Technical Overview of the Lenovo X6 Servers

Last Update: September 2016

Covers the sixth generation Enterprise X-Architecture servers

Support for the latest Intel Xeon E7-8800 v4 "Broadwell EX" processors

Servers scalable up to 8 processors and 12 TB of system memory

Provides technical information about all server features

Ilya Krutov

Click here to check for updates

Abstract

The increasing demand for cloud computing and business analytical workloads to meet business needs drives innovation to find new ways to build informational systems. Clients are looking for cost-optimized fit-for-purpose IT solutions that manage large amounts of data, easily scale performance, and provide reliable real-time access to actionable information.

Built on decades of innovation, Lenovo® introduces the sixth generation of Enterprise X-Architecture technology, Lenovo X6 rack servers. Lenovo X6 servers are designed to be *fast*, *agile*, and *resilient*:

- Fast application performance means immediate access to actionable information.
- Agile system design helps to reduce acquisition costs and provide the ability to host multiple generations of technology in a single server.
- Resilient platforms maximize application uptime and promote easy integration in virtual environments.

Lenovo X6 servers continue to lead the way as the shift toward mission-critical scalable databases, business analytics, virtualization, enterprise applications, and cloud accelerates.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Contents

Introduction to Lenovo X6

The Lenovo X6 product portfolio represents the sixth generation of servers that are built upon Enterprise X-Architecture. Enterprise X-Architecture is the culmination of generations of Lenovo technology and innovation that is derived from the experience in high-end enterprise servers. Now, with the X6 servers, Lenovo scalable systems can be expanded on demand and configured by using a building block approach that optimizes system design for your workload requirements. These servers scale to more processor cores, memory, and I/O than previous systems, enabling them to handle greater workloads than the systems that they supersede. Power efficiency and server density are optimized, making them affordable to own and operate.

Lenovo X6 systems allow your enterprise to grow in processing, input/output (I/O), and memory dimensions. Therefore, you can provision what you need now and expand the system to meet future requirements. System redundancy and availability technologies are more advanced than those previously available in the x86 systems.

The X6 portfolio increases performance and virtualization density while decreasing infrastructure costs and complexity. This function enables you to design faster analytics engines, reign in IT sprawl, and deliver information with high reliability. Lenovo X6 servers are fast, agile, and resilient.

The Lenovo X6 rack server portfolio consists of the following flagship servers of the Lenovo x86 server family:

- Lenovo System x3850 X6
- Lenovo System x3950 X6

Fast application performance

The x3850 X6 and x3950 X6 servers deliver fast application performance thanks to an innovative scalable design and new storage technology that is designed to optimize overall solution performance.

With the new Intel Xeon processor E7-4800 v4 and E7-8800 v4 product families, the x3850 X6 and x3950 X6 servers can deliver up to 6.0 TB or 12 TB of memory and 96 or 192z cores of processing power, respectively. Armed with these capabilities, you can host essential mission-critical applications, implement large virtual machines, or run sizeable in-memory databases without compromises in performance, capacity, or scalability.

Agile design characteristics

Change is inevitable, and managing change is a must to achieve or to maintain market leadership. Changes in IT infrastructure typically drive complexity and cost. Managing an evolving technology, divergent customer needs, and fluctuating costs requires an agile approach to platform design. Having flexible systems to create fit-for-purpose solutions is essential.

The unique, adaptive modular rack design of the X6 family delivers agility that enables you to design a solution that meets your needs. At the same time, you can realize infrastructure cost savings by hosting multiple generations of technology in a single platform, without compromising performance or capacity.

© Copyright Lenovo 2016. All rights reserved.

The X6 platforms provide the following capabilities:

- You can configure the server to fit the unique requirements of your applications and workloads. You can add, modify, or upgrade the X6 platforms easily with selectable modular Book components from each of the types of the X6 Books, one for each of the major subsystems (that is, storage, compute, and I/O).
- You can scale capacity and performance from 4-socket to 8-socket to deliver twice the performance for growing applications without creating IT sprawl.
- You can use Lenovo XClarity Administrator software for automated provisioning of a cluster of servers and realize time-to-value in minutes rather than days.
- You can realize agile system design that is ready to host multiple generations of technology in a single server. For existing X6 customers with E7 v2 or E7 v3 processors, Lenovo offers a simple upgrade path that simply involves replacing the compute books with the latest E7 v4 processor technology. All other components stay the same.

Resilient enterprise platforms

The growth of new applications has elevated database processing and business analytics to the top of the list of crucial x86 workloads for enterprise businesses. These environments demand continuous uptime to rapidly achieve the most valuable result-massive amounts of business-critical data. The enterprise platforms that host these workloads must deliver data at a high velocity and with continuous availability.

Through differentiated the X6 self-healing technology, the X6 servers maximize uptime by proactively identifying potential failures and transparently taking necessary corrective actions. The X6 servers include the following unique features:

- Advanced Page Retire proactively protects applications from corrupt pages in memory, which is crucial for scaling memory to terabytes.
- Advanced Processor Recovery allows the system to automatically switch access and control of networking, management, and storage in the event of a Processor 1 failure, which provides higher availability and productivity.
- Upward Integration Modules for standard hypervisors enable the creation and management of policies to maintain high availability of virtual machines and concurrent updating of the system firmware, with no impact on application performance or availability.
- The X6 modular design reduces service time by enabling quick easy replacement of failed components.

These built-in technologies drive the outstanding system availability and uninterrupted application performance that is needed to host business-critical applications.

Lenovo X6 systems overview

The Lenovo System x3850 X6 server is a 4U rack-optimized server scalable to four sockets, and the Lenovo System x3950 X6 server is a 8U rack-optimized server scalable to eight sockets. These systems are designed for maximum usage, reliability, and performance for compute-intensive and memory-intensive workloads.

Figure 1 on page 5 shows the Lenovo System x3850 X6 server.

Figure 1 Lenovo System x3850 X6 server

The x3850 X6 server has the following key characteristics:

- Up to four Intel Xeon processor E7-4800 v4 or E7-8800 v4 product family processors
- Up to 96 DIMM slots (24 DIMM slots per processor) for up to 6 TB of memory (using 64 GB DIMMs)
- Up to 1600 MHz DDR3 memory speeds and up to 2667 MHz SMI2 link speeds
- Up to eight 2.5-inch hot-swap drives or up to 16 1.8-inch hot-swap solid-state drives (SSDs)
- Support for 12 Gbps SAS connectivity for the internal storage
- Mezzanine LOM slot for the integrated NIC functionality (choice of dual-port 10 GbE or quad-port 1 GbE adapters)
- Up to 11 PCIe 3.0 I/O slots
- Internal USB port for the embedded hypervisor

The x3950 X6 server looks similar to two x3850 X6 servers, where one is placed on top of the other; however, unlike eX5™ servers, x3950 X6 server employs a single server design with a single backplane, without any external connectors and cables.

Figure 2 shows the Lenovo System x3950 X6 server.

Figure 2 Lenovo System x3950 X6 server

The x3950 X6 server has the following key characteristics:

- Up to eight Intel Xeon processor E7-8800 v4 product family processors
- Up to 192 DIMM slots (24 DIMM slots per processor) for up to 12 TB of memory (using 64 GB DIMMs)
- Up to 1866 MHz TruDDR4 memory speeds
- Up to 16 2.5-inch hot-swap drives or up to 30 two 1.8-inch hot-swap SSDs
- Support for 12 Gbps SAS connectivity for the internal storage
- Two mezzanine LOM slots for the integrated NIC functionality (choice of dual-port 10 GbE or quad-port 1 GbE adapters)
- Up to 22 PCIe 3.0 I/O slots
- Two internal USB ports for the embedded hypervisors

Lenovo X6 systems design and architecture

The X6 systems offer the new “bookshelf” design concept that is based on a fixed chassis mounted in a standard rack cabinet. You do not need to pull the chassis in or out of the rack, because you can access all chassis components either from the front or the rear, similar to pulling books from a bookshelf.

The modular component that can be installed in a server is called a *Book*. The following types of books are available:

- Compute Books

  A Compute Book contains one processor and 24 DIMM slots. It is accessible from the front of the server. Compute Books are described in “Processor subsystem” on page 12.

- Storage Books

  A Storage Book contains standard 2.5-inch or eXFlash 1.8-inch hot-swap drive bays. It also provides front USB and video ports, and it has two PCIe slots for internal storage adapters. It is accessible from the front of the server. Storage Books are described in “Storage subsystem” on page 30.

- I/O Books

  An I/O Book is a container that provides PCIe expansion capabilities. I/O Books are accessible from the rear of the server. I/O Books are described in “Networking and I/O” on page 35.

  The following types of I/O Books are available:

  – The *Primary I/O Book* provides core I/O connectivity, including mezzanine LOM slot for an onboard network, three PCIe slots, Integrated Management Module II, and rear ports (USB, video, serial, and management).

  – The hot-swap Full-length I/O Book provides three optional full-length PCIe slots, and two of them are capable of hosting Graphics Processing Unit (GPU) adapters up to 300 W of total power per Book.

  – The hot-swap *Half-length I/O Book* provides three optional half-length PCIe slots.

Modular components

All modular components are hosted in a 4U (the x3850 X6 server) or 8U (the x3950 X6 server) rack drawer. Two passive backplanes (for 4U and 8U chassis respectively) connect all modular components to each other. Books are the same for 4U and 8U servers.

An x3850 X6 4U server can hosts up to four Compute Books, one Storage Book, one Primary I/O Book, and up to two optional I/O Books. In addition, 4U server supports up to four power supplies and up to 10 hot-swap dual-motor fans (eight fans on the front and two fans on the rear). Figure 3 on page 8 and Figure 4 on page 8 show front and rear components of the x3850 X6 server.

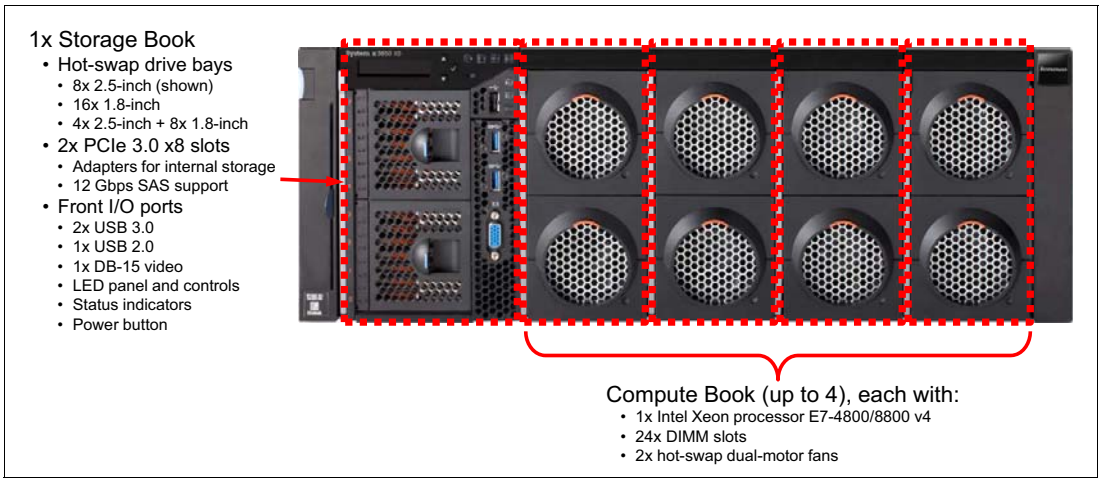

Figure 3 The x3850 X6 server front view

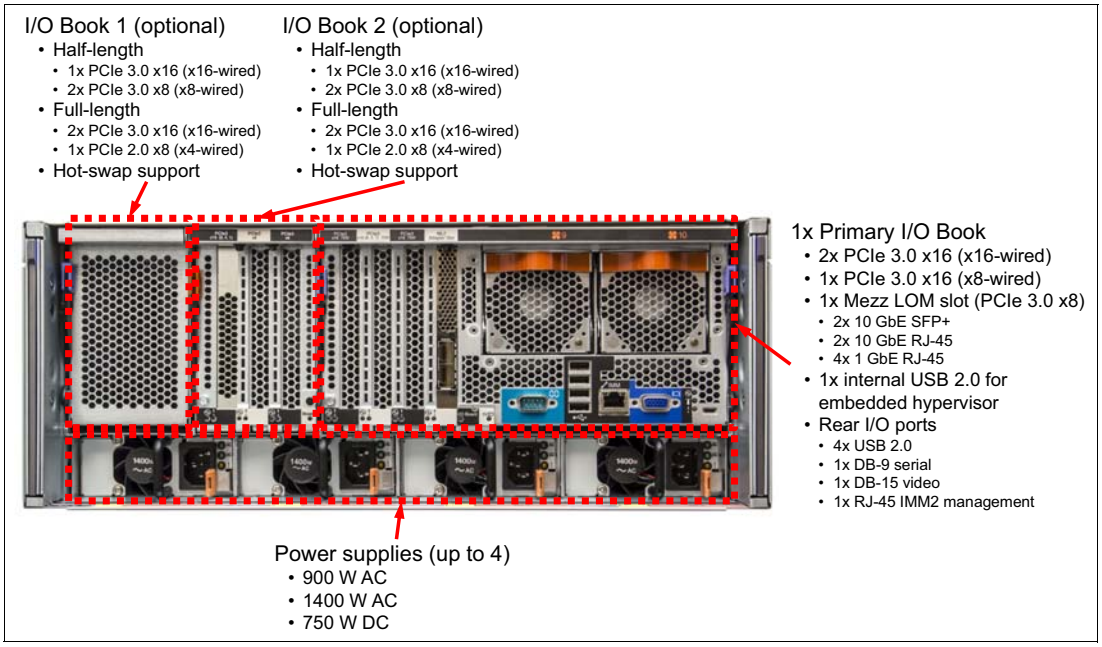

Figure 4 The x3850 X6 server rear view

An x3950 X6 8U server can host up to eight Compute Books (four minimum), two Storage Books, two Primary I/O Books, and up to four optional I/O books. In addition, 8U server supports up to eight power supplies and up to 20 hot-swap dual-motor fans (up to 16 fans on the front and four fans on the rear).

Figure 5 on page 9 and Figure 6 on page 9 show front and rear components of the x3950 X6 server.

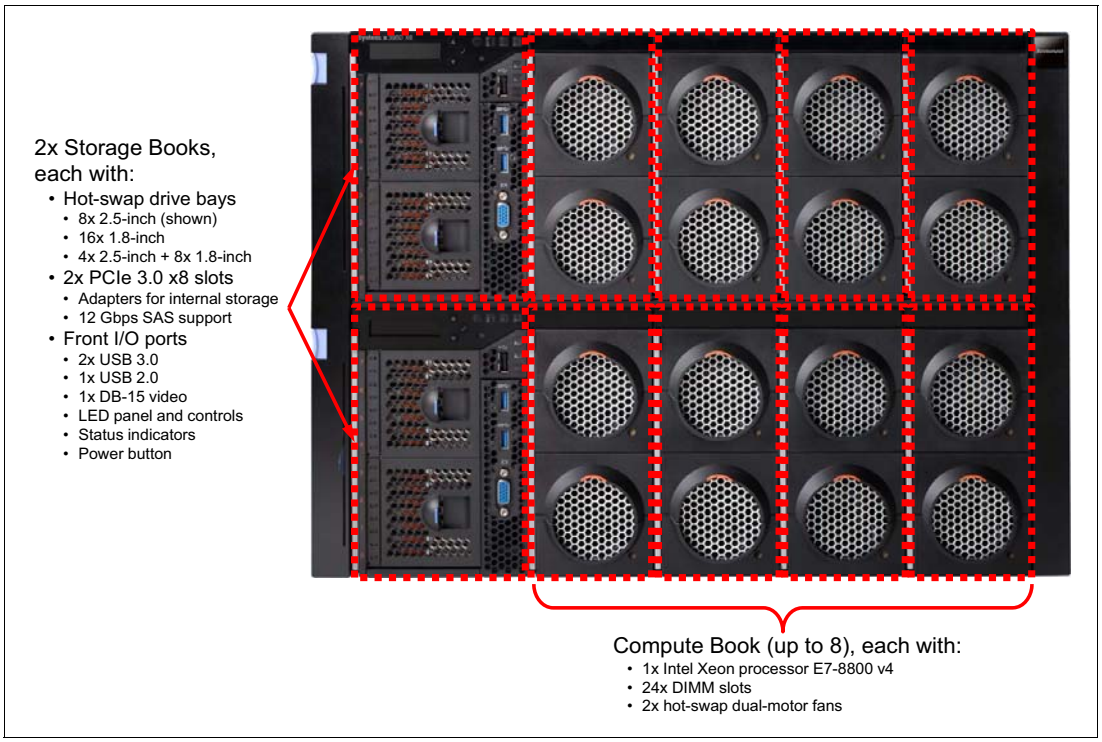

Figure 5 The x3950 X6 server front view

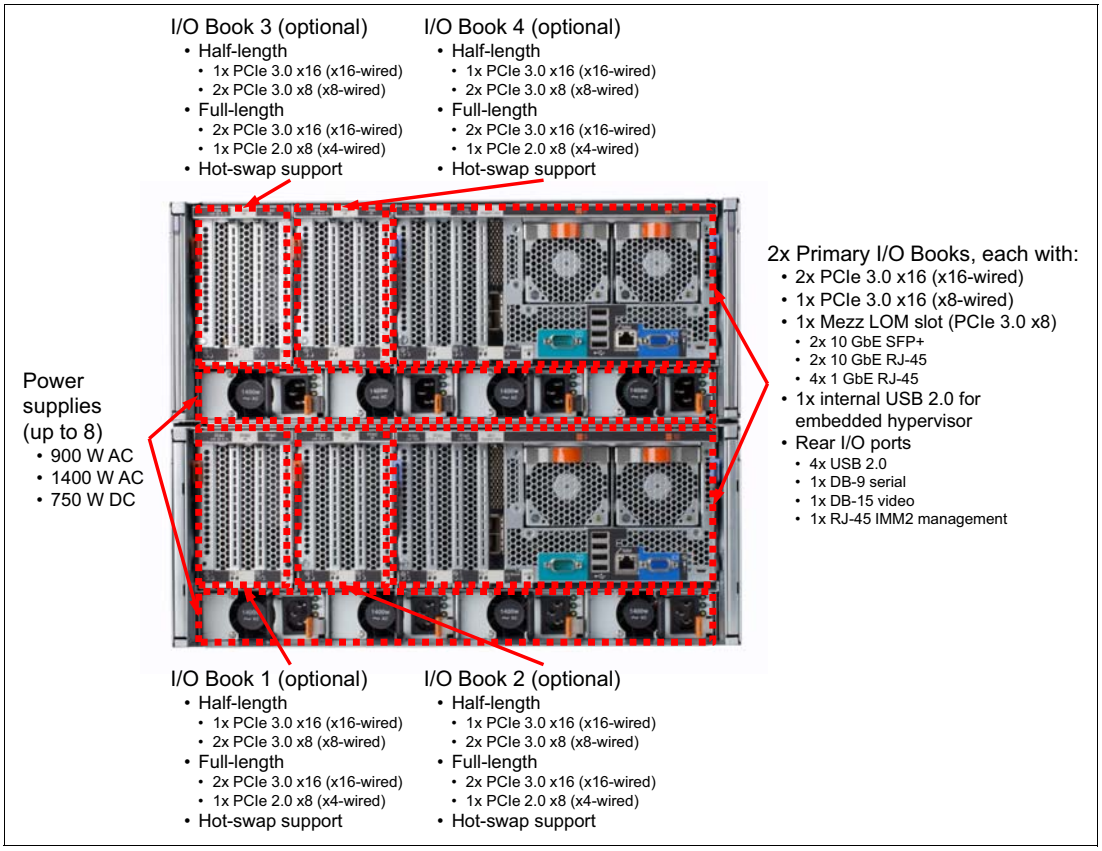

Figure 6 The x3950 X6 server rear view

Systems architecture

Figure 7 shows the system architecture of the x3850 X6 server.

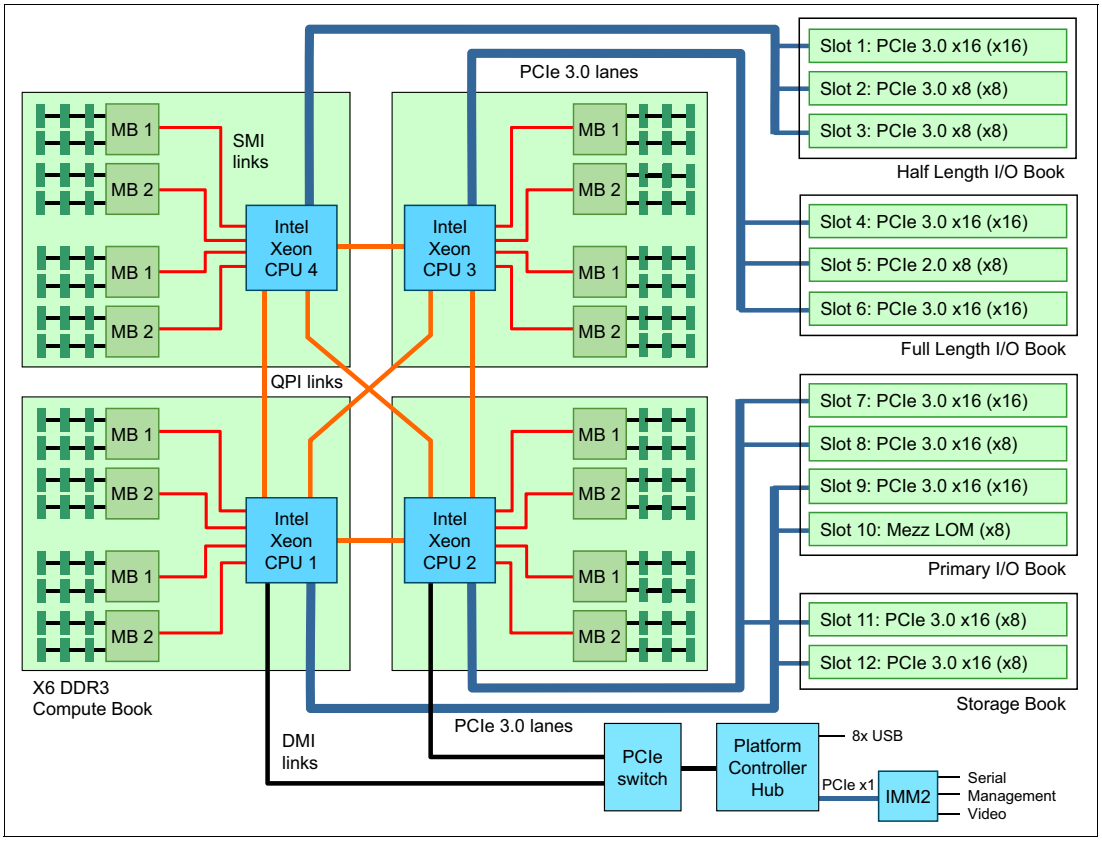

Figure 7 The x3850 X6 system architecture

In the DDR3 X6 Compute Book, each processor has four Scalable Memory Interconnect (SMI) channels (two memory controllers per processor, each with two SMI channels) that are connected to four scalable memory buffers. Each memory buffer (MB) has six DIMM slots (two channels with three DIMMs per channel) for a total of 24 DIMMs (eight channels with three DIMMs per channel) per processor. Compute Books are connected to each other through Quick Path Interconnect (QPI) links.

The Primary I/O Book has three PCIe 3.0 slots, Mezzanine LOM slot, I/O Controller Hub, Integrated Management Module II (IMM2) and peripheral ports (such as USB, video, serial) on the board. Additional I/O Books (Full Length and Half Length) have three PCIe 3.0 slots each and provide ability to hot-add and hot-remove PCIe adapters.

Additional I/O Books: For illustration purposes, Figure 7 shows both Half Length and Full Length I/O Books, where the Half Length I/O Book supplies slots 1, 2, and 3, and the Full Length I/O Book supplies slots 4, 5, and 6. The Half Length I/O Book can also be used to supply slots 4, 5, and 6, and the Full Length I/O Book can also be used to supply slots 1, 2, and 3.

The Primary I/O Book is connected to the Compute Books 1 (CPU 1) and 2 (CPU 2) directly through PCIe links from those processors: PCIe slots 9 and 10 are connected to CPU 1, and PCIe slots 7 and 8 are connected to CPU 2. Also, both CPU 1 and CPU 2 are connected to

the Platform Controller Hub (PCH) through Direct Media Interface (DMI) switched links for redundancy purposes.

The Storage Book is also connected to both Compute Books 1 and 2, however, the PCIe slots 11 and 12 are connected to different processors (CPU 2 and CPU 1 respectively). In addition, certain peripheral ports are routed from the PCH and IMM2 to the Storage Book.

Additional I/O Books are connected to Compute Books 3 and 4 and use PCIe links from CPU 3 and CPU 4. If you need to install additional I/O Book, install the Compute Book in the appropriate slot first.

The System x3850 X6 server is designed so that it can tolerate the failure of CPU 1. This feature decreases downtime in the event of a processor failure, and it requires two Compute Books installed in Compute Book slots 1 and 2. An operating system boot path from CPU 2 is required to successfully disable CPU 1.

When the server detects a failure of CPU 1, the IMM2 reroutes the Platform Controller Hub to CPU 2 instead. The system boots using CPU 2, albeit with reduced functionality. This failover to a surviving processor is automatic and is handled by the system at boot time.

If you want to take advantage of this recovery, it is important to carefully plan your network and storage connectivity, because if CPU 1 fails, PCIe slots 9, 10, and 12 become unavailable.

The x3950 X6 server has the same architecture adapted to 8-socket configuration. Figure 8 and Figure 9 on page 12 shows the system architecture of the x3950 X6 server. (Figure 8 shows the lower half, and Figure 9 on page 12 shows the upper half.)

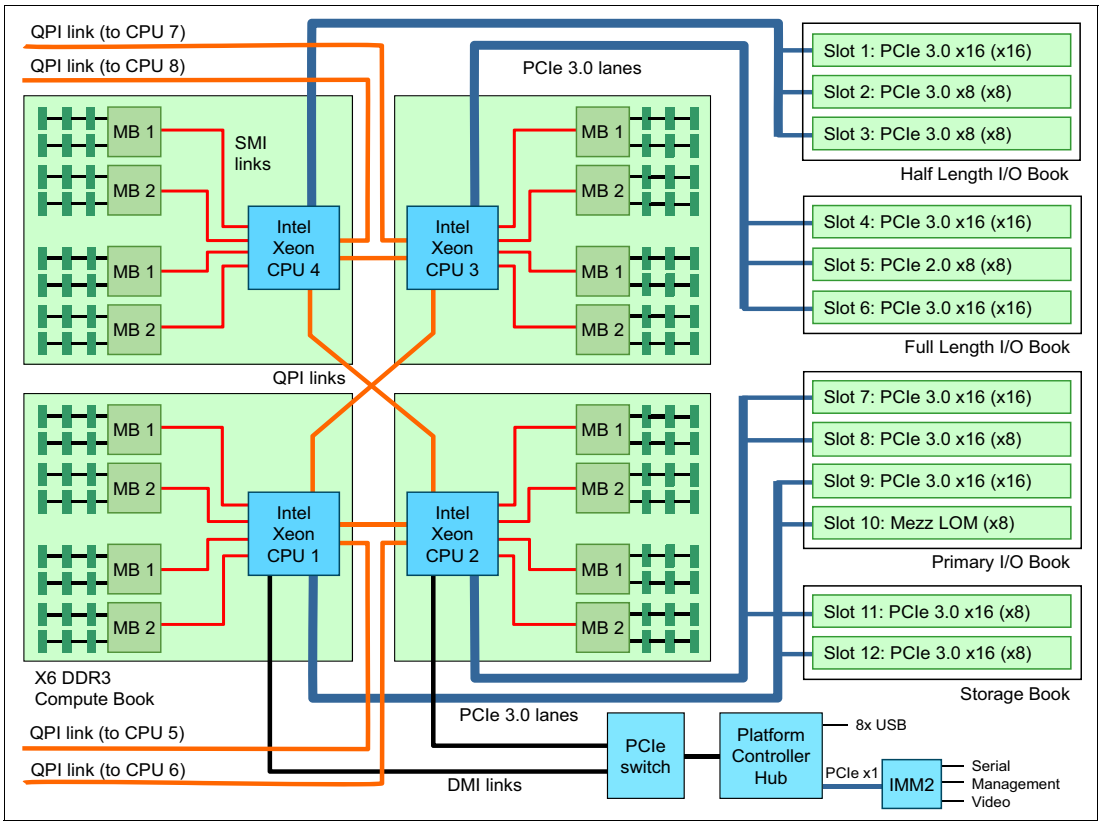

Figure 8 The x3950 X6 system architecture (lower half)

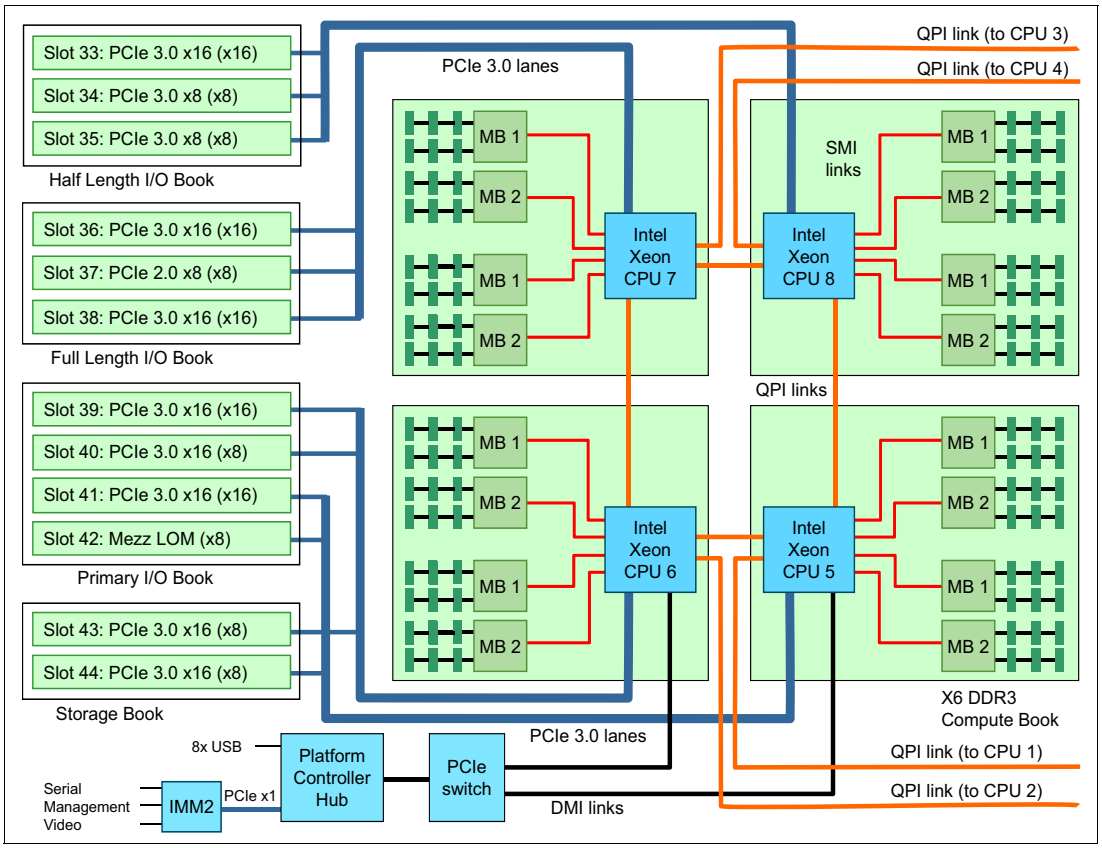

Figure 9 The x3950 X6 system architecture (upper half)

The 8-socket configuration is formed using the native QPI scalability of the Intel Xeon processor E7 family.

In addition, the 8-socket server has the ability to form two independent system that contain four sockets in each node, similar to two independent x3850 X6 servers in one 8U chassis. This partitioning feature can be enabled using the IMM2 interface. While partitioning is enabled, each partition can deploy its own operating system. Each partition uses its own resources and can no longer access the other partition's resources.

Processor subsystem

The current models of the X6 systems use the Intel Xeon E7 processor family. The new Intel Xeon E7 v4 processors feature the new Intel microarchitecture (formerly codenamed "Broadwell-EX") that provides higher core count, larger cache sizes, and DDR4 memory support. Intel Xeon E7 v2, v3 and v4 families support up to 24 DIMMs per processor and provide fast low-latency I/O with integrated PCIe 3.0 controllers.

Intel Xeon processor E7-4800/8800 v4 product family

The following groups of the Intel Xeon processor E7 family are used in the X6 servers:

- The Intel Xeon processor E7-4800/8800 v4 product family is supported in the x3850 X6. This family supports four-processor configurations. (Using E7-8800 v4 processors enables the swapping of Compute Books between x3850 X6 and and x3950 X6 servers.)
- The Intel Xeon processor E7-8800 v4 product family is used in the x3950 X6 to scale to eight-socket configurations.

The X6 systems support the latest generation of the Intel Xeon processor E7-4800 v4 and E7-8800 v4 product family, which offers the following key features:

- Up to 24 cores and 48 threads (by using Hyper-Threading feature) per processor
- Up to 60 MB of shared last-level cache
- Up to 3.2 GHz core frequencies
- Up to 9.6 GTps bandwidth of QPI links
- DDR4 memory interface support, which brings greater performance and power efficiency
- Integrated memory controller with four SMI2 Gen2 channels that support up to 24 DDR4 DIMMs
- Memory channel (SMI2) speeds up to 1866 MHz in RAS (lockstep) mode and up to 3200 MHz in performance mode.
- Integrated PCIe 3.0 controller with 32 lanes per processor
- Intel Virtualization Technology (VT-x and VT-d)
- Intel Turbo Boost Technology 2.0
- Improved performance for integer and floating point operations
- Virtualization improvements with regards to posted interrupts, page modification logging, and VM enter/exit latency reduction
- New Intel Transactional Synchronization eXtensions (TSX)
- Intel Advanced Vector Extensions 2 (AVX2.0) with new optimized turbo behavior
- Intel AES-NI instructions for accelerating of encryption
- Advanced QPI and memory reliability, availability, and serviceability (RAS) features
- Machine Check Architecture recovery (non-running and running paths)
- Enhanced Machine Check Architecture Gen2 (eMCA2)
- Machine Check Architecture I/O
- Resource director technology: Cache monitoring technology, cache allocation technology, memory bandwidth monitoring
- Security technologies: OS Guard, Secure Key, Intel TXT, Crypto performance (ADOX/ADCX), Malicious Software (SMAP), Key generation (RDSEED)

Table 1 compares the Intel Xeon E7-4800/8800 processors that are supported in X6 systems.

Table 1 X6 processors comparisons

Feature	X6 family, Xeon E7 v2	X6 family, Xeon E7 v3	X6 family, Xeon E7 v4
Processor family	Intel Xeon E7-8800 v2 Intel Xeon E7-4800 v2	Intel Xeon E7-8800 v3 Intel Xeon E7-4800 v3	Intel Xeon E7-8800 v4 Intel Xeon E7-4800 v4
Processor codenames	Ivy Bridge EX	Haswell EX	Broadwell EX
Cores per CPU	Up to 15 cores	Up to 18 cores	Up to 24 cores

Feature	X6 family, Xeon E7 v2	X6 family, Xeon E7 v3	X6 family, Xeon E7 v4
Last level cache	Up to 37.5 MB L3 cache	Up to 45 MB L3 cache	Up to 60 MB L3 cache
QPI	QPI 1.1 at 8.0 GT/s max	QPI 1.1 at 9.6 GT/s max	QPI 1.1 at 9.6 GT/s max
CPU TDP rating	Up to 155 W	Up to 165 W	Up to 165 W
DIMM sockets	24 DDR3 DIMMs per CPU	24 DDR3 DIMMs per CPU 24 DDR4 DIMMs per CPU	24 DDR4 DIMMs per CPU
Maximum memory speeds	2667 MHz SMI2	3200 MHz SMI2	3200 MHz SMI2
PCIe technology	PCIe 3.0 (8 GTps)	PCIe 3.0 (8 GTps)	PCIe 3.0 (8 GTps)

Intel Xeon processor E7-4800/8800 v3 product family

Similar to the latest Intel Xeon E7 v4 processors, the Intel Xeon E7 v3 processor family provides two types of processors: E7-4800 v3 and E7-8800 v3 for four- and eight-socket configurations.

The Intel Xeon processor E7-4800 v3 and E7-8800 v3 product family offers the following key features:

- Up to 18 cores and 36 threads (by using Hyper-Threading feature) per processor
- Up to 45 MB of shared last-level cache
- Up to 3.2 GHz core frequencies
- Up to 9.6 GTps bandwidth of QPI links
- Integrated memory controller with four SMI2 channels that support up to 24 DDR3/DDR4 DIMMs
- Up to 1600 MHz DDR3 or 1866 MHz DDR4 memory speeds
- DDR4 memory channel (SMI2) speeds up to 1866 MHz in RAS (lockstep) mode and up to 3200 MHz in performance mode.
- Integrated PCIe 3.0 controller with 32 lanes per processor
- Intel Virtualization Technology (VT-x and VT-d)
- Intel Turbo Boost Technology 2.0
- Intel Advanced Vector Extensions 2 (AVX2)
- Intel AES-NI instructions for accelerating of encryption
- Advanced QPI and memory RAS features
- Machine Check Architecture recovery (non-running and running paths)
- Enhanced Machine Check Architecture Gen2 (eMCA2)
- Machine Check Architecture I/O
- Security technologies: OS Guard, Secure Key, Intel TXT

Intel Xeon processor E7-4800/8800 v2 product family

The Intel Xeon E7 v2 processor family provides two types of processors as well: E7-4800 v2 and E7-8800 v2 for four- and eight-socket configurations.

The Intel Xeon processor E7-4800 v2 and E7-8800 v2 product family offers the following key features:

- Up to 15 cores and 30 threads (by using Hyper-Threading feature) per processor
- Up to 37.5 MB of L3 cache
- Up to 3.4 GHz core frequencies
- Up to 8 GTps bandwidth of QPI links
- Integrated memory controller with four SMI2 channels that support up to 24 DDR3 DIMMs
- Up to 1600 MHz DDR3 memory speeds
- Integrated PCIe 3.0 controller with 32 lanes per processor
- Intel Virtualization Technology (VT-x and VT-d)
- Intel Turbo Boost Technology 2.0
- Intel Advanced Vector Extensions (AVX)
- Intel AES-NI instructions for accelerating of encryption
- Advanced QPI and memory RAS features
- Machine Check Architecture recovery (non-running and running paths)
- Enhanced Machine Check Architecture Gen1 (eMCA1)
- Machine Check Architecture I/O
- Security technologies: OS Guard, Secure Key, Intel TXT

Intel Xeon E7 features

Intel Xeon E7 processors include a broad set of features and extensions. Many of these technologies are common for all Intel Xeon E7 generations; some technologies are unique to the latest Intel Xeon E7 v4 family.

Intel Transactional Synchronization eXtensions

Intel Transactional Synchronization eXtensions (TSX) feature the latest v4 and v3 generations of Intel Xeon E7 processor families and brings hardware transactional memory support. Intel TSX implements a memory-locking approach that is called Hardware Lock Elision (HLE), which facilitates running multithreaded applications more efficiently.

Much TSX-aware software gained great performance boosts by running on Intel Xeon E7 v4 processors. For example, SAP HANA SPS 09 in-memory database showed twice as many transactions per minute with Intel TSX enabled versus TSX disabled on E7 v3 processors and three times more transactions per minute compared to Intel Xeon E7 v2 processors.

For more information about Intel TSX, see the Solution Brief, *Ask for More from Your Data*, which is available here:

http://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/sap-hana-real-time-analytics-solution-brief.pdf

Intel Advanced Encryption Standard

Advanced Encryption Standard (AES) is an encryption standard that is widely used to protect network traffic and sensitive data. Advanced Encryption Standard - New Instructions

(AES-NI), which is available with the E7 processors, implements certain complex and performance intensive steps of the AES algorithm by using processor hardware. AES-NI can accelerate the performance and improve the security of an implementation of AES versus an implementation that is performed by software.

For more information about Intel AES-NI, see this website:

http://software.intel.com/en-us/articles/intel-advanced-encryption-standard-instructions-aes-ni

Intel Virtualization Technology

Intel Virtualization Technology (Intel VT) is a suite of processor and I/O hardware enhancements that assists virtualization software to deliver more efficient virtualization solutions and greater capabilities.

Intel Virtualization Technology for x86 (Intel VT-x) allows the software hypervisors to better manage memory and processing resources for virtual machines (VMs) and their guest operating systems.

Intel Virtualization Technology for Directed I/O (Intel VT-d) helps improve I/O performance and security for VMs by enabling hardware-assisted direct assignment and isolation of I/O devices.

For more information about Intel Virtualization Technology, see this website:

http://www.intel.com/technology/virtualization

Hyper-Threading Technology

Intel Hyper-Threading Technology enables a single physical processor to run two separate code streams (threads) concurrently. To the operating system, a processor core with Hyper-Threading is seen as two logical processors. Each processor has its own architectural state; that is, its own data, segment, and control registers, and its own advanced programmable interrupt controller (APIC).

Each logical processor can be individually halted, interrupted, or directed to run a specified thread independently from the other logical processor on the chip. The logical processors share the running resources of the processor core, which include the running engine, caches, system interface, and firmware.

Hyper-Threading Technology improves server performance. This process is done by using the multithreading capability of operating systems and server applications in such a way as to increase the use of the on-chip running resources that are available on these processors. Application types that make the best use of Hyper-Threading include virtualization, databases, email, Java, and web servers.

For more information about Hyper-Threading Technology, see this website:

http://www.intel.com/technology/platform-technology/hyper-threading

vSphere 5.1 and 8-socket systems: VMware vSphere 5.1 has a fixed upper limit of 160 concurrent threads. Therefore, if you use an 8-socket system with more than 10 cores per processor, you should disable Hyper-Threading.

Turbo Boost Technology 2.0

The Intel Xeon E7-8800/4800 family of processors brings enhanced capabilities for changing processor speed with Intel Turbo Boost 2.0 technology.

Intel Turbo Boost Technology dynamically saves power on unused processor cores and increases the clock speed of the cores in use. Depending on current workload, Intel Turbo Boost Technology allows a dynamic increase in the clock speed of the active cores to gain a performance boost. For example, a 3.4 GHz 15-core processor can overclock the cores up to 3.7 GHz.

Turbo Boost Technology is available on a per-processor basis for the X6 systems. For ACPI-aware operating systems and hypervisors, such as Microsoft Windows 2008/2012, RHEL 5/6, SLES 11, VMware ESXi 4.1, and later, no changes are required to use it. Turbo Boost Technology can be used with any number of enabled and active cores, which results in increased performance of multithreaded and single-threaded workloads.

Turbo Boost Technology dynamically saves power on unused processor cores and increases the clock speed of the cores in use. In addition, it can temporarily increase the speed of all cores by intelligently managing power and thermal headroom. For example, a 2.5 GHz 15-core processor can temporarily run all 15 active cores at 2.9 GHz. With only two cores active, the same processor can run those active cores at 3.0 GHz. When the other cores are needed again, they are turned back on dynamically and the processor frequency is adjusted.

When temperature, power, or current exceeds factory-configured limits and the processor is running above the base operating frequency, the processor automatically steps the core frequency back down to reduce temperature, power, and current. The processor then monitors these variables, and reevaluates whether the current frequency is sustainable or if it must reduce the core frequency further. At any time, all active cores run at the same frequency.

For more information about Turbo Boost Technology, see this website:

http://www.intel.com/technology/turboboost/

QuickPath Interconnect

The Intel Xeon E7 processors implemented in X6 servers include two integrated memory controllers in each processor. Processor-to-processor communication is carried over shared-clock or coherent QPI links. Each processor has three QPI links to connect to other processors.

Figure 10 shows the QPI configurations. On the left is how the four sockets of the x3850 X6 are connected. On the right is how all eight sockets of the x3950 X6 are connected.

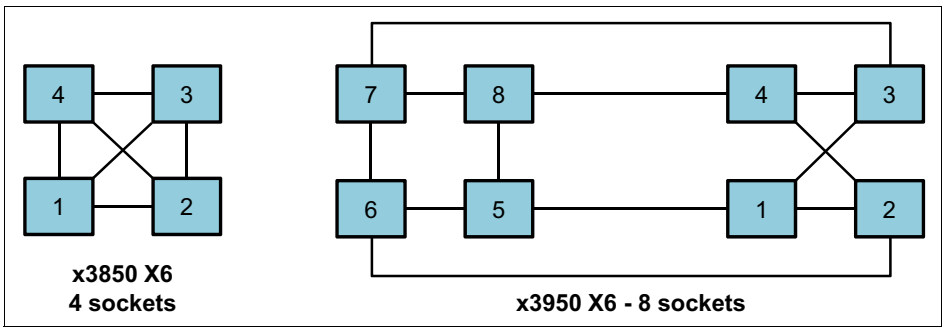

Figure 10 QPI links between processors

Each processor has some memory that is connected directly to that processor. To access memory that is connected to *another* processor, each processor uses QPI links through the other processor. This design creates a non-uniform memory access (NUMA) system. Similarly, I/O can be local to a processor or remote through another processor.

For QPI use, Intel modified the MESI cache coherence protocol to include a forwarding state. Therefore, when a processor asks to copy a shared cache line, only one other processor responds.

For more information about QPI, see this website:

http://www.intel.com/technology/quickpath

Intel Data Direct I/O

For I/O, Intel no longer has a separate I/O hub. Instead, it now integrates PCI Express 3.0 I/O into the processor. Data Direct I/O helps to optimize data transfer between local CPU and PCIe devices. The combination of Data Direct I/O and PCI 3.0 provides a higher I/O performance with lower latencies and reduced power consumption.

For more information about Data Direct I/O, see this website:

http://www.intel.com/content/www/us/en/io/direct-data-i-o.html

RAS features

The Intel Xeon processor E7 family of processors has the following RAS features on their interconnect links (SMI and QPI):

- Cyclic redundancy checking (CRC) on the QPI links

  The data on the QPI link is checked for errors.

- QPI packet retry

  If a data packet on the QPI link has errors or cannot be read, the receiving processor can request that the sending processor try sending the packet again.

- QPI clock failover

  If there is a clock failure on a coherent QPI link, the processor on the other end of the link can become the clock. This action is not required on the QPI links from processors to I/O hubs because these links are asynchronous.

- QPI self-healing

  If persistent errors are detected on a QPI link, the link width can be reduced dynamically to allow the system to run in a degraded mode until repair can be performed. QPI link can reduce its width to a half width or a quarter width, and slowdown its speed.

- Scalable memory interconnect (SMI) packet retry

  If a memory packet has errors or cannot be read, the processor can request that the packet be resent from the memory buffer.

Machine Check Architecture recovery

The Intel Xeon processor E7 family also includes Machine Check Architecture (MCA) recovery, a RAS feature that enables the handling of system errors that otherwise require that the operating system be halted. For example, if a dead or corrupted memory location is discovered but it cannot be recovered at the memory subsystem level and it is not in use by the system or an application, an error can be logged and the operation of the server can continue. If it is in use by a process, the application to which the process belongs can be stopped or informed about the situation.

Implementation of the MCA recovery requires hardware support, firmware support (such as found in the UEFI), and operating system support. Microsoft, SUSE, Red Hat, VMware, and other operating system vendors include or plan to include support for the Intel MCA recovery feature on the Intel Xeon processors in their latest operating system versions.

The new MCA recovery features of the Intel Xeon processor E7-4800/8800 product family include:

- Execution path recovery: The ability to work with hardware and software to recognize and isolate the errors that were delivered to the execution engine (core).
- Enhanced MCA (eMCA) Generation 1: Provides enhanced error log information to the operating system, hypervisor, or application that can be used to provide better diagnostic and predictive failure analysis for the system. This feature enables higher levels of uptime and reduced service costs.
- Enhanced MCA (eMCA) Generation 2: Provides more capabilities for error handling (E7 v3 and v4 processors only).

Security improvements

The Intel Xeon E7-4800/8800 processor families feature the following important security improvements that help to protect systems from different types of security threats:

- Intel OS Guard: Evolution of Intel Execute Disable Bit technology, which helps to protect against escalation of privilege attacks by preventing code execution from user space memory pages while in kernel mode. It helps to protect against certain types of malware attacks.
- #VE2 (Beacon Pass 2 Technology): #VE utilizes ISA-level CPU assists to allow memory-monitoring of antimalicious software performance to scale on virtualized and non-virtualized servers, making deep malicious software detection possible on server platforms.
- Intel Trusted Execution Technology (Intel TXT), Intel VT-x, and Intel VT-d: New hardware-based techniques, with which you can isolate VMs and start VMs in a trusted environment only. In addition, malware-infected VMs cannot affect nother VMs on the same host.
- Intel Secure Key: Provides hardware random number generation without storing any data in system memory. It keeps generated random numbers out of sight of malware, which enhances encryption protection.

For more information, see *Crimeware Protection: 3rd Generation Intel Core vPro Processors*, which is available at this website:

http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/3rd-gen-core-vpro-security-paper.pdf

Compute Books

The core modular element of the X6 design is a Compute Book. It contains:

- One Intel Xeon E7 v2, v3, or v4 processor
- 24 memory DIMM slots
- Two dual-motor fans.

Figure 11 on page 20 shows the DDR3 Compute Book.

Figure 11 The Compute Book

The system board of the Compute Book has two sides on which all components reside.

Figure 12 and Figure 13 show the left and right sides of the Compute Book, respectively. The left side contains one processor and 12 DIMM slots. The right side contains 12 DIMM slots, for a total of 24 DIMMs per Compute Book.

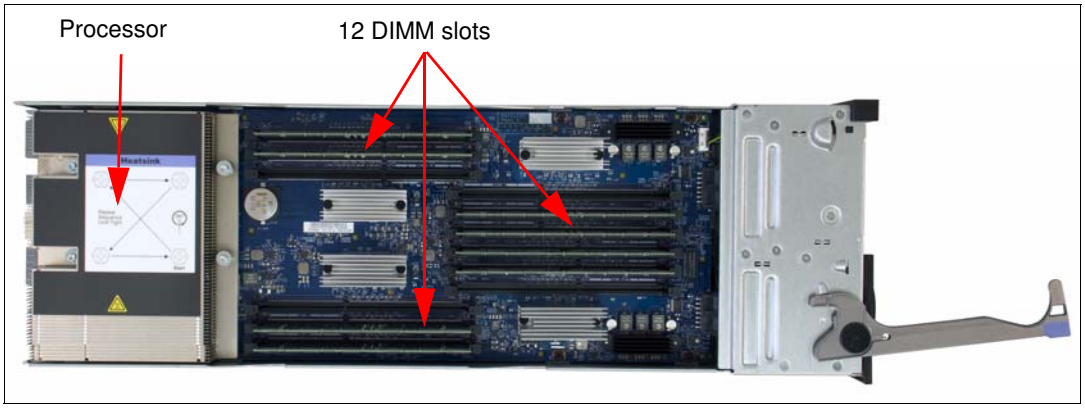

Figure 12 The Compute Book left side

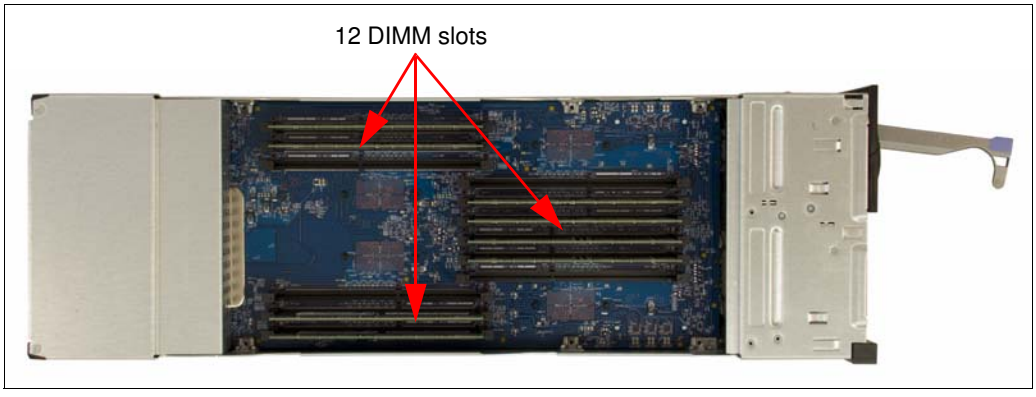

Figure 13 The Compute Book right side

The x3850 X6 server supports up to four Compute Books, and the x3950 X6 server supports up to eight Compute Books. The following configurations are supported:

- 2 or 4 Compute Books for the x3850 X6
- 4, 6, or 8 Compute Books for the x3950 X6

All Compute Books in a server must have the same processors.

Memory subsystem

The System x3850 X6 and x3950 X6 servers support three generations of Intel Xeon E7 processors:

- E7 v4 processors support DDR4 memory only
- E7 v3 processors can use either DDR3 or DDR4 memory
- E7 v2 processors support DDR3 memory only

DDR4 is a new memory standard that is supported by the Intel Xeon E7 v3 and v4 processor families. DDR4 memory modules can run at greater speeds than DDR3 DIMMs, operate at lower voltage, and are more energy-efficient than DDR3 modules.

X6 Compute Books with E7 v3 or v4 processors and DDR4 memory interface support Lenovo TruDDR4 memory modules, which are tested and tuned to maximize performance and reliability. Lenovo TruDDR4 DIMMs can operate at greater speeds and have higher performance than DIMMs that only meet industry standards.

DDR3 and TruDDR4 memory types have ECC protection and support Chipkill and Redundant Bit Steering technologies.

Each processor has two integrated memory controllers, and each memory controller has two Scalable Memory Interconnect generation 2 (SMI2) links that are connected to two scalable memory buffers. Each memory buffer has two memory channels, and each channel supports three DIMMs, for a total of 24 DIMMs per processor.

Figure 14 shows the processor's memory architecture.

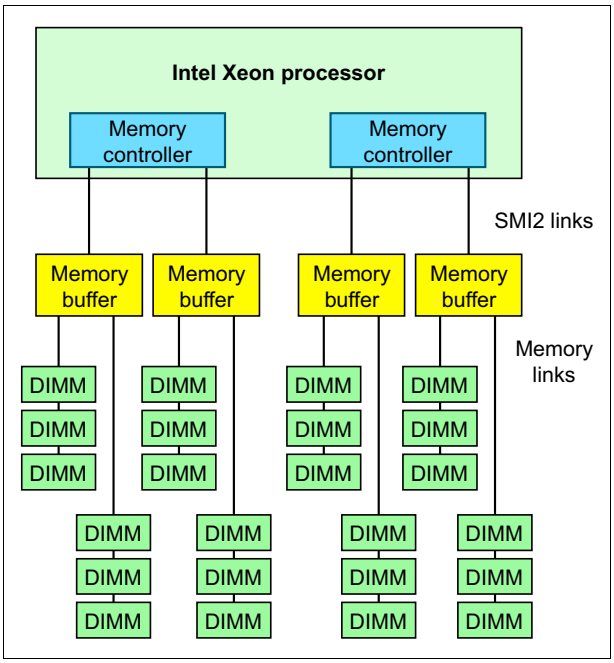

Figure 14 Intel Xeon processor E7-4800/8800 memory architecture

E7 **v2**-based Compute Books each support up to 24 DDR3 memory modules running at speeds up to 1600 MHz. Compute Books with E7 **v3** processors can be DDR3-based with up to 24 DDR3 DIMMs operating at speeds up to 1600 MHz, or DDR4-based with up to 24 TruDDR4 memory modules operating at speeds up to 1866MHz. *DDR3 and TruDDR4 DIMMs cannot be mixed;* all memory across all books within a server must be of the same type (either all DDR3 or all TruDDR4). Compute Books with E7 **v4** processors can support up to 24 TruDDR4 DIMMs operating at speeds up to 1866 MHz. DDR3 memory is *not* supported by E7 **v4** processors.

The x3850 X6 supports up to 96 DIMMs when all processors are installed (24 DIMMs per processor), and the x3950 X6 supports up to 192 DIMMs. The processor and the corresponding memory DIMM slots are on the Compute Book

Operational modes

The following memory modes are supported by the Intel Xeon processor E7 product families:

- Performance mode

  In this operation mode, each memory channel works independently and it is addressed individually via burst lengths of 8 bytes (64 bits). The Intel SMI2 channel operates at twice the memory speed. All channels can be populated in any order and modules have no matching requirements.

  Chipkill (Single Device Data Correction, or SDDC) is supported in Performance mode. Redundant Bit Steering (RBS) is *not* supported in Performance mode.

  Although in this mode DIMMs can be populated in any order, memory modules should be placed based by round-robin algorithm between SMI2 channels and alternating between DDR channels for best performance. For more information about DIMMs population order, see 3.9.2, "Memory population order" on page 90.

- RAS (Lockstep) mode

  In RAS operation mode (also known as Lockstep mode), the memory controller operates two memory channels behind one memory buffer as single channel.

  In RAS mode, the SMI2 channel operates at the memory transfer rate. DIMMs must be installed in pairs.

  Because data is moved by using both channels at once, more advanced memory protection schemes can be implemented to provide protection against single-bit and multibit errors:

  – Chipkill, also known as SDDC
  – RBS (multibit correction)

  The combination of these two RAS features is also known as Double Device Data Correction (DDDC).

Mirroring and sparing are also supported in both modes, as described in , “Memory mirroring and rank sparing” on page 26.

Figure 15 shows the two modes. In RAS mode, both channels of one memory buffer are in lockstep with each other.

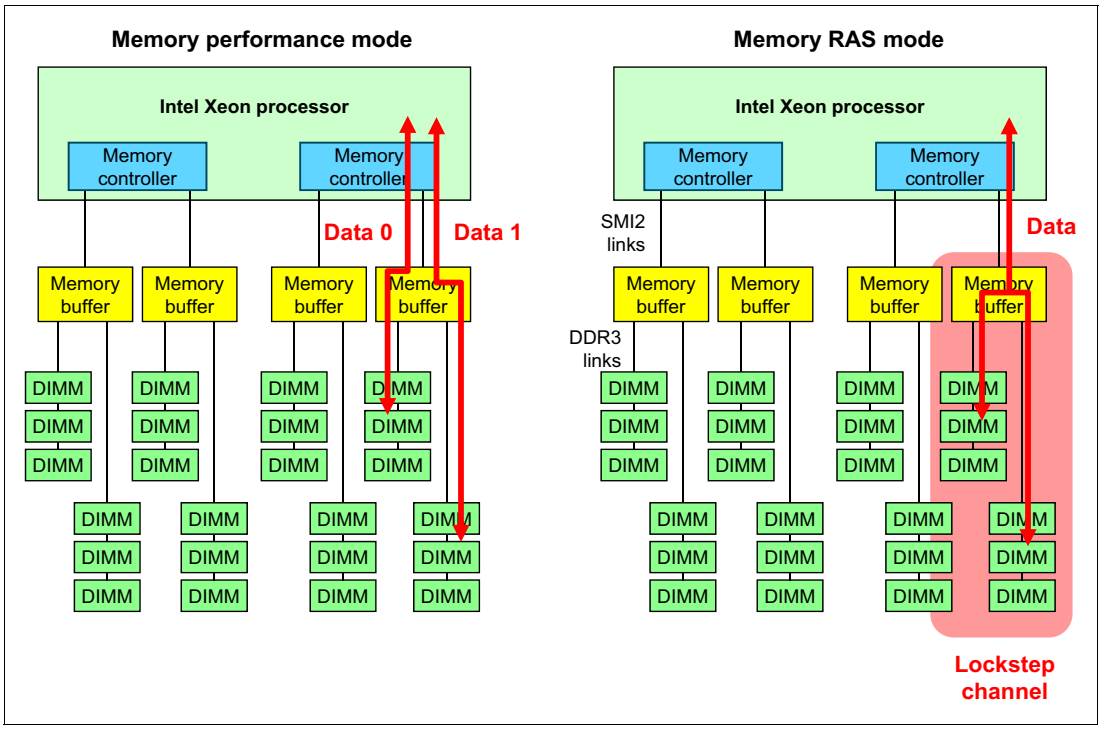

Figure 15 Memory modes: Performance mode (left) and RAS mode (right)

The following tables show the maximum speed and bandwidth for the SMI2 and memory channels in both modes for DDR3 and TruDDR4 memory modules, as well as their operating voltages.

Table 2 DDR3 memory speeds in Performance mode

DIMM type / capacity	**1 DPC 1.35 V**	**1 DPC 1.5 V**	**2 DPC 1.35 V**	**2 DPC 1.5 V**	**3 DPC 1.35 V**	**3 DPC 1.5 V**
DDR3, Performance mode (2:1): Maximum performance						
1RX4, 2Gb, 1600 MHz / 4Gb	1333 MHz		1333 MHz			1333 MHz
1RX4, 4Gb, 1600 MHz / 8Gb	1333 MHz		1333 MHz			1333 MHz
2RX4, 4Gb, 1600 MHz / 16Gb	1333 MHz		1333 MHz			1333 MHz
4RX4, 4Gb, 1600 MHz / 32Gb	1333 MHz		1333 MHz		1333 MHz	
8Rx4, 4Gb, 1333 MHz / 64Gb	1333 MHz		1333 MHz		1333 MHz	
DDR3, Performance mode (2:1): Balanced						
1RX4, 2Gb, 1600 MHz / 4Gb	1067 MHz		1067 MHz		1067 MHz	
1RX4, 4Gb, 1600 MHz / 8Gb	1067 MHz		1067 MHz		1067 MHz	
2RX4, 4Gb, 1600 MHz / 16Gb	1067 MHz		1067 MHz		1067 MHz	
4RX4, 4Gb, 1600 MHz / 32Gb	1067 MHz		1067 MHz		1067 MHz	
8Rx4, 4Gb, 1333 MHz / 64Gb	1067 MHz		1067 MHz		1067 MHz	
DDR3, Performance mode (2:1): Minimal						
1RX4, 2Gb, 1600 MHz / 4Gb	1067 MHz		1067 MHz		1067 MHz	
1RX4, 4Gb, 1600 MHz / 8Gb	1067 MHz		1067 MHz		1067 MHz	
2RX4, 4Gb, 1600 MHz / 16Gb	1067 MHz		1067 MHz		1067 MHz	
4RX4, 4Gb, 1600 MHz / 32Gb	1067 MHz		1067 MHz		1067 MHz	
8Rx4, 4Gb, 1333 MHz / 64Gb	1067 MHz		1067 MHz		1067 MHz	

Table 3 DDR3 memory speeds in RAS mode

DIMM type / capacity	**1 DPC 1.35 V**	**1 DPC 1.5 V**	**2 DPC 1.35 V**	**2 DPC 1.5 V**	**3 DPC 1.35 V**	**3 DPC 1.5 V**
DDR3, RAS mode (1:1): Maximum performance						
1RX4, 2Gb, 1600 MHz / 4Gb		1600 MHz		1600MHz		1333MHz
1RX4, 4Gb, 1600 MHz / 8Gb		1600 MHz		1600 MHz		1333 MHz
2RX4, 4Gb, 1600 MHz / 16Gb		1600 MHz		1600 MHz		1333 MHz
4RX4, 4Gb, 1600 MHz / 32Gb		1600 MHz		1600 MHz	1333 MHz	
8Rx4, 4Gb, 1333 MHz / 64Gb	1333 MHz		1333 MHz		1333 MHz	
DDR3, RAS mode (1:1): Balanced						
1RX4, 2Gb, 1600 MHz / 4Gb	1333 MHz		1333 MHz		1067 MHz	
1RX4, 4Gb, 1600 MHz / 8Gb	1333 MHz		1333 MHz		1067 MHz	
2RX4, 4Gb, 1600 MHz / 16Gb	1333 MHz		1333 MHz		1067 MHz	
4RX4, 4Gb, 1600 MHz / 32Gb	1333 MHz		1333 MHz		1333 MHz	

DIMM type / capacity	**1 DPC 1.35 V**	**1 DPC 1.5 V**	**2 DPC 1.35 V**	**2 DPC 1.5 V**	**3 DPC 1.35 V**	**3 DPC 1.5 V**
8Rx4, 4Gb, 1333 MHz / 64Gb	1067 MHz		1067 MHz		1067 MHz	
DDR3, RAS mode (1:1): Minimal						
1RX4, 2Gb, 1600 MHz / 4Gb	1067 MHz		1067 MHz		1067 MHz	
1RX4, 4Gb, 1600 MHz / 8Gb	1067 MHz		1067 MHz		1067 MHz	
2RX4, 4Gb, 1600 MHz / 16Gb	1067 MHz		1067 MHz		1067 MHz	
4RX4, 4Gb, 1600 MHz / 32Gb	1067 MHz		1067 MHz		1067 MHz	
8Rx4, 4Gb, 1333 MHz / 64Gb	1067 MHz		1067 MHz		1067 MHz	

Table 4 DDR4 memory speeds in Performance mode

DIMM type / capacity	**1 DPC 1.2 V**	**2 DPC 1.2 V**	**3 DPC 1.2 V**
TruDDR4, Performance mode (2:1): Maximum performance			
1RX4, 4Gb, 2133 MHz / 8Gb	1600 MHz	1600 MHz	1600 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1600 MHz	1600 MHz	1600 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1600 MHz	1600 MHz	1600 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1600 MHz	1600 MHz	1600 MHz
TruDDR4, Performance mode (2:1): Balanced			
1RX4, 4Gb, 2133 MHz / 8Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1333 MHz	1333 MHz	1333 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1333 MHz	1333 MHz	1333 MHz
TruDDR4, Performance mode (2:1): Minimal			
1RX4, 4Gb, 2133 MHz / 8Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1333 MHz	1333 MHz	1333 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1333 MHz	1333 MHz	1333 MHz

Table 5 DDR4 memory speeds in RAS mode

DIMM type / capacity	**1 DPC 1.2 V**	**2 DPC 1.2 V**	**3 DPC 1.2 V**
TruDDR4, RAS mode (1:1): Maximum performance			
1RX4, 4Gb, 2133 MHz / 8Gb	1867 MHz	1867 MHz	1600 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1867 MHz	1867 MHz	1600 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1867 MHz	1867 MHz	1600 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1867 MHz	1867 MHz	1600 MHz

DIMM type / capacity	**1 DPC 1.2 V**	**2 DPC 1.2 V**	**3 DPC 1.2 V**
TruDDR4, RAS mode (1:1): Balanced			
1RX4, 4Gb, 2133 MHz / 8Gb	1600 MHz	1600 MHz	1333 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1600 MHz	1600 MHz	1333 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1600 MHz	1600 MHz	1333 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1600 MHz	1600 MHz	1333 MHz
TruDDR4, RAS mode (1:1): Minimal			
1RX4, 4Gb, 2133 MHz / 8Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 4Gb, 2133 MHz / 16Gb	1333 MHz	1333 MHz	1333 MHz
2RX4, 8Gb, 2133 MHz / 32Gb	1333 MHz	1333 MHz	1333 MHz
4RX4, 16Gb, 2133 MHz / 64Gb	1333 MHz	1333 MHz	1333 MHz

Memory mirroring and rank sparing

In addition to Performance and RAS modes, the memory subsystem has the following RAS features that can be enabled from the UEFI:

- Memory mirroring
- Rank sparing

Memory mirroring

To improve memory reliability and availability, the memory controller can mirror memory data across two memory channels. To enable the mirroring feature, you must have both memory channels of a processor populated with the same DIMM type and amount of memory.

Memory mirroring provides the user with a redundant copy of all code and data addressable in the configured memory map. Two copies of the data are kept, similar to the way RAID-1 writes to disk. Reads are interleaved between memory channels. The system automatically uses the most reliable memory channel as determined by error logging and monitoring.

If errors occur, only the alternative memory channel is used until bad memory is replaced. Because a redundant copy is kept, mirroring results in only half the installed memory being available to the operating system. Memory mirroring does not support asymmetrical memory configurations and requires that each channel be populated in identical fashion. For example, you must install two identical 4 GB 2133MHz DIMMs equally and symmetrically across the two memory channels to achieve 4 GB of mirrored memory.

Memory mirroring is a hardware feature that operates independent of the operating system. There is a slight memory performance trade-off when memory mirroring is enabled.

The memory mirroring feature can be used with performance or RAS modes:

- When Performance mode is used, memory mirroring duplicates data between memory *channels* of the two memory buffers connected to one memory controller.
- In RAS (Lockstep) mode, memory mirroring duplicates data between memory *buffers* that are connected to the same memory controller.

Figure 16 shows how memory mirroring is implemented in Performance mode (left) and RAS mode (right).

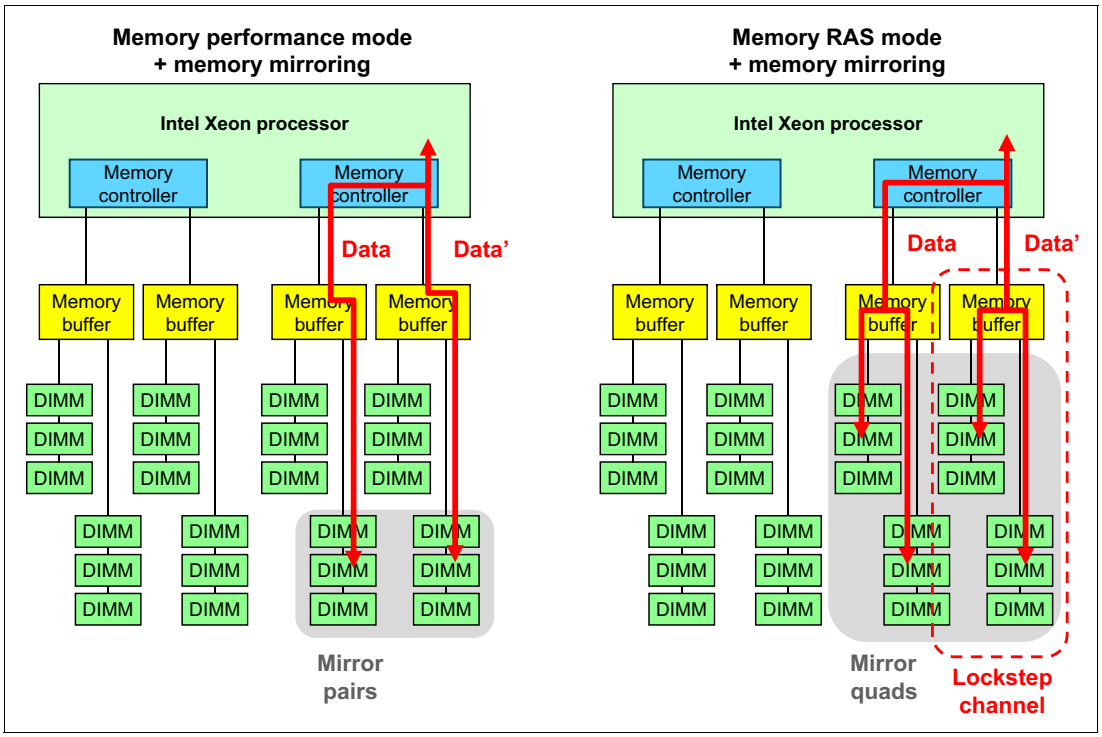

Figure 16 Memory mirroring with used with Performance mode (left) and RAS mode (right)

The following memory mirroring rules apply:

- The server supports single-socket memory mirroring. The Compute Book memory channel 0 mirrors memory channel 1, and memory channel 2 mirrors memory channel 3. This mirroring provides redundancy in memory but reduces the total memory capacity by half.
- DIMMs must be installed in pairs for each Compute Book when the memory mirroring feature is used.
- The DIMM population must be identical (capacity, organization, and so on) for memory channel 0 and memory channel 1, and identical for memory channel 2 and memory channel 3.
- Memory mirroring reduces the maximum available memory by half of the installed memory. For example, if the server has 64 GB of installed memory, only 32 GB of addressable memory is available when memory mirroring is enabled.

Memory Address Range Mirroring

Intel Xeon E7 **v3** and **v4** processors provide an advanced version of the memory mirroring feature named Memory Address Range Mirroring. Unlike basic memory mirroring, when whole memory channels or memory buffers are mirrored, memory address range mirroring designates a memory *region* to mirror. You can choose a range of memory addresses to be mirrored, which allows you to keep critical data secured and save large amounts of memory at the same time. Each processor supports up to two mirror ranges, each mirror range size uses 64 MB granularity. For additional information on Memory Address Range Mirroring, refer to the following paper:
http://software.intel.com/en-us/articles/reliability-availability-and-serviceability-integration-and-validation-guide-for-the-intel

Rank sparing

In rank-sparing mode, one rank is held in reserve as a spare for the other ranks in the same memory channel. There are eight memory channels per processor.

Memory rank sparing provides a degree of redundancy in the memory subsystem, but not to the extent of mirroring. In contrast to mirroring, sparing leaves more memory for the operating system. In sparing mode, the trigger for failover is a preset threshold of correctable errors. When this threshold is reached, the content is copied to its spare rank. The failed rank is then taken offline, and the spare counterpart is activated for use.

In rank sparing mode, one rank per memory channel is configured as a spare. The spare rank must have identical or larger memory capacity than all the other ranks (sparing source ranks) on the same channel.

For example, if dual-rank DIMMs are installed and are all of the same capacity, there are six ranks total for each memory channel (three DIMMs per channel). This configuration means that one of the six ranks are reserved and five of the six ranks can be used for the operating system.

Memory sparing is a hardware featurethat operates independent of the operating system. There is a slight memory performance trade-off when memory sparing is enabled.

The rank sparing feature can be used in addition to performance or RAS modes. Consider the following points:

- When Performance mode is used, rank sparing duplicates data between memory modules of the *same channel* of one memory buffer. If there is an imminent failure (as indicated by a red X in Figure 17 on page 29), that rank is taken offline and the data is copied to the spare rank.
- When RAS (Lockstep) mode is used, rank sparing duplicates data between memory channels of one memory buffer. If there is an imminent failure (as indicated by a red X in Figure 17 on page 29), that rank is taken offline and the data is copied to the spare rank. In addition, the partner rank on the other channel that is connected to the same memory buffer also is copied over.

Figure 17 shows the rank sparing usage with Performance mode (left) and RAS mode (right).

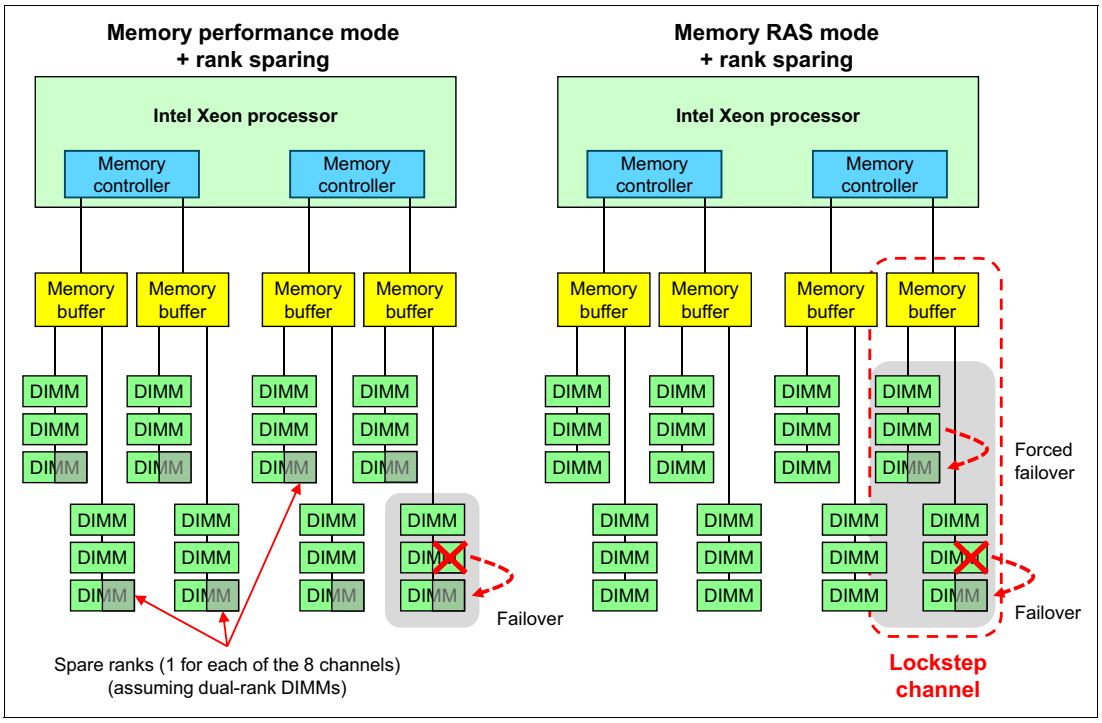

Figure 17 Rank sparing: Performance mode (left) and RAS mode (right)

The following configuration rules apply for rank sparing:

- Memory rank sparing is not supported if memory mirroring is enabled.
- The spare rank must have identical or larger memory capacity than all the other ranks on the same memory channel.
- When single-rank DIMMs are used, a minimum of *two* DIMMs must be installed per memory channel to support memory sparing.
- When multirank DIMMs are used, *one* multirank DIMM can be installed per memory channel to support memory sparing.
- The total memory available in the system is reduced by the amount of memory that is allocated for the spare ranks.

Multiple rank sparing

Intel Xeon E7 v3 and v4 processors provide an advanced version of rank sparing called multiple rank sparing. It is now possible to specify more than one rank as a hot spare. When the memory errors threshold is reached, the content of the failed rank is copied to its spare rank. The failed rank is then taken offline and the spare counterpart is activated for use. If another memory rank fails, the same procedure triggers again to take over to the next standby rank, and so on.

Note that the spare rank(s) must have memory capacity identical to, or larger than, all the other ranks (sparing source ranks). The total memory available in the system is reduced by the amount of memory allocated for the spare ranks.

Chipkill

Chipkill memory technology, which is an advanced form of error correction code (ECC), is available on X6 servers. Chipkill protects the memory in the system from any single memory chip failure. It also protects against multibit errors from any portion of a single memory chip.

Chipkill on its own can provide 99.94% memory availability to the applications without sacrificing performance and with standard ECC DIMMs.

Chipkill is used in Performance and RAS modes.

Redundant bit steering

Redundant bit steering provides the equivalent of a hot-spare drive in a RAID array. It is based in the memory controller and senses when a chip on a DIMM fails and when to route the data around the failed chip.

Redundant bit steering is used in *RAS mode only*.

X6 servers support the Intel implementation of Chipkill plus redundant bit steering, which Intel refers to as DDDC.

Redundant bit steering uses the ECC coding scheme that provides Chipkill coverage for x4 DRAMs. This coding scheme leaves the equivalent of one x4 DRAM spare in every pair of DIMMs. If a chip failure on the DIMM is detected, the memory controller can copy data from the failed chip through the spare x4.

Redundant bit steering operates automatically without issuing a Predictive Failure Analysis (PFA) or light path diagnostics alert to the administrator, although an event is logged to the service processor log. After the second DRAM chip failure on the DIMM in RAS (Lockstep) mode, more single bit errors result in PFA and light path diagnostics alerts.

Advanced Page Retire

Advanced Page Retire is an algorithm to handle memory errors. It is built-in sophisticated error-handling firmware that uses and coordinates memory recovery features, which balances the goals of maximum up time and minimum repair actions.

The algorithm uses short- and long-term thresholds per memory rank with leaky bucket and automatic sorting of memory pages with the highest correctable error counts. First, it uses hardware recovery features, followed by software recovery features, to optimize recovery results for newer and older operating systems and hypervisors.

When recovery features are exhausted, the firmware issues a Predictive Failure Alert. Memory that failed completely is held offline during starts until it is repaired. Failed DIMMs are indicated by light path diagnostics LEDs that are physically at the socket location.

Storage subsystem

The x3850 X6 and x3950 X6 servers support the following types of internal storage:

- Hard disk drive (HDD) storage
  - 2.5-inch SAS/SATA hot-swap HDDs

- Flash memory (or solid-state) storage
  - 1.8-inch SATA SSDs in the eXFlash SSD unit
  - 2.5-inch SAS/SATA SSDs
  - PCIe Flash Storage adapters

Storage Book

The 1.8-inch SSDs in the eXFlash SSD units and 2.5-inch SSDs and HDDs are installed in Storage Book that is shown in Figure 18.

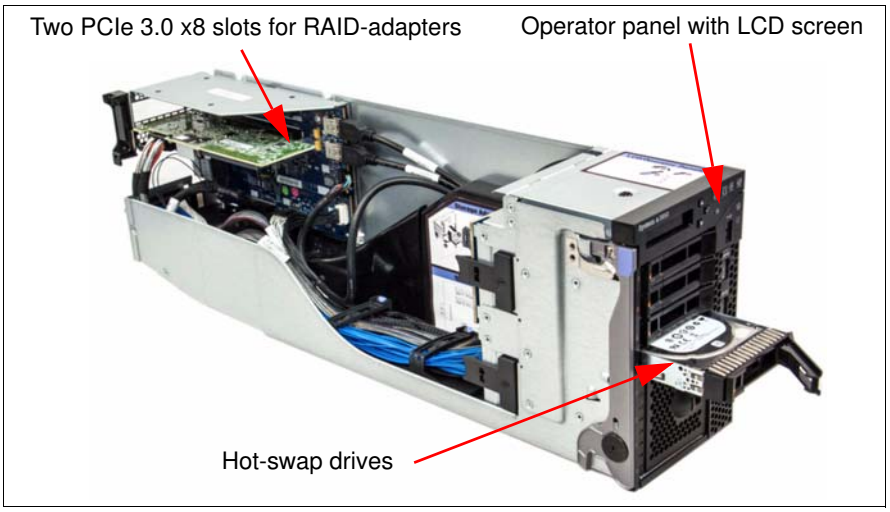

Figure 18 The X6 Storage Book

In addition to the drive bays, Storage Book contains two PCIe 3.0 x8 slots for internal RAID controllers or HBAs.

The X6 servers support 12 Gb SAS/SATA connectivity for internal storage. 12 Gb SAS doubles the data transfer rate compared to 6 Gb SAS solutions to fully unlock the potential of the PCIe 3.0 interface and to maximize performance for storage I/O-intensive applications, including databases, business analytics, and virtualization and cloud environment.

Lenovo 12 Gb SAS/SATA controllers provide performance benefits even for the 6 Gbps SAS drive infrastructure, especially for SSDs, by providing intelligent buffering capabilities. Lenovo 12 Gb SAS solution consists of ServeRAID™ M5210 RAID controllers, performance optimized N2215 SAS/SATA HBAs, and specialized 12 Gb SAS drive backplanes.

The ServeRAID M5210 SAS/SATA Controller has the following specifications:

- Eight internal 12 Gbps SAS/SATA ports
- Two x4 HD mini-SAS internal connectors (SFF-8643)
- Supports connections to SAS/SATA drives and SAS Expanders
- Supports RAID levels 0, 1, and 10
- Supports RAID levels 5 and 50 with optional M5200 Series RAID 5 upgrades
- Supports RAID 6 and 60 with the optional M5200 Series RAID 6 Upgrade
- Supports 1 GB non-backed cache or 1 GB, 2 GB, or 4 GB flash-backed cache
- Up to 12 Gbps throughput per port
- PCIe 3.0 x8 host interface
- Based on the LSI SAS3108 12 Gbps ROC controller

The N2215 SAS/SATA HBA has the following specifications:

- Eight internal 12 Gbps SAS/SATA ports
- Two x4 HD mini-SAS internal connectors (SFF-8643)
- Supports connections to SAS/SATA HDDs and SATA SSDs
- Optimized for SSD performance
- No RAID support
- Up to 12 Gbps throughput per port
- PCIe 3.0 x8 host interface
- Based on the LSI SAS3008 12 Gbps controller

Each Storage Book supports the following configurations:

- 4x 2.5-inch hot-swap drive bays; one RAID controller or HBA
- 8x 2.5-inch hot-swap drive bays; one or two RAID controllers or HBAs
- 4x 2.5-inch hot-swap drive bays + 8x 1.8-inch hot-swap SSD bays; two RAID controllers or HBAs
- 8x 1.8-inch hot-swap SSD bays; one RAID controller or HBA
- 16x 1.8-inch hot-swap SSD bays; two RAID controllers or HBAs

Flash internal storage

Currently, the processor and memory subsystems are well balanced and virtually not considered as performance bottlenecks in the majority of systems. The major source of performance issues is related to the storage I/O activity because of the speed of traditional HDD-based storage systems that still does not match the processing capabilities of the servers. This can lead to a situation when a powerful processor sits idle waiting for the storage I/O requests to complete, therefore wasting its time which negatively impacts user productivity, extends return on investments (ROI) time frame, and increase overall total cost of ownership (TCO).

Flash storage offerings can help to address the issues described previously by combining extreme IOPS performance with low response time. Lenovo X6 servers use flash storage offerings to achieve fastest response time for analytical workloads, transactional databases, and virtualized environments.

The X6 servers include the following Flash storage technologies:

- eXFlash SSD unit: Innovative high-density design of the drive cages and the performance-optimized storage controllers with the reliable high-speed SSD technology.
- High IOPS SSD Adapters: Utilize the latest enterprise-level solid-state storage technologies in a standard PCIe form factor and include sophisticated advanced features to optimize flash storage and help deliver consistently high levels of performance, endurance, and reliability.
- 2.5-inch Enterprise SSDs: Designed to be flexible across a wide variety of enterprise workloads in delivering outstanding performance, reliability, and endurance at an affordable cost.

Table 6 compares features of the Flash storage devices used in X6 servers.

Table 6 Flash storage devices

Feature	2.5-inch SSDs	PCIe SSD adapters	eXFlash SSD unit
Form factor	2.5-inch drive	PCIe adapter	8x 1.8-inch SSDs
Interface	6 Gbps SAS or SATA	PCIe 2.0 x8	6 Gbps SATA

Feature	2.5-inch SSDs	PCIe SSD adapters	eXFlash SSD unit
Capacity	Up to 1.6 TB	Up to 6.4 TB	Up to 6.4 TB
Max. random read IOPS	Up to 100,000	Up to 285,000	Up to 75,000 per SSD
Write latency	Less than 100µs	15 µs	65 µs
Hot-swap capabilities	Yes	No	Yes
RAID support	Yes	Chip-level redundancy; OS SW RAID	Yes

As a general consideration, 2.5-inch SSDs can be an entry-level solution that is optimized for commodity servers with conventional HDD drive tray, with moderate storage IOPS performance requirements.

PCIe SSD adapters are optimized for the storage I/O intensive enterprise workloads, significantly lower write latency.

eXFlash SSD unit

eXFlash technology is a server-based high performance internal storage solution that is based on SSD 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. Figure 19 shows an eXFlash unit, with the status lights assembly on the left side.

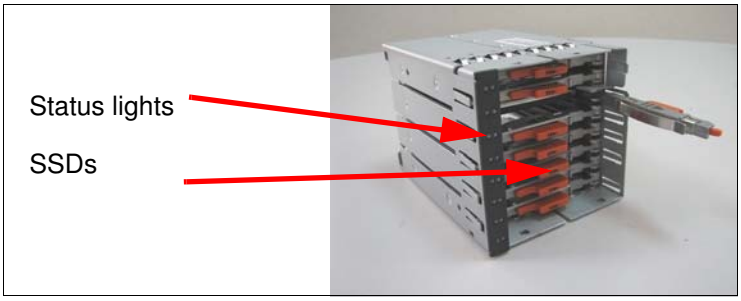

Figure 19 eXFlash unit

Each eXFlash unit can accommodate up to eight 1.8-inch hot-swap front-accessible SSDs. It occupies four 2.5-inch SAS hard disk drive bays. You can install the following number of eXFlash units:

- The x3850 X6 server can have up to 16 1.8-inch SSDs with up to two eXFlash units (up to eight SSDs per eXFlash unit).
- The x3950 X6 server can have up to 32 1.8-inch SSDs with up to four eXFlash units (up to eight SSDs per eXFlash unit).

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.

eXFlash requires the following components:

- eXFlash hot-swap SAS SSD backplane
- 1.8-inch SSDs
- Disk controllers

In environments where RAID protection is required (that is, where eXFlash is used as a master data storage), use the M5210 RAID controller with Performance Accelerator key enabled to ensure the peak IOPS can be reached.

The main advantage of M5210 with Performance Key controllers for SSDs is enabling the Cut Through I/O (CTIO) feature. This feature optimizes highly random read/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 you to achieve up to two times more IOPS compared to the controller with CTIO feature disabled.

Using a single eXFlash unit: A single eXFlash unit requires a dedicated controller (or two controllers).

In a non-RAID environment, use the N2215 HBA to ensure maximum random I/O performance is achieved. Only one N2215 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 two for the x3850 X6 server and four for the x3950 X6 server.

In summary, the eXFlash technology provides the following benefits:

- Lower implementation cost of high performance I/O-intensive storage systems with best cost per IOPS ratio
- Higher performance and lower latency of storage-intensive applications with up to 10 times less response time
- Savings in power and cooling with high performance per watt ratio
- Savings in floor space with extreme performance per U ratio
- Simplified management and maintenance with internal server-based configurations (no external power and information infrastructure needed)

PCIe Flash Storage adapters

The PCIe Flash Storage adapters provide a next level of high-performance storage based on SSD technology. Designed for high-performance servers and computing appliances, these adapters deliver random read throughput of up to 285,000 IOPS, while providing the added benefits of lower power, cooling, and management overhead and a smaller storage footprint.

The PCIe Flash Storage adapters combine high IOPS performance with low latency. As an example, with 4-KB block random reads, the 6400GB Enterprise Value io3 Flash Adapter can deliver up to 285,000 IOPS, compared with 420 IOPS for a 15K RPM 146 GB hard disk drive. The read access latency is about 92 microseconds, which is one hundredth of the latency of a 15K RPM 146 GB disk drive. The write access latency is even less, with about 15 microseconds.

Reliability features include the use of advanced wear-leveling, ECC protection, and Adaptive Flashback redundancy for RAID-like chip protection with self-healing capabilities, providing unparalleled reliability and efficiency. Advanced bad-block management algorithms enable taking blocks out of service when their failure rate becomes unacceptable.

Figure 20 shows the 6400GB Enterprise Value io3 Flash Adapter.

Figure 20 6400GB Enterprise Value io3 Flash Adapter

For more information about High IOPS PCIe SSD Adapters, see *Enterprise Value io3 PCIe Flash Adapters*, TIPS1237, which is available at:

http://lenovopress.com/tips1237

Networking and I/O

The X6 family of servers supports the latest generation of PCI Express (PCIe) protocol, Version 3.0. PCIe 3.0 is evolution of PCI Express I/O standard that brings doubled bandwidth over PCIe 2.0, which preserves compatibility with previous generations of PCIe protocol. Thus, PCIe 1.x and 2.x cards work properly in PCIe 3.0-capable slots, and PCIe 3.0 cards work when plugged into PCIe slots of previous generations.

PCIe 3.0 uses 128b/130b encoding scheme which is more efficient than 8b/10b encoding used in PCIe 2.0 protocol. This approach reduces overhead to less that 2% comparing to 20% of PCIe 2.0, and allows to double bandwidth at 8 GT/s speed.

The x3850 X6 server supports one mezzanine LOM slot (for an ML2 card) and up to 11 PCIe slots with the up to three I/O Books and one Storage Book installed in the 4-socket chassis. The x3950 X6 server supports two mezzanine LOM slots and up to 22 PCIe with the up to six I/O Books and two Storage Books installed in the 8-socket chassis.

Primary I/O Book supplies three PCIe 3.0 slots and one mezzanine LOM slot. PCIe slots are as follows:

- 2x PCIe 3.0 x16 (x16-wired); full height, half length
- 1x PCIe 3.0 x16 (x8-wired); full height, half length

Mezzanine LOM slot (PCIe 3.0 x8 interface) is used for mezzanine network cards that provide the flexibility in choosing integrated NIC option, including:

- 2x 10 Gb Ethernet SFP+ ports
- 2x 10 Gb Ethernet RJ-45 ports
- 4x 1 Gb Ethernet RJ-45 ports

In addition, one port on the ML2 card can be configured as a shared management/data port.

Figure 21 shows the Primary I/O Book location and its components.

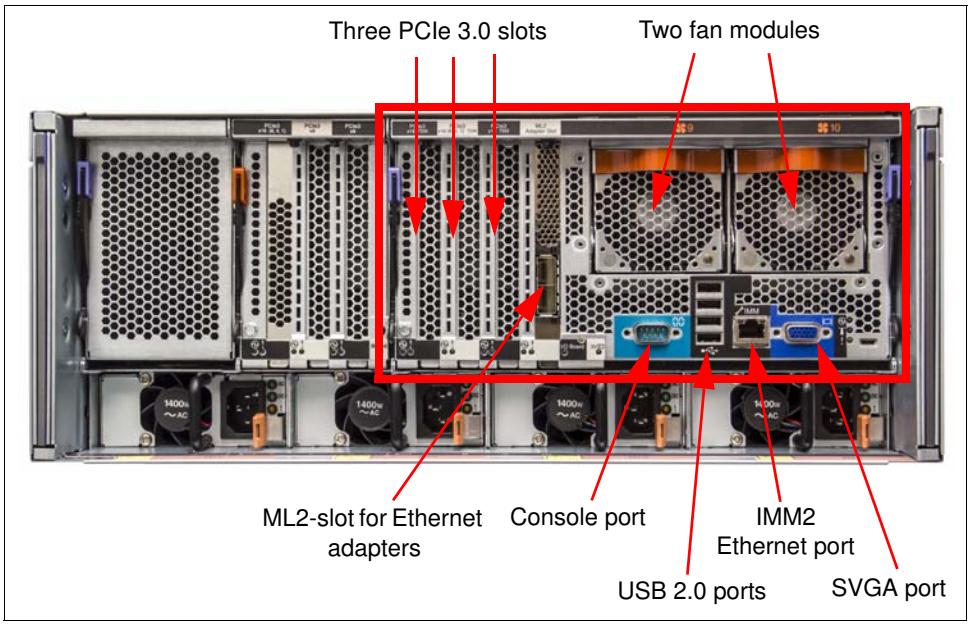

Figure 21 Primary I/O Book location and ports

In addition to PCIe expansion, Primary I/O Book provides I/O ports on the rear of the server. Primary I/O Book requires processors 1 and 2 in the x3850 X6 servers, and processors 1, 2, 5, and 6 in the x3950 X6 servers (two Primary I/O Books).

There are two types of optional I/O Books:

- Full-length
- Half-length

Full-length I/O Book is designed to support high performance GPU adapters (up to 300 W) and other full-length cards, and Half-length I/O Book supports traditional PCIe adapters. Up to two optional I/O Books can be installed in the x3850 X6 server, and they require processors 3 and 4. Up to four optional I/O Books can be installed in the x3950 X6 server, and they require processors 3, 4, 7, and 8. Optional I/O Books are enabled for PCIe hot add and hot replace functionality.

Figure 22 shows the two optional I/O Books.

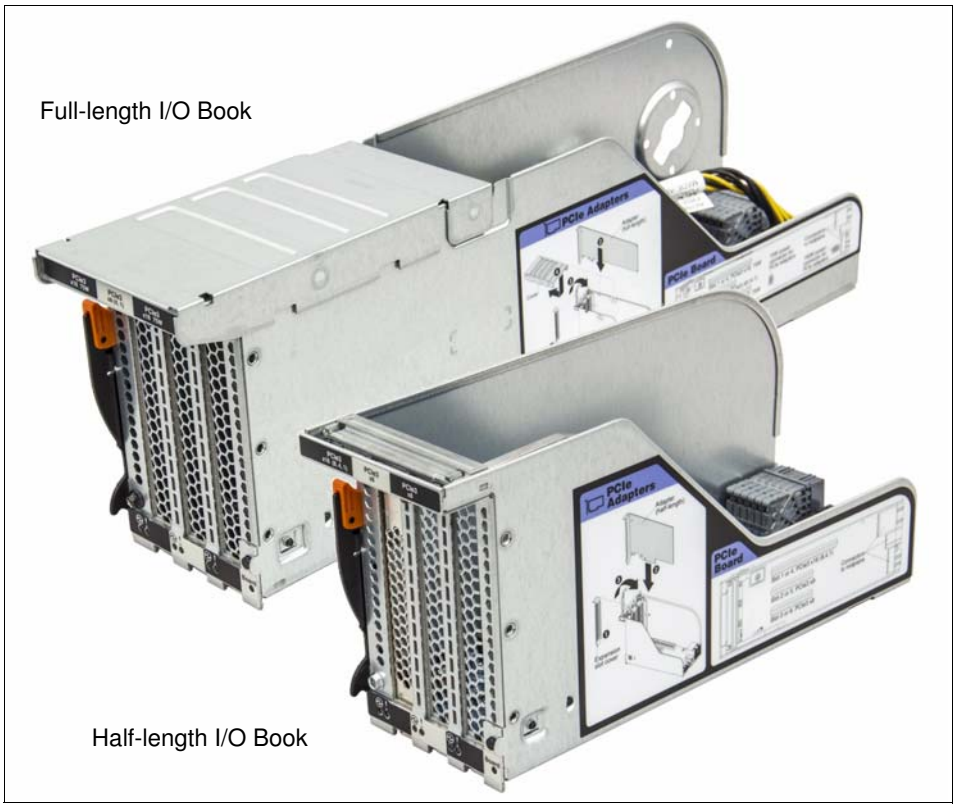

Figure 22 Optional I/O Books

I/O Book slots are as follows:

- Half-length I/O Book
  - Two PCIe 3.0 x8 slots (x8 wired)
  - One PCIe 3.0 x16 slot (x16 wired)
- Full-length I/O Book:
  - Two PCIe 3.0 x16 (x16 wired)
  - One PCIe 2.0 x8 slot (x4 wired)

Both I/O Books accommodate full height adapters.

Scalability

The x3850 X6 and x3950 X6 servers use native QPI scaling capabilities to achieve four-socket and 8-socket configurations. Unlike eX5 systems, there are no external connectors and cables for X6 systems, all interconnects are integrated in the midplane.

Figure 23 shows QPI connectivity between processors for the 4-socket x3850 X6 server and 8-socket x3950 X6 server.

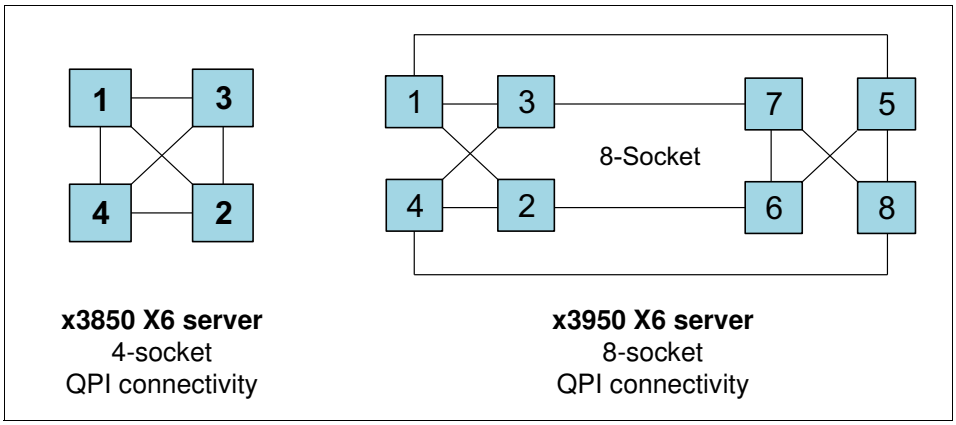

Figure 23 The x3850 X6 server QPI connectivity

The 8-socket configuration that is used in the x3950 X6 server has different connectivity schema with maximum one additional QPI hop between some processors.

The x3950 X6 server can be operated as a single system or as two independent systems, and this feature can be configured remotely without physically accessing the systems. This capability is called *partitioning* and is referred to as *FlexNode technology*. The partitioning is performed by the Integrated Management Module II, and it is configurable remotely through the integrated management module (IMM) web interface.

X6 server RAS features

The ultimate design goal for X6 servers as a part of the Enterprise X-Architecture strategy is to prevent hardware faults from causing an outage. Part selection for reliability, redundancy, recovery, self-healing techniques, and degraded operational modes are used in a fault resilient RAS strategy to avoid application outages.

Hardware fault tolerance

The X6 servers offer a number of hardware-based fault tolerant features, such as:

- Redundant hot-swap power supplies and fans
- Redundant hot-swap disk and solid-state drives with the RAID adapters
- I/O adapter fault-tolerance and hot-replace capabilities
- Integrated Management Module II (IMM2)
- Predictive Failure Analysis
- Light path diagnostics

IMM2 is a part of the RAS strategy. It supports error prevention with Predictive Failure Analysis (PFA), server self-healing, concurrent maintenance, and light path diagnostics for minimizing servicing downtime. In addition, IMM provides self-healing capabilities for the system by isolating faults and restarting a server in degraded mode (for example, if a processor failure occurred, the IMM disables it and reboots the server by using the remaining processors, if possible).

PFA for System x® servers is a collection of techniques that help you run your business with less unscheduled downtime. PFA allows the server to monitor the status of critical subsystems and to notify the system administrator when components appear to be degrading. In most cases, replacement of failing parts can be performed as part of planned maintenance activity. As a result, unscheduled outages can be prevented. Parts identified by PFA are covered for the full duration of the warranty period.

PFA features on the X6 servers include:

- System memory
- System fans
- Processors
- Power supplies
- Voltage regulators on the system board
- Hard disk drives

Each feature is designed to provide local and remote warnings of impending failures that might cause unscheduled downtime. The potential benefit is higher server availability and reduced total cost of ownership (TCO).

Light path diagnostics allows systems engineers and administrators to easily and quickly diagnose hardware problems on the System x servers. Hardware failures that in the past might have taken hours to locate and diagnose can be detected in seconds, avoiding or reducing downtime.

Light path diagnostics constantly monitors selected components within the System x server. If a failure occurs, a light is illuminated on the front panel of the server to alert the systems administrator that there is a problem. An LCD panel on the front of the server indicates the failed component.

Light path diagnostics is performed at a hardware level, and it does not require an operating system to function. These characteristics make it reliable and allow you to diagnose problems before an operating system is even installed.

Intel Xeon processor RAS features

Intel Xeon processor E7 family of processors has additional reliability, availability, and serviceability (RAS) features on their interconnect links (SMI and QPI):

- Cyclic redundancy checking (CRC) on the QPI links

  The data on the QPI link is checked for errors.

- QPI packet retry

  If a data packet on the QPI link has errors or cannot be read, the receiving processor can request that the sending processor try resending the packet.

- QPI clock failover

  If there is a clock failure on a coherent QPI link, the processor on the other end of the link can become the clock. In some cases, clock failover can result in both speed and width reduction.

- QPI self-healing

  If there are persistent errors detected on a QPI link, the link width can be reduced dynamically to allow the system to run in a degraded mode until repair can be performed.

- Scalable memory interconnect (SMI2) packet retry

  If a memory packet has errors or cannot be read, the processor can request that the packet be resent from the memory buffer.

- SMI2 clock and lane failover

  If there are persistent errors detected on a SMI2 link, the link width can be reduced dynamically to allow the system to run in a degraded mode until repair can be performed.

Machine Check Architecture recovery

The Intel Xeon processor E7 family also features Machine Check Architecture (MCA) recovery, a RAS feature that enables the handling of system errors that otherwise require that the operating system be halted. For example, if a dead or corrupted memory location is discovered, but it cannot be recovered at the memory subsystem level, and provided it is not in use by the system or an application, an error can be logged and the operation of the server can continue. If it is in use by a process, the application to which the process belongs can be stopped or informed about the situation.

Implementation of the MCA recovery requires hardware support, firmware support (such as found in the UEFI), and operating system support. Microsoft, SUSE, Red Hat, VMware, and other operating system vendors can describe the Intel MCA recovery feature on the Intel Xeon processors in their latest operating system versions.

Figure 24 illustrates an example of how an application can use the MCA recovery feature to handle system errors to prevent itself from being terminated in case of a system error.

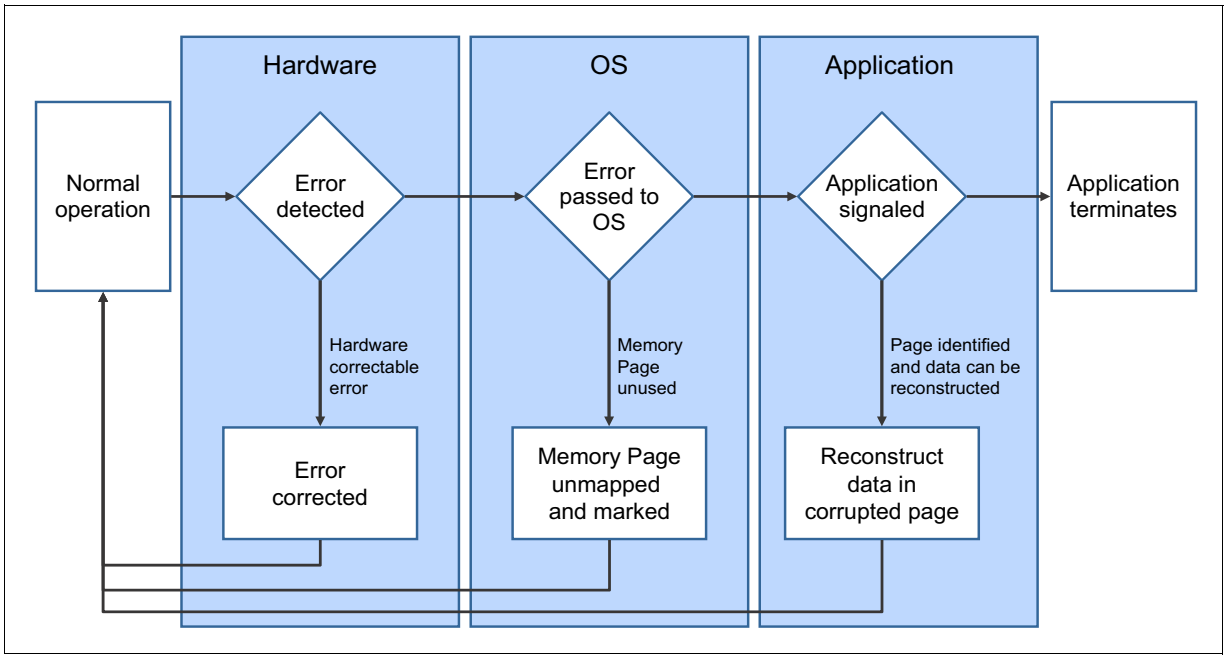

Figure 24 Intel Machine Check Architecture recovery example

If a memory error is encountered that cannot be corrected by the hardware, the processor sends an MCA recovery signal to the operating system. An operating system that supports MCA recovery now determines if the affected memory page is in use by an application. If unused, it unmaps the memory page and marks it as in error. If the page is used by an application, the OS usually holds that application or stops all processing and halts the system.

With the MCA-aware application, the operating system can signal the error situation to the application, giving it the chance to try to repair the effects of the memory error. Using MCA recovery, the application can take an appropriate action at the level of its own data structures to ensure a smooth return to normal operation, avoiding a time-consuming restart or loss of information.

New MCA recovery features of the Intel Xeon E7 processor include:

- Execution path recovery: Ability to work with hardware and software to recognize and isolate the errors that were delivered to the execution engine (core).
- MCA I/O: Ability to report uncorrectable (both fatal and non-fatal) errors to the software layers for further handling, such as determining the root cause of failure or preventive maintenance.
- Enhanced MCA (eMCA) Gen 1: Provides enhanced error log information to the operating system, hypervisor, or application that can be used to provide better diagnostic and predictive failure analysis for the system. This enables higher levels of uptime and reduced service costs.

Memory RAS features

You can enable various memory reliability, availability, and serviceability (RAS) features from the Unified Extensible Firmware Interface (UEFI) shell. These settings can increase the reliability of the system; however, there might be performance trade-offs when these features are enabled. The following sections provide a brief description of each memory RAS technology.

Chipkill

Chipkill memory technology, an advanced form of error correction code (ECC) from Lenovo, is available for the X6 servers. Chipkill (also known as Single Device Data Correction or SDDC) protects the memory in the system from any single memory chip failure. It also protects against multibit errors from any portion of a single memory chip. Chipkill on its own is able to provide 99.94% memory availability[1] to the applications without sacrificing performance and with standard ECC DIMMs.

Redundant bit steering and double device data correction

Memory ProteXion or *redundant bit steering* provides the equivalent of a hot-spare drive in a RAID array. It is based in the memory controller and senses when a chip on a DIMM fails and when to route the data around the failed chip.

Within the system, the models of the X6 servers using the Intel Xeon processor E7 family support the Intel implementation of Chipkill plus redundant bit steering, which Intel calls *double device data correction* (DDDC). Redundant bit steering uses the ECC coding scheme that provides Chipkill coverage for x4 DRAMs. This coding scheme leaves the equivalent of one x4 DRAM spare in every pair of DIMMs. If a chip failure on the DIMM is detected, the memory controller can reconstruct data from the failed chip to the x4 spare.

Redundant bit steering operates automatically without issuing a Predictive Failure Alert (PFA) or light path diagnostics alert to the administrator, although an event is logged to the service processor log. In RAS mode, after the second DRAM chip failure on the DIMM, additional single bit errors result in PFA and light path diagnostics alerts.

[1] *A White Paper on the Benefits of Chipkill-Correct ECC for PC Server Main Memory* by Timothy Dell

Advanced Page Retire

Advanced Page Retire is a unique algorithm to handle memory errors. It is a built-in, sophisticated error handling firmware that uses and co-ordinates memory recovery features by balancing the goals of maximum up time and minimum repair actions. The algorithm uses short- and long-term thresholds per memory rank with leaky bucket and automatic sorting of memory pages with the highest correctable error counts.

It uses hardware recovery features, followed by software recovery features, to optimize recovery results for both newer and older operating systems and hypervisors. When recovery features are exhausted, firmware issues a PFA. Memory that has failed completely is held offline during reboots until repaired. Failed DIMMs are indicated by light path diagnostics LEDs that are physically at the socket location.

Lenovo performs thorough testing to verify the features and co-ordination between the firmware and the operating system or hypervisor.

Memory mirroring

To improve memory reliability and availability beyond ECC and Chipkill (see “Chipkill” on page 41), the memory controller can mirror memory data to two memory channels. To enable mirroring, you must have both memory channels populated with the same DIMM type and amount of memory.

Memory mirroring, or *full array memory mirroring (FAMM) redundancy,* provides a redundant copy of all code and data addressable in the configured memory map. Memory mirroring works within the chip set by writing data to two memory ports on every memory-write cycle. Two copies of the data are kept, similar to the way RAID 1 writes to disk. Reads are interleaved between memory ports. The system automatically uses the most reliable memory port as determined by error logging and monitoring.

If errors occur, only the alternative memory port is used until bad memory is replaced. Because a redundant copy is kept, mirroring results in only half the installed memory being available to the operating system. FAMM does not support asymmetrical memory configurations and requires that each port is populated in identical fashion. For example, you must install 2 GB of identical memory equally and symmetrically across the two memory ports to achieve 1 GB of mirrored memory.

Memory mirroring is independent of the operating system. There is a memory performance trade-off when memory mirroring is enabled.

Memory rank sparing

Sparing provides a degree of redundancy in the memory subsystem, but not to the extent of mirroring. In contrast to mirroring, sparing leaves more memory for the operating system. In sparing mode, the trigger for failover is a preset threshold of correctable errors. When this threshold is reached, the content is copied to its spare. The failed rank is then taken offline, and the spare counterpart is activated for use.

In rank sparing mode, one rank per memory channel is configured as a spare. The ranks must be as large as the rank relative to the highest capacity DIMM for which we are setting up sparing. The size of the two unused ranks for sparing is subtracted from the usable capacity that is presented to the operating system.

Memory sparing is independent of the operating system. There is a memory performance trade-off when memory sparing is enabled.

Upward integration

Upward Integration Modules provide IT administrators with the ability to integrate the management features of the System x servers with the third-party virtualization and systems management tools. Lenovo expands the management capabilities of these environments with System x hardware management functionality, providing affordable, basic management of physical and virtual environments to reduce the time and effort required for routine system administration. It provides the discovery, configuration, event management, and monitoring that is needed to reduce cost and complexity through server consolidation and simplified management.

The following Upward Integration modules are currently available:

- Upward Integration for VMware vCenter
- Upward Integration for Microsoft System Center

The Upward Integration for VMware vCenter includes the following features:

- Provide an overview of the host status, including a system information summary and system health messages.
- Collect and analyze system information to help diagnose system problems.
- Acquire and apply the latest UpdateXpress System Packs™ and individual firmware updates to your ESXi system.
- Monitor and provide a summary of power usage, thermal history, and fan speed and a trend chart of the managed host. Enable or disable the Power Metric function on a host and set the power capping for a power-capping capable host to limit the server power usage. Power throttling is supported, and notification can occur if the server power usage exceeds the specific value.
- Manage the current system settings on the host including IMM, UEFI, and boot order settings for the host.
- Monitor the server hardware status, and receive predictive failure alerts. Enable predictive failure management, and set a management policy for a server based on a predictive failure alert.
- Automate VM migration based on PFAs.
- Schedule rolling firmware updates to a cluster of servers.

Upward Integration for VMware vSphere provides a 90-day trial license by default for these features. When the license expires after 90 days, all of the premium features are disabled. Install the Upward Integration for VMware vSphere license tool to activate the product license. Activation licenses can be purchased by contacting a Lenovo representative or a Lenovo Business Partner.

The Upward Integration for Microsoft System Center provides the following capabilities:

- Monitor both physical and virtual server health through an integrated end-to-end management of System x hardware
- Deploy OS with the latest System x firmware and driver update management
- Receive PFAs and enable predictive failure management.
- Automate VM migration based on server health or power consumption
- Perform hardware configuration and firmware or driver updates and check for the latest updates from the Service and Support website
- Collect Lenovo-specific hardware inventory of System x servers

- Power on and off blades using the Microsoft System Center Console
- Author configuration packs to perform compliance checking on System x servers
- Manage servers remotely independent of operating system state

Upward Integration for Microsoft System Center can be purchased as a one year or 3-year software service and maintenance license. Activation licenses can be purchased by contacting a Lenovo representative or a Lenovo Business Partner.

Summary

The Lenovo X6 family of scalable rack servers consists of the Lenovo System x3850 X6 server, a 4U four-socket server, and the x3950 X6 server, an 8U eight-socket server. Using proven technologies of the previous generations of Enterprise X-Architecture, these servers introduce:

- New levels of fault tolerance and resiliency with advanced RAS features implemented in hardware and software
- Agility and scalability with fit-for-purpose modular “bookshelf” that is ready to support multiple technology upgrades
- Significant improvements in response time with ultralow latency, stretched memory speeds that exceed Intel specifications, and innovative flash memory-channel storage offerings

Lenovo X6 servers continue to lead the way as the shift toward mission-critical scalable databases, business analytics, virtualization, enterprise applications, and cloud accelerates.

Lenovo X6 systems using the Intel Xeon processor E7 v2 family deliver an extensive and robust set of integrated advanced RAS features that prevent hardware faults from causing an outage. Part selection for reliability, redundancy, recovery, and self-healing techniques and degraded operational modes are used in a RAS strategy to avoid application outages. Using this strategy, Lenovo X6 systems can help increase application availability and reduce downtime by enabling 24x7 mission-critical capabilities to run your core business services.

Related publications

The following publications provide additional information about the topic in this document.

- *Lenovo System x3850 X6 Product Guide,* TIPS1250

  http://lenovopress.com/tips1250
- *Lenovo System x3950 X6 Product Guide,* TIPS1251

  http://lenovopress.com/tips1251
- Lenovo System x3850 X6 and x3950 X6 Planning and Implementation Guide

  http://lenovopress.com/sg248208

Author

Ilya Krutov is a Project Leader at Lenovo Press. He manages and produces pre-sale and post-sale technical publications on various IT topics, including x86 rack and blade servers, server operating systems, virtualization and cloud, networking, storage, and systems management. Ilya has more than 15 years of experience in the IT industry, backed by professional certifications from Cisco Systems, IBM, and Microsoft. During his career, Ilya has held a variety of technical and leadership positions in education, consulting, services, technical sales, marketing, channel business, and programming. He has written more than 200 books, papers, and other technical documents. Ilya has a Specialist's degree with honors in Computer Engineering from the Moscow State Engineering and Physics Institute (Technical University).

Thanks to the following people for their contributions to this project:

- Randy Kolvick
- Randall Lundin
- Doug Petteway
- David Watts

Notices

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

Lenovo 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:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO 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 jurisdictions 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. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

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

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

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

© Copyright Lenovo 2016. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by Global Services Administration (GSA) ADP Schedule Contract

This document was created or updated on September 3, 2016.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/redp5059

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo 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 Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at http://www.lenovo.com/legal/copytrade.html.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

eX5™	Lenovo®	System x®
eXFlash™	Lenovo(logo)®	UpdateXpress System Packs™
Flex System™	ServeRAID™	

The following terms are trademarks of other companies:

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

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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

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