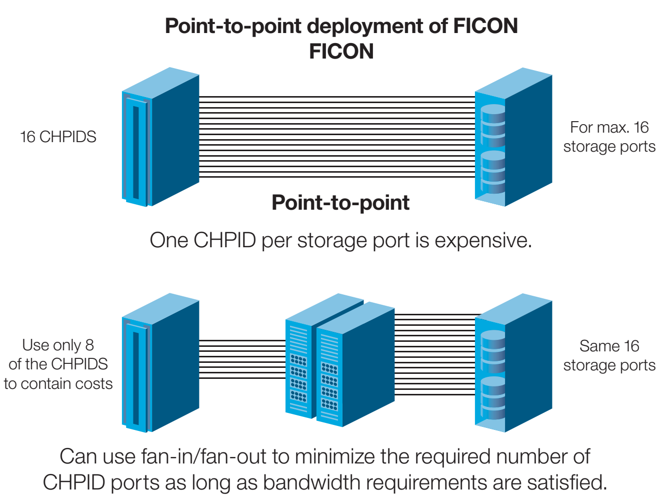

Figure 6. Switched FICON drives improved channel use, while preserving Channel Path Identifiers (CHPIDs) for growth

It also is very important to keep traffic for tape drives streaming, and avoid stopping and starting the tape drives, as this leads to unwanted wear and tear of tape heads, cartridges, and the tape media itself. This is accomplished by using FICON acceleration and emulation techniques as described earlier. A configuration similar to the one shown in Figure 7 can also be implemented. Such a configuration requires solid analysis and planning, but it will pay dividends for an organization's FICON tape environment.

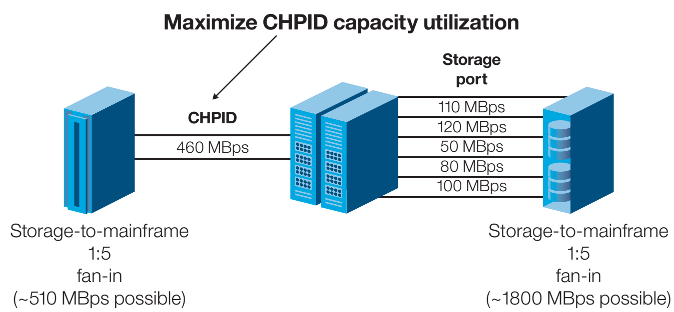

Figure 7. A well-planned configuration can maximize CHPID capacity use for FICON tape efficiency

Finally, switches facilitate fan-in—allowing different hosts and LPARs whose I/O subsystems are not shared to share the same assets. While some benefit may be realized immediately, the potential for value in future equipment planning can be even greater. With the ability to share assets, equipment that would be too expensive for a single environment can be deployed in a cost-saving manner.

The most common example is to replace tape farms with virtual tape systems. By reducing the number of individual tape drives, maintenance (service contracts), floor space, power, tape handling, and cooling costs are reduced. Virtual tape also improves reliable data recovery, allows for significantly shorter Recovery Time Objectives (RTO) and nearer Recovery Point Objectives (RPO), and offers features such as peer-to-peer copies. However, without the ability to share these systems, it may be difficult to amass sufficient cost savings to justify the initial cost of virtual tape. And the only practical way to share these standalone tape systems or tape libraries is through a switch.

With disk subsystems, in addition to sharing the asset, it is sometimes desirable to share the data across multiple systems. The port limitations on DASD may prohibit or limit this capability using direct-attached (point-to-point) FICON channels. Again, the switch can provide a solution to this issue.

Even when there is no need to share devices during normal production, this capability can be valuable in the event of a failure. Data sets stored on tape can quickly be read by CPUs picking up workload that is already attached to the same switch as the tape drives. Similarly, data stored on DASD can be available as soon as a fault is determined.

Switch features such as preconfigured port prohibit and allow matrix tables can ensure that access intended only for a disaster scenario is prohibited during normal production.

Investments made in switches for disaster recovery and business continuance are likely to pay the largest dividends. Having access to alternative resources and multiple paths to those resources can result in significant savings in the event of a failure.

Summary

Direct-attached zHPF might appear to be a great way to take advantage of FICON technology's advances. A closer examination, however, shows that switched zHPF, similar to switched FICON, is a better, more robust architecture for enterprise data centers. Switched zHPF offers:

- Better use of host channels and their performance capabilities
- Scalability to meet growth requirements
- Improved reliability, problem isolation, and availability
- Flexible connectivity to support evolving infrastructures
- More robust business continuity implementations through cascaded FICON
- Improved distance connectivity, with improved performance over extended distances
- New mainframe I/O technology enhancements such as NPIV, FICON DCM, zDAC, and zHPF

Switched zHPF also provides many business advantages and potential cost savings, including:

- The ability to perform massive server, I/O, and SAN consolidation, dramatically reducing capital and operating expenses
- Local-like application performance over any distance, allowing host and storage resources to reside wherever business dictates
- More effective resource sharing, improving use and reducing costs

FICON Express16S channels using default FICON on z13 processors can have up to 64 concurrent I/Os (open exchanges) to different devices. FICON Express16S channels running zHPF can have up to 750 concurrent I/Os on the z13 processor family. Only when a director or switch is used between the host and storage device can the true performance potential inherent in these channel bandwidth and I/O processing gains be fully exploited.

With the growing trend toward increased usage of Linux on z Systems, and the cost advantages of NPIV implementations and PIM SAN architectures, direct-attached storage in a mainframe environment is becoming a thing of the past. The advantages of a switched zHPF infrastructure are simply too great to ignore.

For more information

To learn more about switched zHPF/FICON architecture please contact your IBM representative or IBM Business Partner, or contact Raymond Newsom at rnewsom@us.ibm.com.

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

IBM Systems
Route 100
Somers, NY 10589

Produced in the United States of America
February 2016

[1] https://www-03.ibm.com/press/us/en/pressrelease/21524.wss

TSW03333-USEN-00