Redpaper

Ilya Krutov

Choosing eXFlash Storage on IBM eX5 Servers

Introduction

In keeping with their leadership position providing continuous technology improvement and innovation, IBM® recently refreshed the eXFlash offering to expand its capabilities. Compared to the previous generation, the new features of the IBM eXFlash solution include:

- Hot-swap capability
- IOPS performance improved by six times to 240,000 IOPS per eXFlash unit
- Storage capacity increased by four times to 1.6 TB per eXFlash unit

You can expect further expansion of IBM eXFlash capabilities for both IOPS and throughput performance and capacity over time as SSD technology improvements are developed and adopted by the industry.

The intent of this IBM Redpaper™ document is to discuss the benefits and advantages of IBM eXFlash storage technology within the IBM eX5 portfolio, demonstrate how eXFlash fits into a multi-tiered storage design approach, and provide recommendations on the most effective use of eXFlash storage within the enterprise information infrastructure.

This paper covers the following topics:

- “Application requirements for storage”
- “IBM System x server and storage products”
- “IBM eXFlash deployment scenarios”

Application requirements for storage

Choosing the right storage for application data can be a complex task because you must ensure that critical business and application requirements are met while costs are kept optimized. In particular, storage performance capabilities should match the processing capabilities of the server itself to ensure the most efficient utilization of system resources. There is no “one size fits all” approach possible because different applications have different storage data access patterns.

In general, the factors to consider during the planning process for application data storage include:

- Importance of data (Can I accept the loss of data?)
- Sensitivity of data (Do I need an advanced data protection and security?)
- Availability of data (Do I need the data 24 hours per day, 7 days per week?)
- Security of data (Who can read, modify, and delete the data?)
- Data access speed (How quickly do I need to insert and extract the data?)
- Performance or workload capacity (How many IOPS for I/O-intensive workloads and how many MBps for throughput-intensive workloads do I need?)
- Storage capacity (How much space do I need to store the data?)
- Frequency of access (How often do I need the data?)
- Backup and recovery strategy (How much time do I need to backup and restore the data?)
- Retention policy (How long should I keep the data?)
- Scalability for future growth (Do I expect the workload increase in the near future?)
- Storage deployment: internal or external (If external, then JBOD or storage controller? If storage controller, then SAS, iSCSI, FC, or FCoE?)
- Data access pattern (How does the application access the data?)
  - Read or write intensive
  - Random or sequential access
  - Large or small I/O requests

Answers to these questions will help you to formalize the performance, availability, and capacity requirements for your applications, and match these requirements with the tiered storage design model.

Multi-tiered storage architecture

As we mentioned previously, the planning of information infrastructure includes choosing the most cost-effective way to fulfill the application requirements for storage access with respect to speed, capacity, and availability. To describe these requirements, and to establish the framework for the deployment of the storage infrastructure, the storage tiering approach was established.

Each *storage tier* defines a set of characteristics to meet the application requirements. There are four tiers, each with performance, availability, capacity, and access pattern characteristics for the data residing on that tier. Knowing your data access requirements will help you to place data on the appropriate storage tier, thereby ensuring that your storage infrastructure is capable of running your workloads in a cost-efficient manner.

The storage tiers, their corresponding characteristics, suitable storage media types, and relative cost per gigabyte are listed in Table 1.

Table 1 Storage tiers and characteristics

Storage tier	Characteristic	Storage media type	Cost per gigabyte
Tier 0 (SSD)	Random access I/O-intensive Extreme performance Extreme availability Frequent access	SSD, IBM eXFlash	Very high
Tier 1 (Online)	Random access I/O-intensive High performance High availability Frequent access	SAS, FC HDD	High
Tier 2 (Nearline)	Sequential access Throughput-intensive Capacity High availability Infrequent access	NL SAS, NL SATA HDD	Moderate
Tier 3 (Offline)	Sequential access Throughput-intensive Large capacity Long-term retention No direct access	Tape	Low

Tiers 0, 1, and 2 are considered *primary* storage, meaning data on these tiers can be directly accessed by the application. Tier 3 is *secondary* storage, that is, data that cannot be accessed directly; in order to access this data it must be moved to primary storage. Tier 0 has been specifically added for the enterprise solid state drives.

Data storage closer to the main memory (that is, closer to the application processes residing in the main memory) costs more to implement than storage that is farther away. In other words, the price per GB of data storage increases from tier 3 to tier 0. To keep costs optimized, the most demanded data (also referred to as *hot* data) from the working data set should be placed closest to the main memory, whereas less demanded data can be placed on a higher (more distant) storage tier. From a planning standpoint, the rules that define the policy of placing data onto different storage tiers are part of the overall information life cycle management strategy for the organization. From a technology standpoint, data management and relocation policy can be implemented either manually by administrators or automatically by management software that supports policy-based data relocation, for example, IBM GPFS™, or by integrated features of storage systems, like the IBM Easy Tier™ feature of IBM Storwize® V7000 external storage.

Storage performance considerations for applications

Currently, the processor, memory, and I/O subsystem are well balanced and virtually not considered as performance bottlenecks in the majority of systems. The main source of performance issues tends to be related to storage I/O activity because the speed of traditional HDD-based storage systems still does not match the processing capabilities of the servers.

Various caching technologies can be implemented to increase HDD storage access speed. Despite the overall size of a stored data set only a portion of its data is actively used during normal operations at certain time intervals. Data caching algorithms ensure that the most demanded data (most frequently used) is always kept as close to the application as possible to provide the fastest response time.

Caching exists at several levels. Storage controllers use fast DRAM cache to keep the most frequently used data from disks; however, the cache size is normally limited to several GBs. Operating systems and certain applications keep their own disk cache in the fast system memory, but the cost per GB of RAM storage is high.

With the introduction of solid state drives there is an opportunity to dramatically increase the performance of disk-based storage to match the capabilities of other server subsystems, while keeping costs optimized because the solid state drives have a lower cost per MB compared to DRAM memory, and lower latency compared to traditional hard drives. This is illustrated in Figure 1.

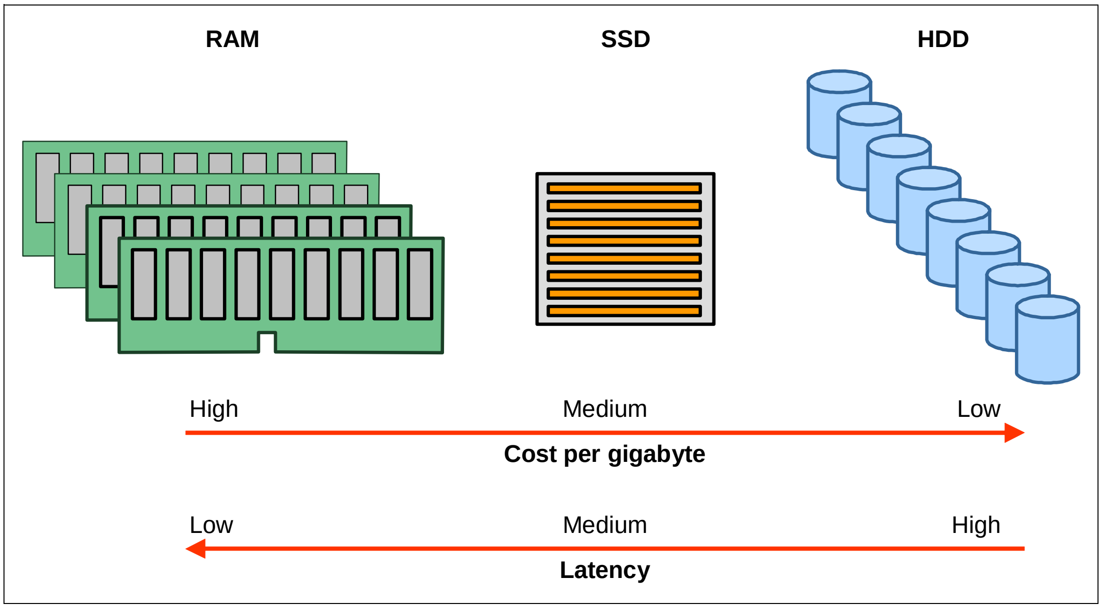

Figure 1 Cost per gigabyte and latency for RAM, SSD, and HDD

In general, there are two key types of storage applications based on workload they generate:

- *I/O-intensive* applications require the storage system to process as many host's read and write requests (or I/O requests) per second as possible given the average I/O request size used by this application, typically 8 - 16 KB. This behavior is most common for OLTP databases.
- *Throughput-intensive* applications require storage system to transfer to or from host as many gigabytes of information per second as possible, and they typically use I/O request size of 64 - 128 KB. These characteristics commonly inherent to file servers, multimedia streaming, and backup.

Therefore, there are two key performance metrics to evaluate storage system performance: *input/output requests per second (IOPS)* and *throughput* depending on application workload.

Another important factor to take into account is the *response time* (or *latency*), or how much time does the application spend waiting for the response from the storage system after submitting a particular I/O request. In other words, response time is the amount of time required by the storage system to complete an I/O request. Response time has direct impact on the productivity of users who work with the application (because of how long it takes to get the requested information) and also on the application itself. For example, slow response to database write requests might cause multiple record locks and further performance degradation of the application.

Key factors affecting the response time of the storage system are how quickly the required data can be located on the media (seek time) and how quickly it can be read from or written to the media, which in part depends on the size of the I/O requests (reading or writing more data normally takes more time).

In addition, the majority of applications generate many storage I/O requests at the same time, and these requests might spend some time in the queue if they cannot be immediately handled by the storage system. The number of I/O requests that can be concurrently sent to the storage system for execution is called *queue depth*. This refers the *service* queue, that is, the queue of requests currently being processed by the storage system. If the number of outgoing I/O requests exceeds the parallel processing capabilities of the storage system (I/O queue depth), the requests are put into the *wait* queue, and then moved to the service queue when a place becomes available. This also affects the overall response time.

From the traditional spinning HDD perspective, improvement of latency is limited by mechanical design. Despite an increase in rotational speed of disk plates and higher density of stored data, the response time of HDD is still several milliseconds, which effectively limits its maximum IOPS. For example, a single 2.5 in.15K rpm SAS HDD is capable of approximately 300 IOPS.

With SSD-based eXFlash latency is measured in dozens of microseconds (or almost 100 times lower latency than for hard drives), which in turn leads to the 240,000 IOPS capabilities identified earlier. Higher IOPS capabilities also mean higher queue depth and therefore better response time for almost all types of storage I/O-intensive applications.

In other words, if the application is multi-user, heavily loaded, and accesses storage with random I/O requests of a small size, then this application is a good candidate to consider to put its entire data set (or part of it) on an IBM eXFlash or external SSD-based storage system. Conversely, if an application transfers large amounts of data, like backups or archiving, then eXFlash might not provide any advantage because the limiting factor will be the bandwidth of the SSD interface.

The knowledge of how the application accesses data—read-intensive or write-intensive, and random data access or sequential data access—helps you design and implement the most cost-efficient storage to meet required service level agreements (SLAs). Table 2 summarizes the relationship between typical application workload patterns and application types.

Table 2 Typical application workload patterns

	Workload type						
Application type	**Read intensive**	**Write intensive**	**I/O intensive**	**Throughput intensive**	**Random access**	**Sequential access**	**Good for eXFlash**
OLTP Database	Yes	Yes	Yes		Yes		Yes
Data warehouse	Yes		Yes		Yes		Yes
File server	Yes			Yes	Yes		
Email server	Yes	Yes	Yes		Yes		Yes
Medical imaging	Yes			Yes	Yes		Yes
Document imaging		Yes		Yes		Yes	
Streaming media	Yes			Yes		Yes	

	Workload type						
Application type	Read intensive	Write intensive	I/O intensive	Throughput intensive	Random access	Sequential access	Good for eXFlash
Video on demand	Yes			Yes	Yes		Yes
Web/Internet	Yes	Yes	Yes		Yes		Yes
CAD/CAM	Yes		Yes		Yes		Yes
Archives/Backup		Yes		Yes		Yes	

As a general rule, to deploy the most efficient storage that satisfies application performance requirements given the required storage capacity, you should consider:

- For I/O-intensive workloads: A higher number of hard drives (more drives of smaller capacities because adding drives provides an almost linear increase in IOPS), or eXFlash with solid state drives.
- For throughput-intensive workloads: A higher bandwidth between the host controller and storage arrays, utilizing more host ports on a controller and higher port speeds (for example, 6 Gbps rather than 3 Gbps for SAS or 8 Gbps rather than 4 Gbps for Fibre Channel) with a sufficient number of drives in the array to put the workload on these links.

Based on Table 2 on page 5, the following application types benefit from the deployment based on IBM eXFlash and other SSD-based storage:

- Databases
- Data warehouse
- Email
- Medical imaging
- Video on demand
- Web
- CAD/CAM storage

Typical IBM eXFlash usage scenarios include:

- High-speed read cache in a local or SAN-based storage environment
- Temporary local storage space for mid-tier applications and databases
- Main (Tier 0) local data storage in a single server environments or in a distributed scale-out environment with local-only storage or mixed local and SAN-based storage

Typical IBM SSD usage in case of external storage includes Tier 0 main data storage with automated data movement capabilities like Easy Tier in IBM Storwize V7000.

IBM System x server and storage products

The storage deployment scenarios described in this paper reference IBM eX5 systems, and entry and midrange IBM System Storage® products. This section provides a brief overview of these offerings.

IBM eX5 architecture and portfolio

The IBM eX5 product portfolio represents the fifth generation of servers built upon IBM Enterprise X-Architecture®. Enterprise X-Architecture is the culmination of generations of IBM technology and innovation derived from our experience in high-end enterprise servers. Now with eX5, IBM scalable systems technology for Intel processor-based servers has also been delivered to blades. These servers can be expanded on demand and configured by using a building block approach that optimizes system design servers for your workload requirements.

As a part of the IBM Smarter Planet™ initiative, our IBM Dynamic Infrastructure® charter guides us to provide servers that improve service, reduce cost, and manage risk. These servers scale to more CPU cores, memory, and I/O than previous systems, enabling them to handle greater workloads than the systems they supersede. Power efficiency and machine density are optimized, making them affordable to own and operate.

The ability to increase the memory capacity independently of the processors means that these systems can be highly utilized, yielding the best return from your application investment. These systems allow your enterprise to grow in processing, I/O, and memory dimensions, so that you can provision what you need now, and expand the system to meet future requirements. System redundancy and availability technologies are more advanced than the technologies that were previously available in the x86 systems.

The IBM eX5 product portfolio is built on Intel® Xeon® processor E7-8800/4800/2800 product families. With the inclusion of these processors, the eX5 servers became faster, more reliable, and more power efficient. As with previous generations of IBM Enterprise X-Architecture systems, these servers have delivered many class leading benchmarks, including the highest TPC-E result for a system of any architecture.

IBM eX5 systems

The four systems in the eX5 family are the x3850 X5, x3950 X5, x3690 X5, and the HX5 blade. The eX5 technology is primarily designed around three major workloads: database servers, server consolidation using virtualization services, and Enterprise Resource Planning (application and database) servers. Each system can scale with additional memory by adding an IBM MAX5 memory expansion unit to the server, and the x3850 X5, x3950 X5, and HX5 can also be scaled by connecting two systems to form a 2-node scale.

Figure 2 on page 8 shows the IBM eX5 family.

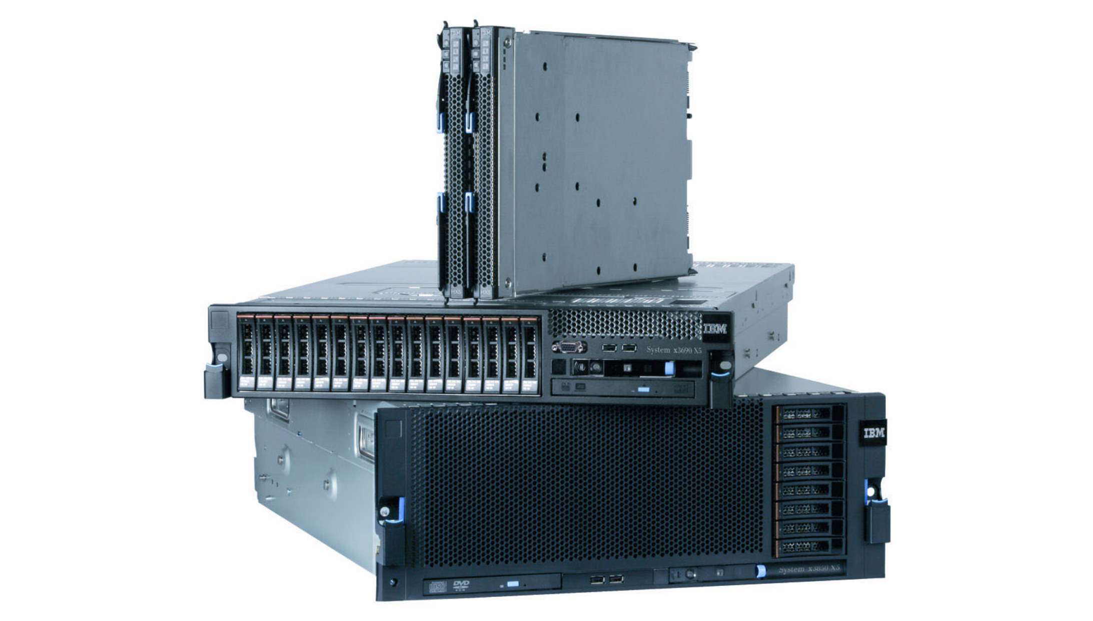

Figure 2 eX5 family (top to bottom): IBM BladeCenter® HX5 (2-node), IBM System x3690 X5, and IBM System x3850 X5 (the IBM System x3950 X5 looks the same as the x3850 X5)

The IBM System x3850 X5 and x3950 X5 are 4U highly rack-optimized servers. The x3850 X5 and the workload-optimized x3950 X5 are the new flagship servers of the IBM x86 server family. These systems are designed for maximum utilization, reliability, and performance for computer-intensive and memory-intensive workloads. These servers can be connected together to form a single system with twice the resources, or support memory scaling with the attachment of a MAX5. With the new Intel Xeon E7 series processors, the x3850 X5 and x3950 X5 now can scale to a two server plus two MAX5 configuration.

The IBM System x3690 X5 is a 2U rack-optimized server. This machine brings features and performance to the middle tier, as well as a memory scalability option with MAX5.

The IBM BladeCenter HX5 is a single-wide (30 mm) blade server that follows the same design as all previous IBM blades. The HX5 brings unprecedented levels of capacity to high-density environments. The HX5 is expandable to form either a two-node system with four processors, or a single-node system with the MAX5 memory expansion blade.

When compared to other machines in the IBM System x® portfolio, these systems represent the upper end of the spectrum, are suited for the most demanding x86 tasks, and can handle jobs which previously might have been run on other platforms. To assist with selecting the ideal system for a given workload, IBM designed workload-specific models for virtualization and database needs.

Table 3 gives an overview of the features of IBM eX5 systems.

Table 3 Maximum configurations for the eX5 systems

Maximum configurations		x3850 X5/x3950 X5	x3690 X5	HX5
Processors	1-node	4	2	2
	2-node	8	Not available	4
Memory	1-node	2048 GB (64 DIMMs)[a]	1024 GB (32 DIMMs)[b]	256 GB (16 DIMMs)
	1-node with MAX5	3072 GB (96 DIMMs)[a]	2048 GB (64 DIMMs)[b]	640 GB (40 DIMMs)
	2-node	4096 GB (128 DIMMs)[a]	Not available	512 GB (32 DIMMs)
	2-node with MAX5	6144 GB (192 DIMMs)[a]	Not available	Not available
Disk drives (non-SSD)[c]	1-node	8	16	Not available
	2-node	16	Not available	Not available
SSDs	1-node	16	24	2
	2-node	32	Not available	4
Standard 1 Gb Ethernet interfaces	1-node	2	2	2
	2-node	4	Not available	4
Standard 10 Gb Ethernet interface	1-node	2[d]	2[d]	0
	2-node	4	Not available	0

a. Requires full processors to install and use all memory.
b. Requires that the memory mezzanine board is installed along with processor 2.
c. For the x3690 X5 and x3850 X5, additional backplanes might be needed to support these numbers of drives.
d. Standard on most models.

IBM eX5 chip set

The members of the eX5 server family are defined by their ability to use IBM fifth-generation chip sets for Intel x86 server processors. IBM engineering, under the banner of Enterprise X-Architecture (EXA), brings advanced system features to the Intel server marketplace. Previous generations of EXA chip sets powered System x servers from IBM with scalability and performance beyond what was available with the chip sets from Intel.

The Intel QuickPath Interconnect (QPI) specification includes definitions for the following items:

- Processor-to-processor communications
- Processor-to-I/O hub communications
- Connections from processors to chip sets, such as eX5, referred to as *node controllers*

To fully utilize the increased computational ability of the new generation of Intel processors, eX5 provides additional memory capacity and additional scalable memory interconnects (SMIs), increasing bandwidth to memory. The eX5 also provides these additional reliability, availability, and serviceability (RAS) capabilities for memory: Chipkill, Memory ProteXion, and Full Array Memory Mirroring.

QPI uses a source snoop protocol. This technique means that a CPU, even if it knows another processor has a cache line it wants (the cache line address is in the snoop filter, and it is in the shared state), must request a copy of the cache line and wait for the result to be returned from the source. The eX5 snoop filter contains the contents of the cache lines and can return them immediately.

Memory that is directly controlled by a processor can be accessed faster than through the eX5 chip set, but it is connected to all processors and introduces less delay than accesses to memory controlled by another processor in the system.

The eX5 chip set also has, as with previous generations, connectors to allow systems to scale beyond the capabilities provided by the Intel chip sets. We call this scaling Enterprise X-Architecture (EXA) scaling. You can use EXA scaling to connect two x3850 X5 servers and two MAX5 memory expansion units together to form a single system image with up to eight Intel Xeon E7 processors and up to 6TB of RAM.

Intel Xeon processors

The latest models of the eX5 systems use Intel Xeon E7 processors. Earlier models use Intel Xeon 7500 or 6500 series processors.

The Intel Xeon E7 family of processors used in the eX5 systems (more precisely, the E7-2800, E7-4800 and E7-8800 series) are follow-ons to the Intel Xeon 6500 and 7500 family of processors. Whereas the processor architecture is largely unchanged, the lithography size was reduced from 45 nm to 32 nm, allowing for more cores (and thus more threads with Hyper-Threading Technology), and more last level cache, while staying within the same thermal design profile (TDP) and physical package size.

The three groups of the E7 family of processors support scaling to different levels:

- The E7-2800 family is used in the x3690 X5 and BladeCenter HX5. Members of this series only support two-processor configurations, so they cannot be used in a two-node HX5 configuration. Most processors in this family support connection to a MAX5 (except E7-2803 and E7-2820).
- The E7-4800 family is primarily used in the HX5 and the x3850 X5. This series supports four-processor configurations, so can be used for single node x3850 X5 and two node HX5s. All members of the E7-4800 family support connection to a MAX5, and can also be used for two-node x3850 X5 with MAX5 configurations. Such configurations use EXA scaling, which the E7-4800 processors support.
- The E7-8800 family processors are used in the x3850 X5 to scale to two nodes without MAX5s. There are specific high frequency and low power models of this processor available for the x3690 X5 and HX5 as well.

These scalability capabilities are summarized in Table 4 on page 11.

Table 4 Comparing the scalability configurations of the Intel Xeon E7 processors

	E7-2800	E7-4800	E7-8800
x3690	Yes	Yes	Yes
x3690 X5 with MAX5	Yes[a]	Yes	Yes
HX5	Yes	Yes	Yes
HX5 with MAX5	Yes[a]	Yes	Yes
HX5 2-node	Not supported	Yes	Yes
x3850 X5	Not supported	Yes	Yes
x3850 X5 with MAX5	Not supported	Yes	Yes
x3850 X5 2-node without MAX5	Not supported	Not supported	Yes
x3850 X5 2-node with MAX5	Not supported	Yes (EXA scaling)	Yes (EXA scaling)

[a] E7-2803 and E7-2820 processors do not support MAX5

For additional information about the IBM eX5 portfolio refer to the following publication:

- IBM eX5 Portfolio Overview: IBM System x3850 X5, x3950 X5, x3690 X5, and BladeCenter HX5

  http://www.redbooks.ibm.com/abstracts/redp4650.html

IBM eXFlash technology

IBM eXFlash technology is a server-based high performance internal storage solution that is based on Solid State Drives (SSDs) and performance-optimized disk controllers (both RAID and non-RAID).

A single eXFlash unit accommodates up to eight hot-swap SSDs, and can be connected to up to two performance-optimized controllers. eXFlash is supported on IBM System x3690 X5, x3850 X5, and x3950 X5 servers.

Figure 3 shows an eXFlash unit, with the status lights assembly on the left side.

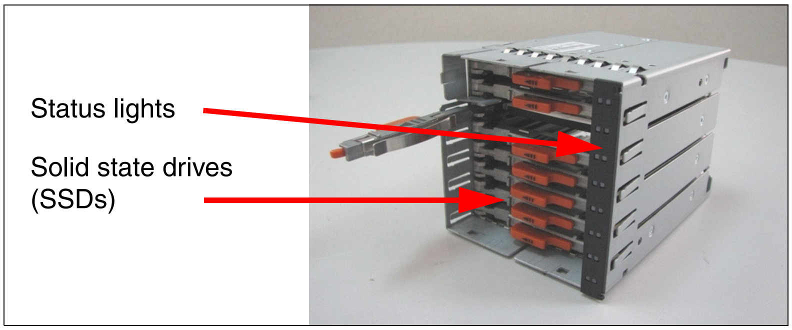

Figure 3 IBM eXFlash unit

Each eXFlash unit occupies four 2.5-inch SAS hard disk drive bays. The eXFlash units can be installed in the following configurations:

- The x3850 X5 can have up to sixteen 1.8-inch SSDs with up to two eXFlash units (up to eight SSDs per eXFlash unit).
- The x3950 X5 database-optimized models have two eXFlash units standard with sixteen 200 GB SSDs installed (for the models with Intel Xeon E7 series processors) for a total of 32 SSDs in a dual node configurations.
- The x3690 X5 can have up to twenty four 1.8-inch SSDs with up to three eXFlash units (up to eight SSDs per eXFlash unit).

A single IBM eXFlash unit has the following characteristics:

- Up to eight 1.8-inch hot-swap front-accessible SSDs
- Up to 240,000 random read IOPS
- Up to 2 GBps of sustained read throughput
- Up to 1.6 TB of available storage space with IBM 200 GB 1.8-inch eMLC SSDs or up to 400 GB with IBM 50 GB 1.8-inch eMLC SSDs

In theory, the random I/O performance of a single eXFlash unit is equivalent to that of a storage system consisting of about 800 traditional spinning HDDs. Besides the HDDs themselves, building such a massive I/O-intensive high-performance storage system requires external deployment with many additional infrastructure components including host bus adapters (HBAs), switches, storage controllers, disk expansion enclosures, and cables. Consequently, this leads to more capital expenses, floor space, electrical power requirements, and operations and support costs. Because eXFlash is based on internal server storage, it does not require all those additional components and their associated costs and environmental requirements.

In summary, an IBM eXFlash solution provides the following benefits:

- Significantly lower implementation cost (up to 97% lower) for high performance I/O-intensive storage systems with the best IOPS/$ performance ratio
- Significantly higher performance (up to 30 times or more) for I/O-intensive applications like databases and business analytics with up to nine times less response time
- Significant savings in power and cooling (up to 90%) with a high performance per watt ratio
- Significant savings in floor space (up to 30 times less) with extreme performance per rack U space ratio
- Simplified management and maintenance with internal server-based configurations (no external power and information infrastructure needed)

IBM eXFlash is optimized for a heavy mix of random read and write operations, such as transaction processing, data mining, business intelligence and decision support, and other random I/O-intensive applications. In addition to its superior performance, eXFlash offers superior uptime with three times the reliability of mechanical disk drives. SSDs have no moving parts to fail. They use Enterprise Wear-Leveling to extend their use even longer. All operating systems that are listed in IBM ServerProven® for each machine are supported for use with eXFlash.

The eXFlash SSD backplane uses two long SAS cables, which are included with the backplane option. When more than one SSD backplane is installed, each backplane must be connected to a separate disk controller.

In environments where RAID protection is required, that is, eXFlash is used as a master data storage, use two RAID controllers per backplane to ensure the peak IOPS can be reached. Although use of a single RAID controller results in a functioning solution, peak IOPS can be reduced by a factor of approximately 50%. Use the ServeRAID M5014 or M5015 with the ServeRAID M5000 Performance Accelerator Key when all drives are SSDs. Alternatively, the ServeRAID B5015 SSD Controller can be used instead.

The main advantage of B5015 and M5014 or M5015 with Performance Key controllers for SSDs is a *Cut Through I/O (CTIO)* feature enabled. CTIO optimizes highly random read and write I/O operations for small data blocks to support the high IOPS capabilities of SSD drives and to ensure the fastest response time to the application. For example, enabling CTIO on a RAID controller with SSDs allows the controller to achieve up to two times more IOPS compared to a controller with the CTIO feature disabled.

Note: A single eXFlash unit requires a dedicated controller (or two controllers). When used with eXFlash, these controllers cannot be connected to the HDD backplanes. The ServeRAID B5015 SSD Controller is only supported with SSDs.

In a non-RAID environment where eXFlash can be used as a high-speed read cache, use the IBM 6 Gb Performance Optimized HBA to ensure maximum random I/O read performance is achieved. Only one 6 Gb SSD HBA is supported per single SSD backplane.

It is possible to mix RAID and non-RAID environments; however, the maximum number of disk controllers that can be used with all SSD backplanes in a single system is four.

IBM eXFlash requires the following components:

- IBM eXFlash hot-swap SAS SSD backplane (if not already installed on a standard pre-configured model)
- IBM solid state drives (SSDs)
- IBM disk controllers

Table 5 shows ordering information for eXFlash backplanes.

Note: IBM System x3850 X5 and x3690 X5 have different eXFlash backplanes.

Table 5 IBM eXFlash 8x 1.8-inch HS SAS SSD Backplane

Part number	**Feature code**	**Description**
59Y6213	4191	IBM eXFlash 8x 1.8-inch HS SAS SSD Backplane for x3850 X5
60Y0360	9281	IBM eXFlash 8x 1.8-inch HS SAS SSD Backplane for x3690 X5

Table 6 lists the 1.8-inch solid state disk (SSD) options that are supported in the eX5 systems. These drives are supported with the eXFlash SSD backplane, part number 59Y6213.

Table 6 Supported 1.8-inch SSDs

Part number	Feature code	Description
43W7726	5428	IBM 50 GB SATA 1.8-inch MLC SSD
43W7746	5420	IBM 200 GB SATA 1.8-inch MLC SSD

Table 7 lists the supported controllers.

Table 7 Controllers supported with the eXFlash SSD backplane option

Part number	Feature code	Description
46M0912	3876	IBM 6 Gb Performance Optimized HBA
46M0916	3877	ServeRAID M5014 SAS/SATA Controller[a]
46M0829	0093	ServeRAID M5015 SAS/SATA Controller[a]
46M0969	3889	ServeRAID B5015 SSD Controller
81Y4426	A10C	ServeRAID M5000 Series Performance Accelerator Key[b]

[a] Requires M5000 Performance Accelerator key when used with eXFlash
[b] Adds Cut Through I/O (CTIO) for SSD FastPath optimization on ServeRAID M5014, M5015, and M5025 controllers

For more information about the devices mentioned here, see the relevant IBM Redbooks® at-a-glance guides:

- Solid State Drives for IBM BladeCenter and System x servers

  http://www.redbooks.ibm.com/abstracts/tips0792.html
- IBM 6 Gb Performance Optimized HBA

  http://www.redbooks.ibm.com/abstracts/tips0744.html
- ServeRAID B5015 SSD Controller

  http://www.redbooks.ibm.com/abstracts/tips0763.html
- ServeRAID M5015 and M5014 SAS/SATA Controllers

  http://www.redbooks.ibm.com/abstracts/tips0738.html
- ServeRAID M5000 Series Performance Accelerator Key for IBM System x

  http://www.redbooks.ibm.com/abstracts/tips0799.html

IBM System Storage products

The IBM System Storage disk products portfolio covers the needs of a wide spectrum of possible implementations, from entry-level to large enterprise. It combines the high performance of the IBM System Storage DS8000® series and XIV® enterprise storage systems with the Storwize V7000, N Series, and DS5000 series of midrange systems, and with the DS3500 series low priced entry systems.

The family is further complemented by a range of expansion enclosures to expand the disk storage capacities of individual systems into hundreds of terabytes (TB), or even to a petabyte (PB). Furthermore, a full range of IBM System Storage capabilities such as advanced copy services, management tools, and virtualization services are available to help protect data.

For the purpose of this paper, we compare the key capabilities of the entry SAN (DS3500 series), midrange SAN (DS5000 series and V7000), and NAS (midrange N Series and Scale Out Network Attached Storage - SONAS) storage families from the performance, scalability, capacity, and advanced features points of view. Table 8 on page 15 compares DS3524, V7000, DS5300, and N6270, the top models of the respective storage families. The purpose of this comparison is to provide a brief overview of the relative capabilities of various IBM System Storage families so that you can understand their positioning and evaluate which ones are most suitable for your information infrastructure to handle projected workloads.

Table 8 IBM System Storage feature comparison: SAN storage

Storage family Feature	SAN			NAS or File Storage
	DS3524	V7000	DS5300	N6270
Scalability and capacity				
Host connectivity	FC, iSCSI, SAS	FC, iSCSI	FC, iSCSI	FC, iSCSI, FCoE, NAS
Host interface	8 Gb FC, GbE, 10 GbE, 6 Gb SAS	8 Gb FC, GbE, 10 GbE	8 Gb FC, GbE, 10 GbE	4 Gb FC, 8 Gb FC, GbE, 10 GbE
Max number of host ports	12	32	16	56
Max number of drives	192	240 (LFF), 480 (SFF)	480	960
Drive types	SAS, NL SAS, SSD	SAS, SSD	FC, SATA, SSD	FC, SAS, SATA
Max raw capacity, TB	384	480 (LFF), 288 (SFF)	960	2,880
Max cache size, GB	4	32	64	32
Performance				
SPC-1 IOPS	24,449[a]	56,511[b]	62,243[c]	Not available
SPC-2 throughput, MBps	2,510	3,132	5,543	Not available
SPECsfs2008_nfs throughput, operations/s	Not available	Not available	Not available	101183[d]
Advanced features				
Snapshots	Yes	Yes	Yes	Yes
Remote replication	Yes	Yes	Yes	Yes
Automated storage tiering	No	Yes	No	Yes

[a] 96x 300GB 10K RPM 2.5” SAS HDDs
[b] 240x 300GB 10K RPM 2.5” SAS HDDs
[c] 256x 146.8GB/15K RPM DDMs
[d] 360x 450GB 15K RPM SAS HDDs

For more information about SPC benchmarks and results visit:

- Storage Performance Council (SPC)

  http://www.storageperformance.org

For more information about SPECsfs2008 benchmarks and results visit:

- Standard Performance Evaluation Corporation (SPEC)

  http://www.spec.org/sfs2008/

For more information about available IBM System Storage offerings refer to:

- IBM System Storage Solutions Handbook

  http://www.redbooks.ibm.com/abstracts/sg245250.html

IBM eXFlash deployment scenarios

The following subsections provide some ideas on where to deploy SSD-based IBM eXFlash solutions locally or in conjunction with external storage to get significant performance benefits while keeping costs optimized. The following scenarios are discussed:

- Transaction processing (OLTP databases)
- Data warehousing (OLAP databases)
- Corporate email
- Actively connected users (Web 2.0)
- Medical imaging
- Video on demand

Each subsection includes a workload description, possible sources of storage I/O performance issues assuming no other bottlenecks exist in the system, and recommended internal IBM eXFlash-based or external IBM System Storage solutions that also include high-availability and scalability considerations.

OLTP databases

Online transaction processing (OLTP) is a multi-user, memory-, CPU-, and storage I/O-intensive, random workload. It is characterized by a large number of small read and write storage I/O requests (typically four or eight kilobytes and 70/30 read/write ratio) generated by transactions originated by multiple users. The transactions are relatively simple; however, every single transaction can generate dozens of physical storage I/O requests depending on transaction type, application architecture, and business model used.

The key performance indicator of transactional systems is the response time: the client expects to get the requested product information or place an order quickly. If the expectations are not met then the chance that the client will go to a competitor is high. Because of that, storage I/O performance is considered an important factor to ensure the response time goals are met, and to keep other system resources (CPU and memory) at good utilization and not waiting for the data.

Typically, the OLTP workloads can be classified as light, medium, or heavy based on the number of transactions per second (tps) as follows:

- Light workload: Up to 100 tps
- Medium workload: Up to 1,000 tps
- Heavy workload: More than 1,000 tps

For the purpose of this use case, assume that we have a local 1 TB database, OLTP workload of 1,000 tps, and each OLTP transaction generates about 25 physical storage I/O requests on average. This scenario in turn translates into 25,000 IOPS of physical storage I/O requests that we should get from our storage system to ensure acceptable response time and balanced performance. Because IOPS directly depend on the number of drives used in a traditional HDD-based system we must ensure that we have a sufficient number of HDDs to support 25,000 IOPS. We assume that a single HDD is capable of 300 IOPS with OLTP workload. We also assume that the single four-socket x3850 X5 is capable of processing about 3,000 transactions of this kind per second.

Figure 4 illustrates this scenario using two approaches: achieving required IOPS with traditional HDDs and with IBM eXFlash.

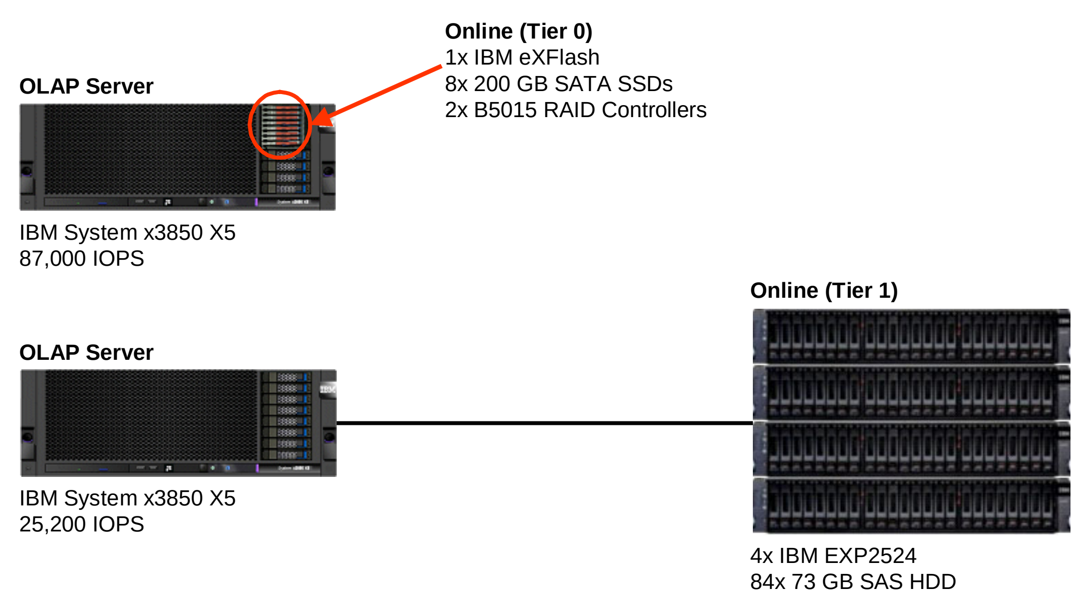

Figure 4 IBM eXFlash versus traditional HDDs: OLTP databases

Using the traditional HDD-based approach we need to deploy one x3850 X5 connected to four IBM System Storage EXP2524 with 84 hard drives by IBM ServeRAID M5025 SAS/SATA Controller.

With IBM eXFlash, which is capable of up to 87,000 IOPS for OLTP workloads, there is no need for external disk storage at all because all I/O performance requirements are met internally. Although both configurations can handle 25,000 IOPS, the eXFlash-based configuration has much higher IOPS capacity of 87,000 IOPS, whereas the HDD-based configuration is just about at its maximum of 25,200 IOPS for the given number of drives. In addition, the response time of the IBM eXFlash-based configuration is significantly better.

As you can see, IBM eXFlash is able to significantly reduce power and cooling requirements, occupied rack space, and management and operational costs, all while providing better reliability and the same or better performance levels. Table 9 summarizes characteristics of each scenario and highlights IBM eXFlash advantages.

Table 9 Traditional HDDs versus IBM eXFlash: Local OLTP workload scenario

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Number of drives	84	8	Less components to acquire and maintain, higher reliability, server-based management
Drive type	73 GB 15K 2.5” SAS HDD	200 GB 1.8” SATA SSD	
Location	External	Internal	
Maximum IOPS capacity	25,200	87,000	Transparent performance scaling at no additional cost, significantly better IOPS/$ ratio, ~3 times faster response time
Used IOPS capacity	25,000	25,000	
Processing time (25,000 I/O requests)	1 sec	0.29 sec	

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Raw storage capacity	6.1 TB	1.6 TB	Significantly better storage space utilization, no wasted storage space, comparable utilized GB/$ ratio
Effective RAID capacity	5.3 TB	1.2 TB	
Used capacity	1 TB	1 TB	
Power consumption units of measurement	kW	W	99% savings in power and cooling costs, no external power infrastructure required
Rack U space	12U	4U	No additional rack space required
Acquisition costs	Comparable to eXFlash	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

To ensure the high availability requirements are met in case of a node failure, several techniques can be utilized depending on the database vendor. These techniques include:

- Log shipping
- Replication
- Database mirroring

The partitioning feature of many databases (for example, IBM DB2®) can help to split the workload between several nodes, thereby increasing overall performance, availability, and capacity.

If, for some reason, the entire database cannot be placed onto IBM eXFlash, consider putting at least part of the data there. The areas to look at include:

- Log files
- Temporary table space
- Frequently-accessed tables
- Partition tables
- Indexes

Some databases (for example, Oracle) support extension of their own data buffers to the SSDs, which provides significant cost-efficient performance increase.

There are more complex configurations, where tiered storage is implemented with both internal and external storage, and the data is moved between storage tiers based on defined policies. The data movement can be implemented manually using the available database tools (like the DB2 Reorg utility in DB2) or automatically using the appropriate storage management software, for example, IBM GPFS.

Data warehouses

Data warehouses are commonly used with online analytical processing (OLAP) workloads in decision support systems, for example, financial analysis. Unlike OLTP, where transactions are typically relatively simple and deal with small amounts of data, OLAP queries are much more complex and process large volumes of data. By its nature, OLAP workload is sequential read-intensive and throughput-intensive; however, in multipurpose multi-user environments it becomes truly random, and therefore, sensitive to IOPS given the I/O request size.

OLAP databases are normally separated from OLTP databases, and OLAP databases consolidate historical and reference information from multiple sources. Queries are submitted to OLAP databases to analyze consolidated data from different points of view to make better business decisions in a timely manner.

For OLAP workloads, it is critical to have a fast response time to ensure that business decisions support an organization's strategy and are made in a timely manner in response to changing market conditions; delay might significantly increase business and financial risks. Because of that, storage I/O capabilities must match the performance of other server subsystems to ensure that queries are processed as quickly as possible.

For illustration purposes, consider the following scenario. Multiple business analysts need to evaluate current business performance and discover new potential opportunities. They submit ten queries, and their queries need to cumulatively process 500 GB of historical data. Possible approaches to implement this solution are shown on Figure 5.

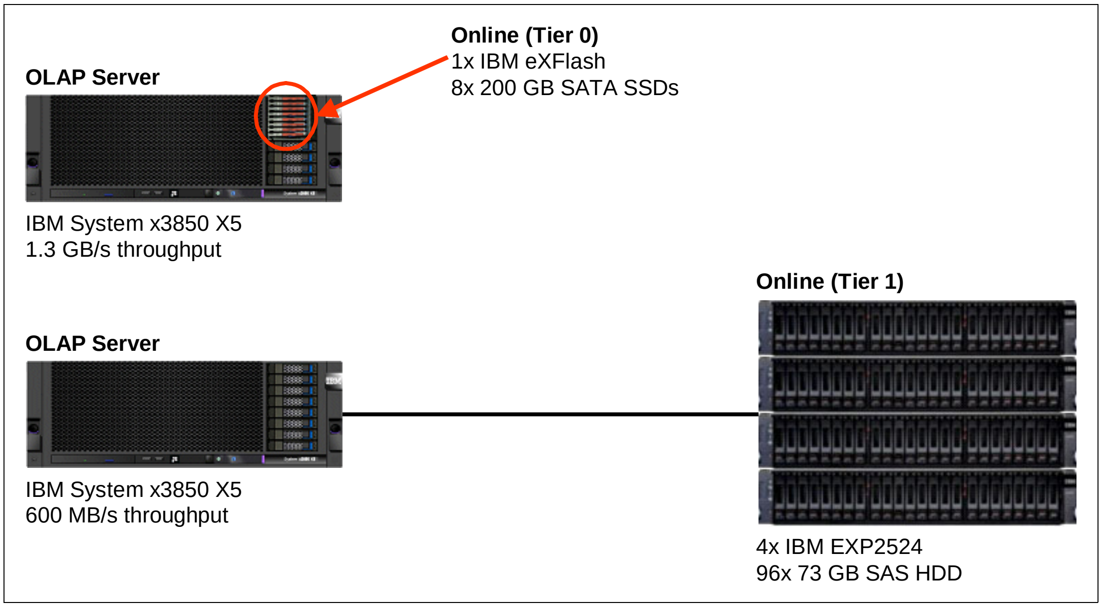

Figure 5 IBM eXFlash versus traditional HDDs: Data warehouses

In one case, the storage system used with an OLAP server consists of 96 hard disk drives and is able to deliver 600 MBps of throughput in RAID-5 arrays with random 16 KB I/O requests. Given this, the queries will be completed by the storage system in approximately 14 minutes, or 1.4 minutes per query on average.

Alternatively, with two IBM eXFlash units, where each of them is capable of approximately 1,300 MBps of throughput in a RAID-5 array with random 16 KB I/O requests, the time to complete the tasks decreased by more than four times, to 3.2 minutes or about 20 seconds per query. IBM eXFlash also provides additional benefits and advantages at comparable acquisition costs, as shown in Table 10.

Table 10 Traditional HDDs versus IBM eXFlash: Local OLAP workload scenario

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Number of drives	96	16	Fewer components to acquire and maintain, higher reliability, server-based management
Drive type	73 GB 15K 2.5" SAS HDD	200 GB 1.8" SATA SSD	
Location	External	Internal	

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Throughput, MBps	600	2,600	Significantly faster query execution
Processing time, sec	84	20	
Raw storage capacity	7.0 TB	3.2 TB	Better storage space utilization, less wasted storage space, comparable utilized GB/$ ratio
Effective RAID capacity	6.4 TB	2.8 TB	
Used capacity	500 GB	500 GB	
Power consumption units of measurement	kW	W	99% savings in power and cooling costs, no external power infrastructure required
Rack U space	12U	4U	No additional rack space required
Acquisition costs	Comparable to eXFlash	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

Corporate email

Corporate email applications like IBM Lotus® Domino® or Microsoft Exchange use databases to store messages and attachments, logging features for recovery purposes, and indexes for fast searching of data. The email workload is multi-user, random, storage I/O-intensive, and characterized by moderate to low numbers of small read and write I/O requests per second.

Although the storage IOPS are relatively low for email workloads compared to OLTP, they play an important role in increasing the utilization of the system resources and providing fast response time for the users working with email.

Consider the following scenario. You need to deploy 10,000 active mailboxes across an organization, each with a 250 MB disk space quota, and you expect heavy email exchange resulting in about 8,000 peak storage IOPS. You also decided to implement local-only storage for email servers, and you use replication of data between servers for high availability purposes.

We assume that a single IBM System x3690 X5 server can host up to 10,000 active mailboxes, and a single 2.5” 15K RPM SAS HDD is capable of 300 IOPS.

Given the scenario with traditional HDDs you need four IBM x3690 X5 servers to meet the requirements outlined previously. In such a case, each server can host up to 5,000 mailboxes because the number of mailboxes is limited by the capabilities of local HDD storage (approximately 4,000 IOPS with 14 HDDs). The storage space required for 5,000 mailboxes is approximately 1.3 TB. During normal operations, each server hosts 2,500 mailboxes (2,000 IOPS), and the remaining capacity is reserved for a failover scenario where the server would pick up workload from a failed node.

With IBM eXFlash, one IBM x3690 X5 server can host 10,000 mailboxes because storage I/O is no longer a limiting factor. In this case you need two x3690 X5 servers with two eXFlash units in each server, and each server will support 5,000 mailboxes during normal operations, and can accommodate 10,000 mailboxes in case of failover.

These scenarios are illustrated on Figure 6.

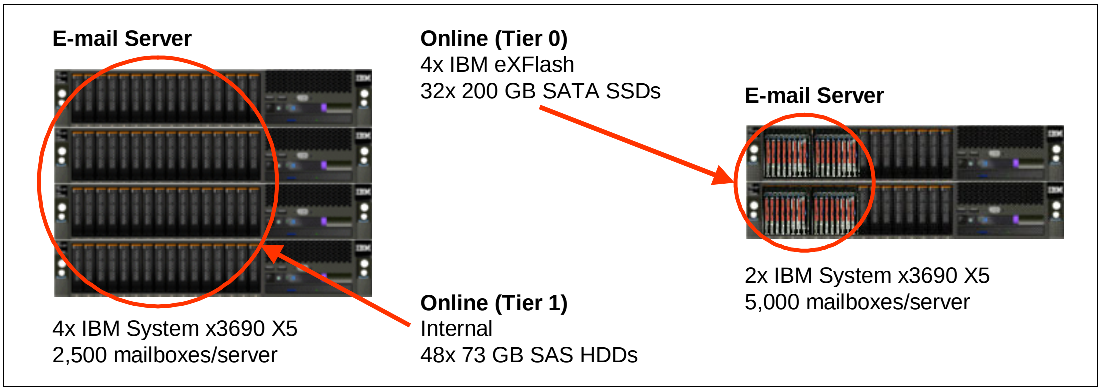

Figure 6 IBM eXFlash versus traditional HDDs: Corporate email

Table 11 summarizes the characteristics of different local online storage approaches.

Table 11 Traditional HDDs versus IBM eXFlash: Local email workload scenario

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Number of servers	4	2	Fewer components to acquire and maintain, higher reliability, lower management, maintenance, and support costs
Number of drives	56	32	
Drive type	146 GB 15K 2.5” SAS HDD	200 GB 1.8” SATA SSD	
Location	Internal	Internal	
Maximum IOPS	16,000	174,000	Significantly faster response time (~10 times) and higher IOPS capacity
Processing time (8,000 I/O requests)	0.5 sec	0.05 sec	
Raw storage capacity	8.2 TB	6.4 TB	Better storage space utilization, less wasted storage space, comparable utilized GB/$ ratio
RAID-5 capacity	7.7 TB	5.6 TB	
Used capacity	5.0 TB	5.0 TB	
Storage power consumption units of measurement	W	W	Three times lower power and cooling costs
Rack U space	8U	4U	Less rack space required
Acquisition costs	Comparable to eXFlash[a]	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

[a] This includes the acquisition cost of additional server hardware

Actively connected users

In a shared collaborative environment, where many users work together, they might produce and operate with a lot of structured and unstructured content, and this requires large storage capacity and throughput and IOPS performance. This is especially true for large Web 2.0 deployments with hundreds, thousands, or even millions of users who participate in online gaming, photo and video sharing, social networking, and other activities through the web interface. This workload is highly random, both read- and write-intensive, and extremely I/O-intensive. In addition, because of the highly heterogeneous nature, workloads can be difficult to predict, and they also require a large amount of data to be stored.

Usually, to achieve the best response time and to meet capacity requirements, the storage systems for such workloads are deployed using multi-tiered scale-out storage architecture with automated storage tiering management. This approach allows the provision of petabytes of storage capacity and hundreds of thousands of operations per second with dozens of gigabytes per second of throughput to maintain fast response times for the users.

Consider this scenario. An organization provides the following online services for their web-based users: gaming, messaging, chatting, and photo and video sharing. The total number of active users is 250,000, and there are 50,000 concurrent users that generate 25,000 storage I/O requests per second. Each user has a storage space quota of 1 GB. There are four IBM System x3690 X5 servers with load balancing that process user requests, assuming each server is capable of hosting up to 12,500 concurrent users.

IBM Storwize V7000 is used as an external disk storage system. The nearline storage tier is built with 144x 2 TB 7.2K RPM 3.5” SATA HDDs providing 288 TB of total space in a RAID-5 configuration (12 RAID-5 arrays of 12 HDDs each). The online storage tier is built with 96x 73 GB 15K RPM 2.5” SAS HDDs providing 6.4 TB of active storage space (eight RAID-5 arrays with 12 HDDs each and assuming that concurrent users actively work with 100 MB of data (10% of quota) for the total of 5 TB for 50,000 concurrent users) and up to 28,800 IOPS to meet concurrent IOPS requirements. IBM GPFS is used as a scale-out file system, and it also provides policy-based automated storage tier management.

With IBM eXFlash, the nearline storage tier is still based on 144 SATA HDDs with IBM Storwize V7000, but the online tier is implemented with eight internal 200 GB eXFlash SSDs installed in each of four IBM System x3690 X5 servers. This provides 5.6 TB of tier 0 online storage (1.4 TB with RAID-5 per server), and 174,000 IOPS (43,500 IOPS per server with one IBM ServeRAID B5015 Controller). Automated tier management and scale-out capabilities are also provided by IBM GPFS.

Figure 7 on page 23 illustrates these use cases.

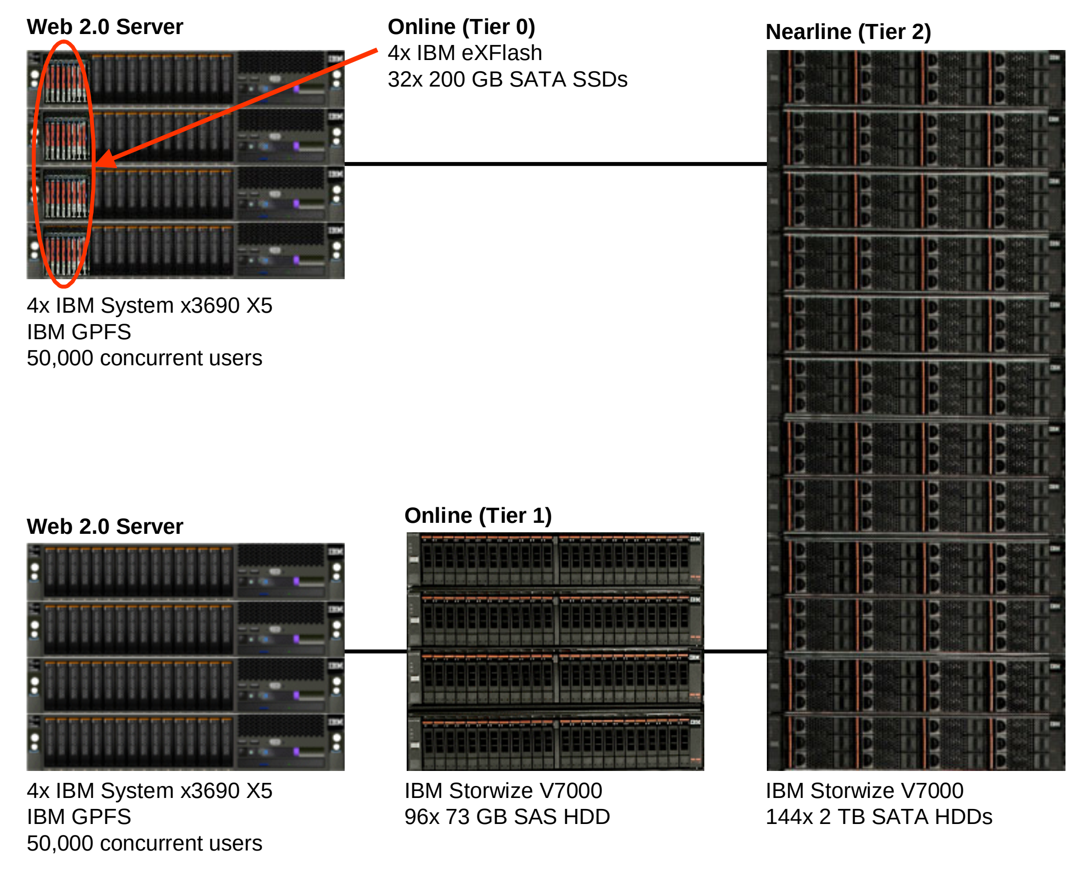

Figure 7 IBM eXFlash versus traditional HDDs: Actively connected users

Table 12 compares an IBM eXFlash-based approach with traditional HDD disk systems.

Table 12 Traditional HDDs versus IBM eXFlash: Actively connected users

Characteristic	**Traditional HDD**	**IBM eXFlash**	**IBM eXFlash advantage**
Number of servers	4	4	Fewer components to acquire and maintain; higher reliability; lower management, maintenance, and support costs
Number of online drives	96	32	
Online drive type	73 GB 15K 2.5” SAS HDD	200 GB 1.8” SATA SSD	
Location	External	Internal	
Maximum IOPS	28,800	174,000	Significantly faster response time (6 times faster) and higher IOPS capacity
Storage processing time (25,000 I/O requests)	0.87 sec	0.14 sec	
Raw storage capacity	7.0 TB	6.4 TB	Better storage space utilization, less wasted storage space, comparable utilized GB/$ ratio
RAID-5 capacity	6.4 TB	5.6 TB	
Used capacity	5.0 TB	5.0 TB	
Storage power consumption units of measurement	kW	W	99% savings in power and cooling costs, no external power infrastructure required

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Rack U space (Tier 0/1)	8U	0U	Less rack space required
Acquisition costs	Comparable to eXFlash	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

Medical imaging

Medical imaging is widely used for diagnostics purposes, and it includes many sources of such information, for example, magnetic resonance imaging, computed tomography, digital X-ray, positron emission tomography, ultrasound, and digital cardiology. All this data must be stored for a long period of time, and this requires large storage space, sometimes petabytes of digital data, because the study sizes can range from dozens of megabytes to several hundred megabytes. At the same time, for faster diagnostics, there is a need to quickly get the required images when needed.

Medical imaging is a multi-user random read-intensive workload. The key performance goal is to achieve high throughput in MBps with random reads to ensure the quick delivery of current information. The tiered storage design model with scale-out approach and automated tier management capabilities fits well to support such workload because it provides required throughput and response time together with high data storage capacity. In this case, after the images are acquired they are placed on Tier 0 or Tier 1 storage (or short-term cache) to be ready for examination by diagnosticians. After a certain period, for example when the patient has been treated, these images are moved to the nearline tier (long-term cache). In addition, when patients are scheduled to visit the doctor by appointment, their studies can be retrieved from the long-term cache and put into short-term cache if needed.

Consider the following scenario. There are 500 patients treated every day in the hospital, 50 patients are treated by doctors at the same time, and the average patient’s study is 100 MB. The throughput required from the Picture Archiving and Communication System (PACS) server to fulfill doctors’ requests for studies should be 5 GBps to deliver the data in one second or 2.5 GBps to deliver the data in two seconds. The space required to store the short-term data is 5 GB, and about 1.5 TB of date is added to the medical archive on a monthly basis.

The main archive is based on nearline storage consisting of IBM System Storage N6270 with 200x 2 TB SATA drives providing 400 TB of available storage space. If the short-term cache is built on traditional HDDs, then 240 HDDs are required to achieve 5 GBps or 120 HDDs to achieve 2.5 GBps. With IBM eXFlash, you need three units to meet 5 GBps throughput, or two units to meet 2.5 GBps.

These scenarios are shown on Figure 8.

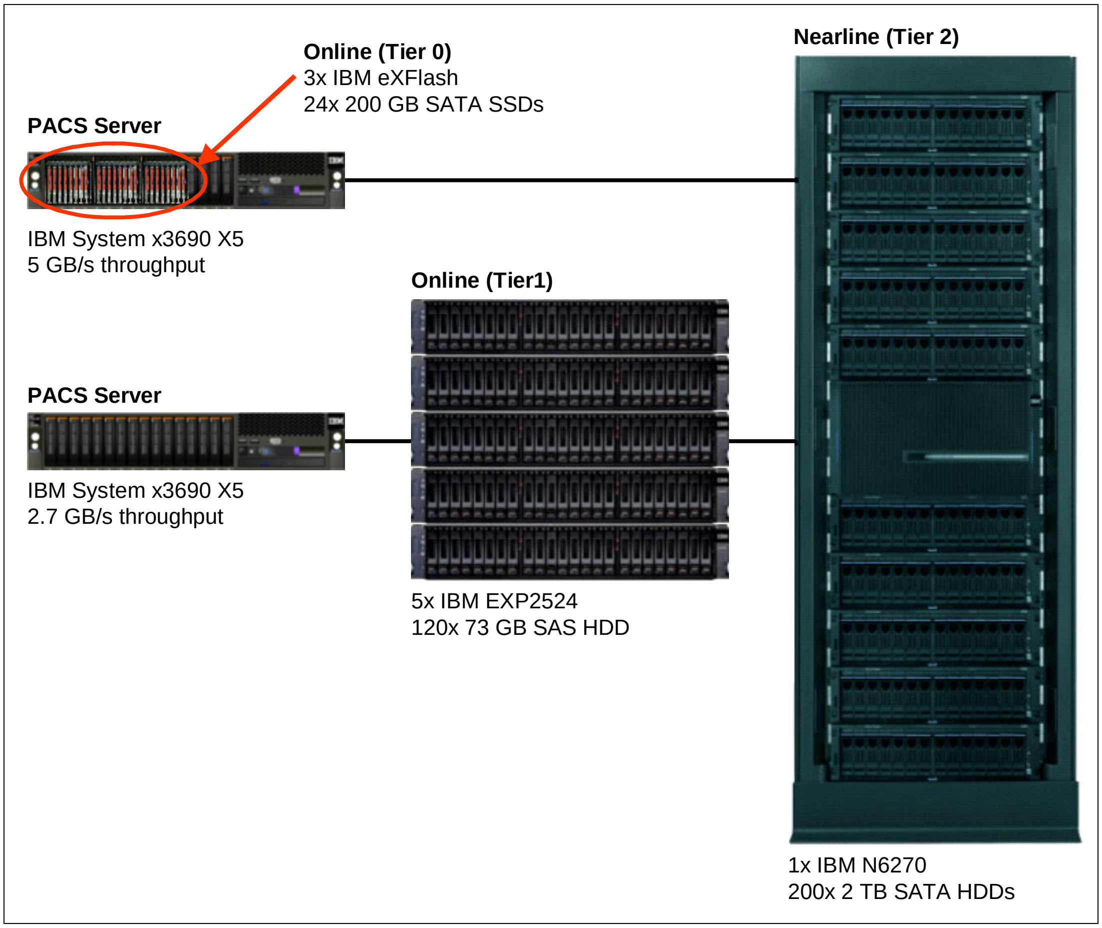

Figure 8 IBM eXFlash versus traditional HDDs: Medical imaging

Table 13 summarizes the characteristics of each scenario.

Table 13 Traditional HDDs versus IBM eXFlash: Medical imaging

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Number of servers	1	1	Fewer components to acquire and maintain; higher reliability; lower management, maintenance, and support costs
Number of online drives	120	24	
Online drive type	73 GB 15K 2.5” SAS HDD	200 GB 1.8” SATA SSD	
Location	External	Internal	
Maximum throughput	2.7 GBps	6 GBps	Higher throughput capacity, faster transfer rates
Study load time	1.9 sec	0.8 sec	
Raw storage capacity	8.8 TB	4.8 TB	Better storage space utilization, less wasted storage space, comparable utilized GB/$ ratio
RAID-5 capacity	8.0 TB	4.2 TB	
Used capacity	5 GB	5 GB	

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Storage power consumption units of measurement	kW	W	99% savings in power and cooling costs, no external power infrastructure required
Rack U space (Tier 0/1)	10U	0U	Less rack space required
Acquisition costs	Comparable to eXFlash	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

Video on demand

While video on demand is traditionally sequential throughput-intensive workload, in a multi-user environment, where every user receives their own data stream watching different content or even the same content with some delay (for example, a recently published new movie), the workload becomes randomized, and this requires faster response time to ensure a better user experience and smoother video playback. In general, video streaming applications use 64 KB I/O blocks to interact with the storage system.

As with medical imaging, video libraries require a significant amount of storage space and sufficient throughput. This can be achieved using a tiered storage design approach with automated tier management capabilities. With such an approach, movie libraries reside on nearline storage, and the movies currently being watched are placed on the online (Tier 0 or Tier 1) storage.

Consider the following scenario. A provider of on demand video content has 50,000 subscribers, and 10,000 movies in their video library. Five thousand subscribers are active at the same time, and they watch 1,000 videos simultaneously (that is, one video is watched by five subscribers on average). The provider uses SD video that requires about 4 Mbps or 0.5 MBps per stream, and the average movie size is 4 GB.

Let’s assume that a single IBM System x3690 X5 server used in the scenario is capable of handling 5,000 simultaneous video streams. The total throughput for 5,000 concurrent streams is 2.5 GBps. To store 1,000 videos we need 4 TB of storage space.

For the main video library storage, we use a nearline tier with IBM System Storage DS3524 with 48x 1 TB 7.2K RPM 2.5” SATA HDDs configured as a RAID-5 array. For the online tier we can choose either SAS HDDs (Tier 1) or IBM eXFlash with SSDs (Tier 0), as shown on Figure 9 on page 27.

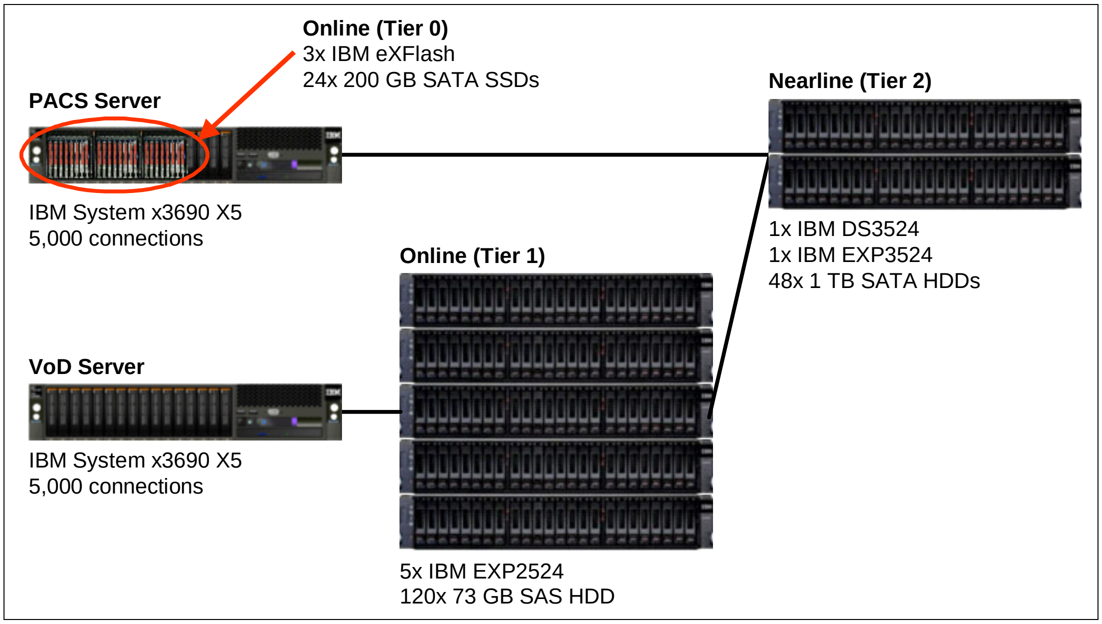

Figure 9 IBM eXFlash versus traditional HDDs: Video on demand

With the traditional HDD scenario, to achieve such a throughput of 2.5 GBps with random I/O reads and 4 TB of storage space we need about 120x 73 GB 15K RPM SAS HDDs in Tier 1 storage. With the IBM eXFlash scenario, we need three eXFlash units to meet the capacity requirements, and we also meet the throughput requirement because a single eXFlash is capable of 2 GBps of read throughput. Table 14 summarizes the characteristics of these scenarios and highlights IBM eXFlash advantages.

Table 14 Traditional HDDs versus IBM eXFlash: Video on demand

Characteristic	Traditional HDD	IBM eXFlash	IBM eXFlash advantage
Number of servers	1	1	Fewer components to acquire and maintain; higher reliability; lower management, maintenance, and support costs
Number of online drives	120	24	
Online drive type	73 GB 15K 2.5” SAS HDD	200 GB 1.8” SATA SSD	
Location	External	Internal	
Maximum throughput	2.7 GBps	6 GBps	Higher throughput capacity, faster transfer rates
Raw storage capacity	8.8 TB	4.8 TB	Better storage space utilization, less wasted storage space, comparable utilized GB/$ ratio
RAID-5 capacity	8.0 TB	4.2 TB	
Used capacity	4.0 TB	4.0 TB	
Storage power consumption units of measurement	kW	W	99% savings in power and cooling costs, no external power infrastructure required
Rack U space (Tier 0/1)	10U	0U	Less rack space required
Acquisition costs	Comparable to eXFlash	Comparable to HDD	Higher scalability, reliability, and performance, and significantly lower power consumption at the comparable acquisition cost

Summary

We described several scenarios with different workload patterns where IBM eXFlash was able to provide obvious benefits compared to traditional HDD-based approaches. In summary, IBM eXFlash helps to:

- Significantly decrease implementation costs (up to 97% lower) of high performance I/O-intensive storage systems with best IOPS/$ performance ratio
- Significantly increase performance (up to 30 times or more) of I/O-intensive applications like databases and business analytics
- Significantly save on power and cooling (up to 90%) with a high performance per watt ratio
- Significantly save on floor space (up to 30 times less) with extreme performance per rack U space ratio

In addition, the majority of current systems use a tiered storage approach, where "hot" data is placed close to the application on the online storage (Tier 0 or Tier 1), "warm" data is placed on the nearline storage (Tier 2), and "cold" data is placed on the offline storage (Tier 3). In such a case, IBM eXFlash can be utilized as Tier 0 storage.

In general, typical IBM eXFlash usage scenarios include:

- High-speed read cache in a local or SAN-based storage environment
- Temporary local storage space for mid-tier applications and databases
- Main (Tier 0) local data storage in single server environments or in a distributed scale-out environment with local-only storage or mixed local and SAN-based storage

The author who wrote this paper

This paper was produced by a technical specialist working at the International Technical Support Organization, Raleigh Center.

Ilya Krutov is an Advisory IT Specialist and project leader at the International Technical Support Organization, Raleigh Center, and has been with IBM since 1998. Prior roles in IBM included STG Run Rate Team Leader, Brand Manager for IBM System x and BladeCenter, Field Technical Sales Support (FTSS) specialist for System x and BladeCenter products, and instructor at IBM Learning Services Russia (IBM x86 servers, Microsoft NOS, Cisco). He graduated from the Moscow Engineering and Physics Institute, and holds a Bachelor's degree in Computer Engineering.

Portions of this paper were taken from *IBM eX5 Portfolio Overview: IBM System x3850 X5, x3950 X5, x3690 X5, and BladeCenter HX5*, REDP-4650. Thanks to the authors of that document:

- David Watts
- Duncan Furniss
- Scott Haddow
- Jeneea Jervay
- Eric Kern
- Cynthia Knight

Thanks to the following for his contributions to this project:

- Paul Nashawaty

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-4807-00 was created or updated on December 9, 2011.

Send us your comments in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

BladeCenter®
DB2®
Domino®
DS8000®
Dynamic Infrastructure®
Easy Tier™
GPFS™
IBM®
Lotus®
Redbooks®
Redpaper™
Redbooks (logo) ®
ServerProven®
Smarter Planet™
Storwize®
System Storage®
System x®
X-Architecture®
XIV®

The following terms are trademarks of other companies:

Microsoft, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.