Ilya Krutov

Reliability, Availability, and Serviceability Features of the IBM eX5 Portfolio

This IBM® Redpaper™ publication describes advanced reliability, availability, and serviceability (RAS) features that are available in IBM eX5 server systems. It describes how those features differ from standard RAS features found in other dual-socket systems, and what advantages they provide for building a resilient and 24x7 operational environment.

This paper explains what RAS is, why it is important, and how IBM approaches RAS in system design. It then provides an architectural overview and describes the RAS features of IBM eX5 servers, and describes the differences between advanced and standard RAS features and their impact on overall system availability. It provides an example that illustrates IBM eX5 availability advantages by comparing the scale up approach to the scale out approach.

The following topics are covered:

- RAS: Reliability, availability, and serviceability
- IBM eX5 architecture and portfolio overview
- RAS features: Advanced versus standard
- Availability evaluation
- Conclusions

Executive summary

Core mission-critical business services and applications deployed in modern data centers must be available 24 hours per day, 7 days per week (24x7). If such an application stops because a certain event occurred and the mission-critical service is disrupted, then this downtime can be costly. This situation is especially important in the case of large-scale server deployments where, statistically, component failures might occur often. Therefore, server reliability, availability, and serviceability (RAS) features become important in a platform that hosts mission-critical applications.

IBM eX5 systems, which use the Intel Xeon processor E7 family of processors, deliver an extensive and robust set of integrated RAS features to prevent hardware faults from causing an outage. Part selection for reliability, redundancy, recovery, and self-healing techniques and degraded operational modes are used in an IBM RAS strategy to avoid application outages. With this strategy, IBM eX5 systems can increase application availability and reduce downtime, enabling the 24x7 mission-critical capabilities that run your core business services.

Key availability features of IBM eX5 systems include:

- Extensive memory protection with IBM Chipkill, Redundant Bit Steering (RBS), and Double Device Data Correction (DDDC).

  The combination of IBM Chipkill and RBS (also known as DDDC) provides robust memory protection that sustains to two sequential memory DRAM chip failures without affecting overall system performance.

- Redundant processor-to-I/O hub interconnect links enable self-recovery from a processor failure.

  If the primary processor (the one used for booting the operating system) fails, then an eX5 system can use a second processor to boot the OS, as the system has still access to the integrated I/O devices through the redundant links between the processors and I/O hubs.

- Single image 8-way systems that have two interconnected 4-way nodes that provide self-healing capabilities in case of a single node failure.

  Two independent nodes form a resilient 8-way configuration. In a single node failure, the system can be restarted in degraded mode, thus eliminating downtime that requires a service specialist to perform recovery.

- Predictive Failure Analysis (PFA), also found in high volume IBM System x® servers.

  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, replacements of failing parts can be performed as part of planned maintenance activity. As a result, unscheduled outages can be prevented and your system continues to run.

As a result, in scalable implementations, IBM eX5 systems can help you achieve more than three times better overall availability (as expressed in the number of repair actions compared to standard systems), reducing the number of system outages related to hardware failures and lowering maintenance and support costs.

RAS: Reliability, availability, and serviceability

Core business services and applications deployed in modern data centers must be available 24x7. Because of their 24x7 availability requirements, we call them *mission-critical applications*. If such an application stops because a certain event occurred and the mission-critical service is disrupted, then this downtime can be costly. This situation is especially important in the case of large-scale server deployments where, statistically, component failures might occur often. Typical examples of mission-critical applications include core databases, business intelligence, corporate applications, such as Enterprise Resource Planning (ERP), and consolidated virtualized environments that host multiple business-critical application and infrastructure services. These environments are the ones where RAS is important; better RAS means fewer outages.

Regarding failures in computing systems, in most cases, system downtime is caused by server hardware failures, and memory-related failures are the top reasons for these failures[1]. Therefore, we focus on hardware failures in the servers and what IBM offers to prevent them or minimize their impact on a production environment.

There are many ways to define reliability, availability, and serviceability. A useful definition of RAS for server hardware is as follows:

- *Reliability*: How infrequently a defect or fault is seen in a server.
- *Availability*: How infrequently the functionality of a system or application is impacted by a fault or defect.
- *Serviceability*: How well faults and their impacts are communicated to users and service personnel and how efficiently and nondisruptively they are repaired.

Using this definition, reliability in hardware is about how often a hardware fault requires a system to be serviced. The less frequent the failures, the greater the reliability. Availability is how infrequently such a failure impacts the operation of the system or application. For high levels of availability, correct system operation must not be affected adversely by hardware faults. A highly available system design ensures that most hardware failures do not result in an application outage. Serviceability concerns itself with identifying what fails and ensuring an efficient repair.

Ultimately, no matter how well it is designed and built, hardware fails. The intent of 24x7 availability is to reduce the impact of these failures on system operations. IBM traditionally classifies hardware failures in multiple ways:

- Repair Action (RA): RAs are related to the industry standard definition of Mean Time Between Failure (MTBF). A RA is any hardware event that requires service on a system. Repair actions include incidents that affect system availability and incidents that are concurrently repaired.
- Interrupt Repair Action (IRA): An IRA is a hardware event that requires a scheduled system outage to repair.

[1] *A Large-Scale Study of Failures in High-Performance Computing Systems*, found at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4775906&isnumber=5624749

- Unscheduled Incident Repair Action (UIRA): A UIRA is a hardware event that causes a system to be rebooted in full or degraded mode. The system experiences an unscheduled outage. The restart might include some level of capability loss, but the remaining resources are made available for productive work.
- High Impact Outage (HIO): A HIO is a hardware failure that triggers a system crash that is not recoverable by immediate reboot. This failure is usually caused by failure of a component that is critical to system operation and is, in some sense, a measure of system single points of failure. HIOs result in the most significant availability impact on the system, because repairs cannot be effected without a service call.

The clear design goal for IBM System x servers as a part of the IBM X-Architecture® strategy is to 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 (Figure 1).

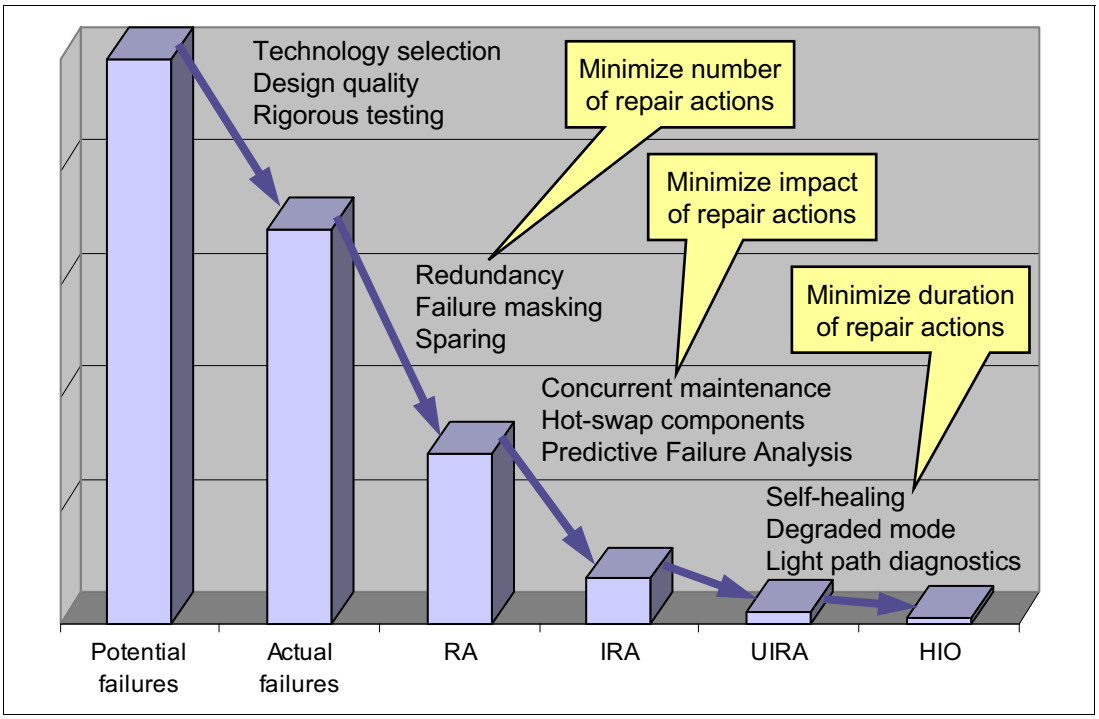

Figure 1 IBM RAS strategy

The core principles guiding IBM engineering design are reflected in the RAS architecture. The goal of any server design is to:

- Achieve a highly reliable design through extensive use of highly reliable components and rigorous compatibility testing.
- Clearly identify, early in the server design process, those components that have the highest opportunity for failure. Employ a server architecture that allows the system to recover from intermittent errors in these components and fail over to redundant components when necessary.
- Develop server hardware that can detect and report on failures and impending failures.
- Create server hardware that is self-healing, that is, it automatically initiates actions to effect error correction, repair, or component replacement.

Hardware failures can be grouped into two categories:

- *Transient* (or soft) errors
- *Permanent* (or hard) errors

Soft errors include faults caused by random events that affect electronic circuits, such as power line disturbances, static electrical discharge, alpha particles, energy released into circuits by cosmic radiation, and occasional circuit disturbance due to other sources of random noise. Soft errors are most common for DRAM modules, but they can also occur on interconnects such as memory and I/O buses. Soft errors can be corrected, if detected, without replacing a hardware component. In general, there are two approaches to correct soft errors: error correcting code (for example, ECC DIMMs and memory data path) and checksum validation with retry operations (for example, a cyclic redundancy check (CRC) on PCIe and QPI links, and a parity check on the processor's memory address bus).

Hard errors represent a permanent failure of the component or its part. They can be corrected only by replacing the failed component either physically (for example, replacing a failed power supply) or logically by using a previously enabled spare component (for example, IBM Chipkill together with IBM Memory ProteXion enables the usage of a spare memory chip on a DIMM module).

Standard approaches to server RAS focus on errors that can be handled at the hardware level. In such a case, any uncorrectable error causes a system outage. In this respect, another design element of the IBM RAS strategy is to enable software layers (operating system, hypervisors, and applications) to participate in the recovery process. The idea is to make the software that runs on a system both hardware fault-aware (both soft and hard errors) and use its capabilities to prevent service failure or to restore service operations. This combination of hardware and software self-recovery techniques are part of advanced RAS features that increase the availability of services that must be 24x7.

Figure 2 shows the different operational states of a server that follows the IBM RAS strategy.

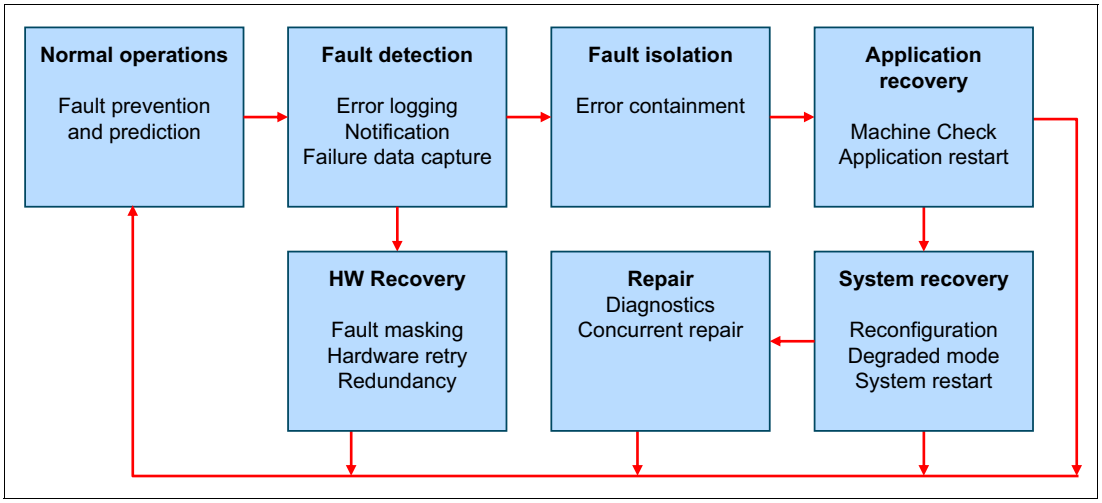

Figure 2 IBM server system RAS operations

One more thing to consider is which building block approach is used for creating a scalable server infrastructure. There are two types of the scalable system design approach:

- Scale up
- Scale out

The scale up approach uses 4-way SMP server building blocks or large systems with highly scalable shared internal resources and advanced RAS capabilities, while the scale out approach is based on 2-way server building blocks with dedicated resources and standard RAS capabilities.

There is no strict rule about what approach is better, as each of them has its own advantages and drawbacks. The approach you use depends on the specific client environment and requirements. However, from the RAS point of view, the scale up approach appears to be more reliable in terms of number of hardware failures compared to the scale out approach. The building blocks are equipped with advanced RAS features and have fewer physical components (processors, system boards, adapters, hard disk drives, fans, and power supplies).

We focus on a scale up approach and its building blocks. For more information about this RAS concept, see “RAS features: Advanced versus standard” on page 9.

IBM eX5 architecture and portfolio overview

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

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

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

The IBM eX5 product portfolio is built on the Intel Xeon processor E7-8800/4800/2800 product families. With the inclusion of these processors, the eX5 servers became faster, more reliable, and more power efficient.

IBM eX5 systems

The four systems in the eX5 family are the IBM System x3850 X5, IBM System x3950 X5, IBM System x3690 X5, and the IBM BladeCenter® HX5 blade. The eX5 technology is primarily designed around three major workloads: database servers, server consolidation by using virtualization services, and Enterprise Resource Planning (application and database) servers. Each system can scale by adding more memory, and you add this memory by adding an IBM MAX5 memory expansion unit to the server. The x3850 X5 and x3950 X5 servers can also scale from four to eight sockets, and the HX5 scales from two to four sockets by combining two like servers into a single two-node system.

Figure 3 shows the IBM eX5 family.

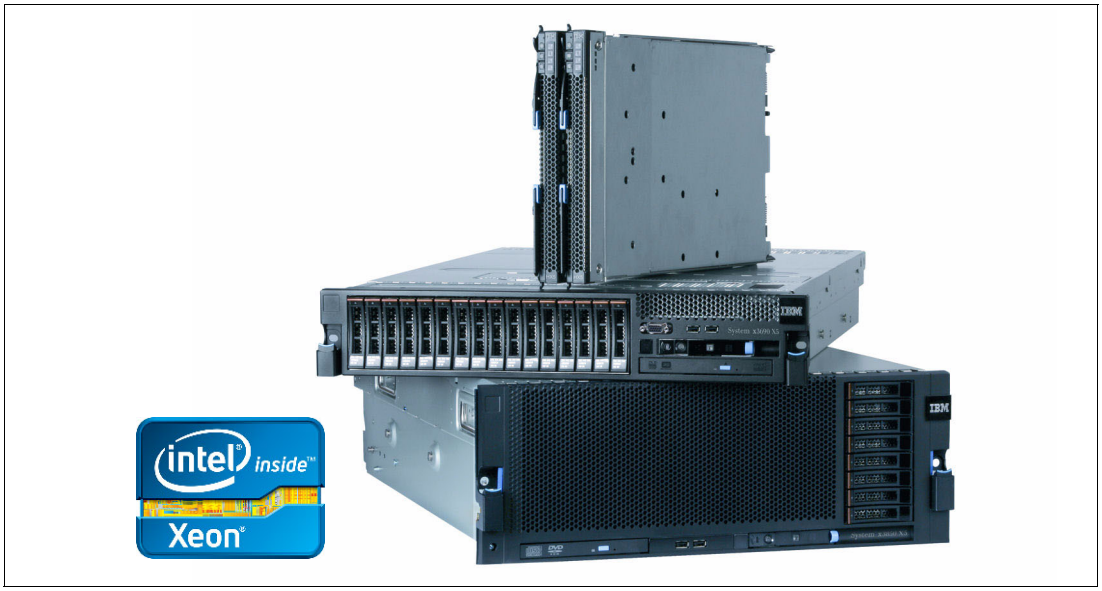

Figure 3 eX5 family (top to bottom): IBM BladeCenter HX5 (2-node), IBM System x3690 X5, and IBM System x3850 X5 (the IBM System x3950 X5 looks the same as the x3850 X5)

When compared to other machines in the IBM System x portfolio, these systems represent the upper end of the spectrum, are suited for the most demanding x86 tasks that require mission-critical availability, and can handle jobs that previously might have run on other platforms.

Table 1 gives an overview of the processor and memory scalability features of IBM eX5 systems.

Table 1 Maximum configurations for the eX5 systems

Maximum configurations		**x3850 X5/x3950 X5**	**x3690 X5**	**HX5**
Processors	1-node	4	2	2
	2-node	8	Not available	4
Memory	1-node	2048 GB (64 DIMMs)[a]	1024 GB (32 DIMMs)[b]	256 GB (16 DIMMs)
	1-node with MAX5	3072 GB (96 DIMMs)[a]	2048 GB (64 DIMMs)[b]	640 GB (40 DIMMs)
	2-node	4096 GB (128 DIMMs)[a]	Not available	512 GB (32 DIMMs)
	2-node with MAX5	6144 GB (192 DIMMs)[a]	Not available	Not available

a. Requires full processors to install and use all memory.
b. Requires that the memory mezzanine board is installed along with processor 2.

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

Figure 4 shows how you can use eX5 system scaling to connect two x3850 X5 servers and two MAX5 memory expansion units together to form a single system image. This image has up to eight Intel Xeon processor E7 family processors and up to 6 TB of memory.

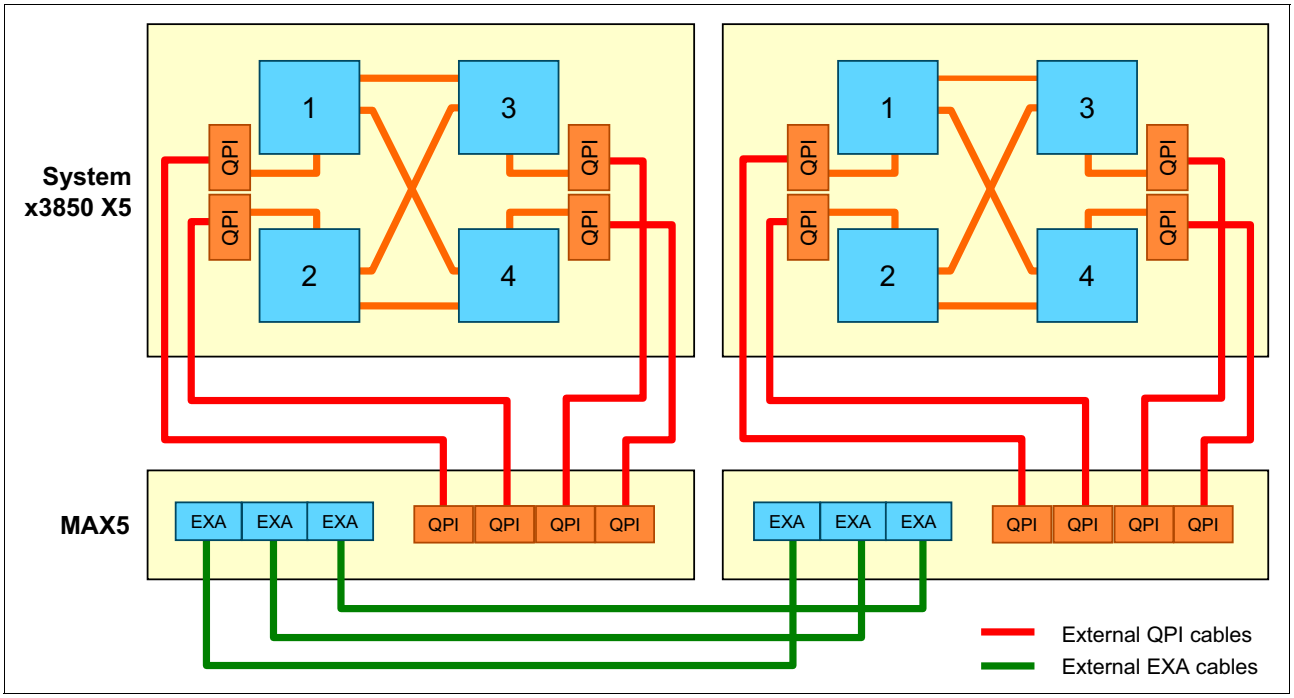

Figure 4 IBM System x3850 X5 scaling

Intel Xeon processors

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

The three groups of the Intel Xeon processor E7 family support scaling to different levels:

- The Intel Xeon processor E7-2800 product family is used in the x3690 X5 and BladeCenter HX5 systems. This family supports only two processor configurations, so it cannot be used in a two-node HX5 configuration. Most processors in this family support connection to a MAX5 system (except Intel Xeon processor E7-2803 and Intel Xeon processor E7-2820).
- The Intel Xeon processor E7-4800 product family is primarily used in the HX5 and x3850 X5 systems. This family supports four-processor configurations, so it can be used for single-node x3850 X5 and two-node HX5 systems. All of the members of the Intel Xeon processor E7-4800 product family support connection to a MAX5 system, and can also be used for two-node x3850 X5 systems with MAX5 configurations.
- The Intel Xeon processor E7-8800 product family is used in x3850 X5 systems to scale to two nodes with or without MAX5 configurations. There are specific high frequency and low-power models of this processor available for the x3690 X5 and HX5 systems as well.

These scalability capabilities are summarized in Table 2.

Table 2 Comparing the scalability configurations of the Intel Xeon processor E7 family

System	Intel Xeon processor E7-2800 product family	Intel Xeon processor E7-4800 product family	Intel Xeon processor E7-8800 product family
x3690 X5	Yes	Yes	Yes
x3690 X5 with MAX5	Yes[a]	Yes	Yes
HX5	Yes	Yes	Yes
HX5 with MAX5	Yes[a]	Yes	Yes
HX5 2-node	Not supported	Yes	Yes
x3850 X5	Not supported	Yes	Yes
x3850 X5 with MAX5	Not supported	Yes	Yes
x3850 X5 2-node without MAX5	Not supported	Not supported	Yes
x3850 X5 2-node with MAX5	Not supported	Yes	Yes

a. Intel Xeon processor E7-2803 and Intel Xeon processor E7-2820 do not support MAX5 configurations.

For more information about IBM eX5 portfolio, see *IBM eX5 Portfolio Overview: IBM System x3850 X5, x3950 X5, x3690 X5, and BladeCenter HX5*, REDP-4650.

RAS features: Advanced versus standard

In this section, we describe the differences in standard and advanced RAS features and how they impact overall server infrastructure availability. For our purpose, we use an IBM System x3850 X5 serve using the Intel Xeon processor E7 family to illustrate the advanced RAS capabilities of IBM eX5 systems in a scale up solution. Standard RAS features are commonly found in Intel Xeon processor E5 family-based platforms such as IBM System x3650 M4 system, and we use it as an example to highlight standard RAS capabilities with a scale out approach. We also refer to the advanced RAS features of the x3690 X5 system and how they compare to the traditional dual socket platforms using members of the Intel Xeon processor E5 family.

There is no single answer to the question about what type of system and type of RAS to choose; it depends on specific client requirements and the environment. Although advanced RAS provides application availability because of better hardware availability, it also costs more. At the same time, application availability can be provided by an infrastructure software (for example, high availability or load balancing clusters) and there might not be any need for advanced RAS features in such a case.

Table 3 summarizes the RAS features available in x3850 X5, x3690 X5, and x3650 M4 systems.

Table 3 x3850 X5, x3690 X5 and x3650 M4 RAS feature comparison

Feature	x3850 X5	x3690 X5	x3650 M4
Intel Xeon processor product family	E7-4800/8800	E7-2800/4800/8800	E5-2600
Maximum processor quantity	8	2	2

Feature	x3850 X5	x3690 X5	x3650 M4
DIMM slots	192	64	24
Maximum memory capacity	6 TB	2 TB	768 GB
RAS type	Advanced	Advanced	Standard
Processor RAS			
ECC cache	Yes	Yes	Yes
Memory address parity	Yes	Yes	Yes
Resilient QPI links	Yes	Yes	Yes
Redundant processor-to-I/O hub links	Yes	Yes	No
Machine Check Architecture recovery	Yes	Yes	No
QPI RAS			
QPI protocol CRC protection	Yes	Yes	Yes
QPI self-healing	Yes	Yes	No
QPI clock failover	Yes	Yes	No
QPI packet retry	Yes	Yes	Yes
Memory RAS			
IBM Chipkill	Yes	Yes	Yes
Redundant Bit Steering	Yes	Yes	No
Memory scrubbing	Yes	Yes	Yes
Memory mirroring	Yes	Yes	Yes
Memory DIMM sparing	Yes	Yes	No
Memory rank sparing	Yes	Yes	Yes
SMI RAS			
SMI protocol CRC protection	Yes	Yes	No
SMI lane failover	Yes	Yes	No
SMI clock failover	Yes	Yes	No
SMI packet retry	Yes	Yes	No
System-wide RAS features			
Hot-swap components (HDDs, power supplies, fans)	Yes	Yes	Yes
Integrated Management Module	Yes	Yes	Yes
Predictive Failure Analysis	Yes	Yes	Yes
Light path diagnostics	Yes	Yes	Yes
Dynamic System Analysis	Yes	Yes	Yes

The following subsections explain the RAS features listed in Table 3 on page 9 in more detail.

We begin with overall system architecture, then continue with memory RAS as the most important IBM eX5 RAS differentiation. We then describe interconnect links RAS followed by Machine Check Architecture. We then recap common RAS features across the entire IBM System x portfolio.

System architecture

The overall system architecture and component interconnects have a great impact on the RAS features of a server. Figure 5 shows the internal architecture and advanced RAS features of an IBM System x3850 X5 server.

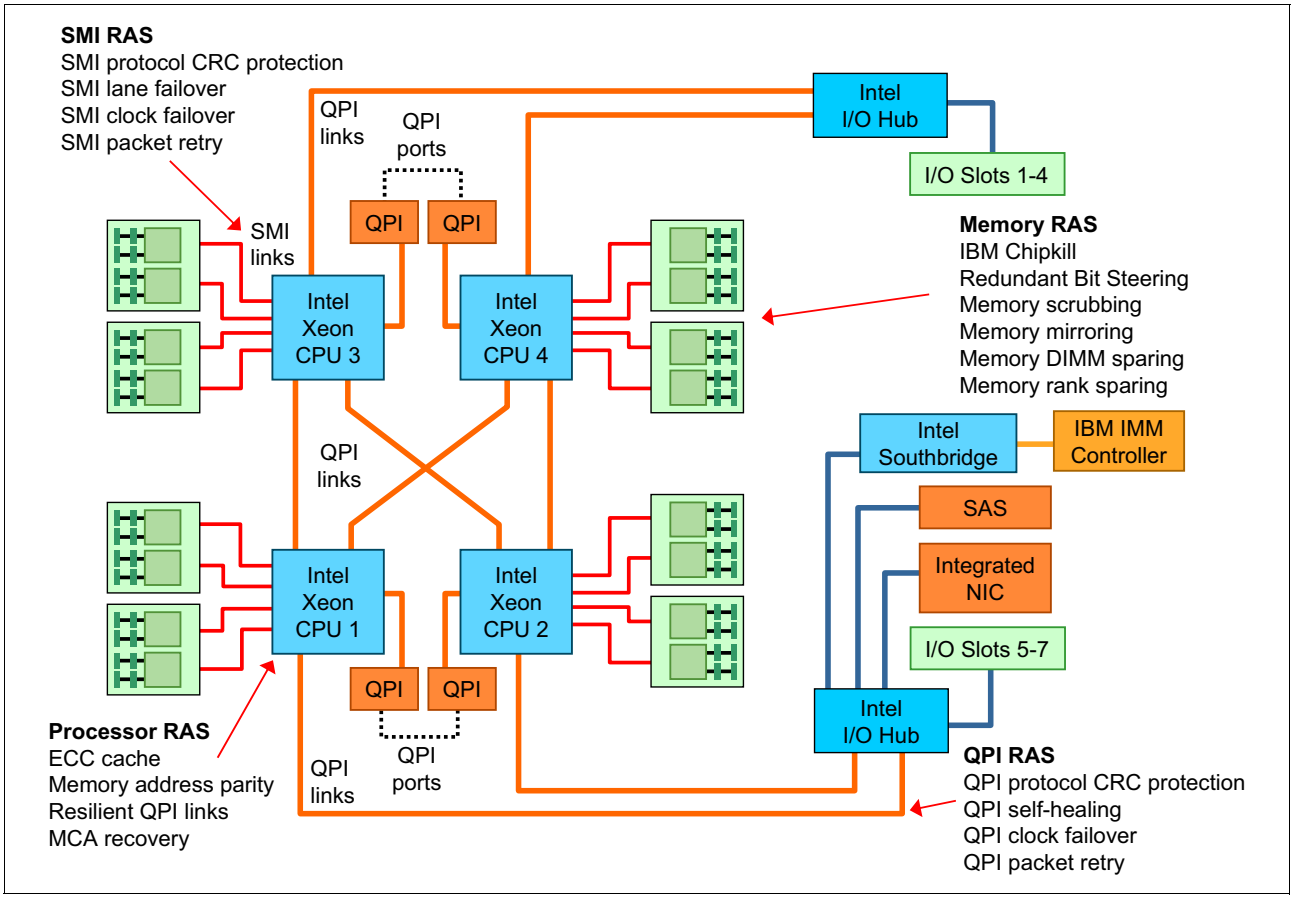

Figure 5 Advanced RAS features of an IBM System x3850 X5 server

The architecture and RAS features of an x3690 X5 server are shown in Figure 6.

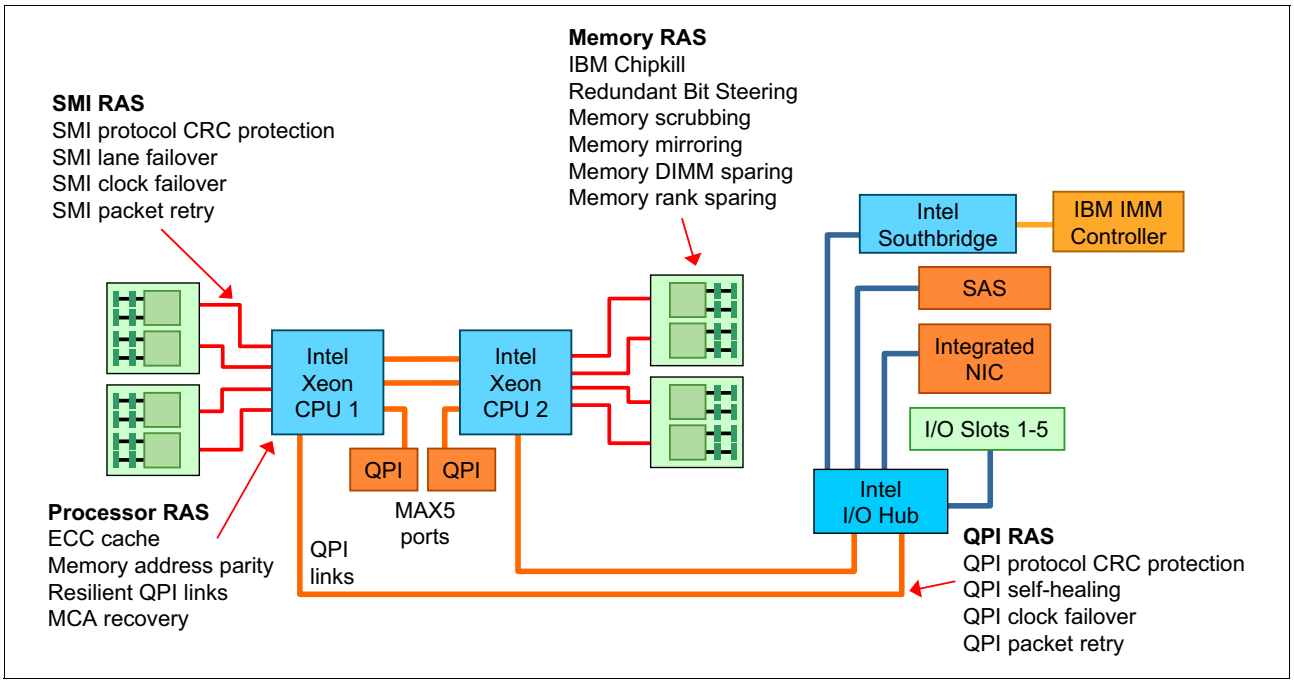

Figure 6 Advanced RAS features of x3690 X5 server

The internal system architecture of the HX5 blade is similar to the x3690 X5 server, but there are differences in the number of DIMM slots and I/O slots.

Figure 7 shows the x3650 M4 architecture and its standard RAS features.

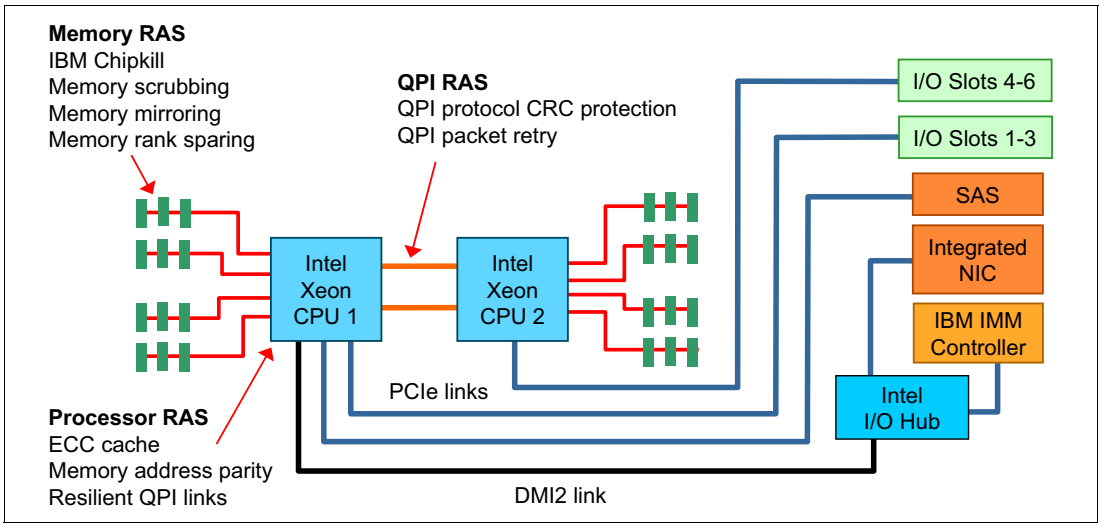

Figure 7 Standard RAS features of IBM System x3650 M4

As we compare the block diagrams listed above, we can see the clear differences in the systems' architecture, interconnect links, and RAS features. Specifically, eX5 systems that are based on the Intel Xeon processor E7 family have multiple redundant processor-to-I/O hub links to ensure that the boot device and integrated management module are available to the system in case of a processor failure. Unlike eX5 systems, traditional 2-way systems that are based on the Intel Xeon processor E5 family do not have such a feature, and in case of a processor 1 failure, they are not able to boot.

IBM eX5 systems also have better memory protection technologies, including both IBM Chipkill and Memory ProteXion (also known as Redundant Bit Steering) to recover from two sequential DRAM chip failures on a memory DIMM. QPI and SMI links support CRC error detection and hardware retry capabilities, as well as link self-healing and link failover. These points are important to take into account when comparing dual socket IBM x3690 X5 systems to the traditional 2-way systems, such as the x3650 M4 system.

Memory RAS

Various memory reliability, availability, and serviceability (RAS) features can be enabled 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.

IBM Chipkill

Chipkill memory technology, an advanced form of error checking and correcting (ECC) from IBM, is available for the eX5 and other 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 is able to provide 99.94% memory availability[2] to the applications without sacrificing performance and with standard ECC DIMMs.

Redundant bit steering and double device data correction

IBM Memory ProteXion or Redundant Bit Steering (RBS) provides the equivalent of a hot-spare drive in a RAID array. It is based in the memory controller, and it 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 eX5 servers using the Intel Xeon processor E7 family support the Intel implementation of Chipkill + RBS, which Intel calls *double device data correction* (DDDC).

The external MAX5 memory expansion unit also supports Chipkill and RBS when x4 memory DIMMs are used. The x8 DIMMs do not support RBS. RBS is automatically enabled in the MAX5 memory port if all DIMMs installed to that memory port are x4 DIMMs.

RBS 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 by memory scrubbing, the memory controller can reroute data around that failed chip through these spare bits. DIMMs using x8 DRAM technology use a separate ECC coding scheme that does not leave spare bits, which is why RBS is not available on x8 DIMMs.

RBS 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, PFA and light path diagnostics alerts occur on that DIMM normally.

The combination of IBM Chipkill and RBS provides robust memory protection that sustains to two sequential memory DRAM chip failures without affecting overall system performance.

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

Memory mirroring

To improve memory reliability and availability beyond error checking and correcting (ECC) and Chipkill (see “IBM Chipkill” on page 13), the chip set 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 is supported on both eX5 and 2-way systems.

Note: MAX5 is required to enable memory mirroring in two-node x3850 X5 configurations.

Memory mirroring, or *full array memory mirroring (FAMM) redundancy,* provides the user with 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 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. Depending on the type of sparing (DIMM or rank), when this threshold is reached, the content is copied to its spare. The failed DIMM or rank is then taken offline, and the spare counterpart is activated for use. There are two sparing options:

- DIMM sparing (supported on eX5 systems only)

  Two unused DIMMs are configured as spares per memory card. These DIMMs must have the same rank and capacity as the largest DIMMs for which we are setting up sparing. The size of the two unused DIMMs for sparing is subtracted from the usable capacity that is presented to the operating system. DIMM sparing is applied to all memory cards in the system.

- Rank sparing (supported on both eX5 and 2-way systems)

  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.

Interconnect links

The 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. This action is not required on the QPI links from processors to I/O hubs, as these links are asynchronous.

- 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 (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.

- SMI clock failover

  Directs forwarded clocks to the clock failover lane in the case of a forwarded clock failure.

- SMI lane failover

  When an SMI link exceeds the preset error threshold, it is disabled, and memory transfers are routed through the other SMI link to the memory buffer.

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 include support for the Intel MCA recovery feature on the Intel Xeon processors in their latest operating system versions.

SAP HANA is the first application that uses the MCA recovery feature to handle system errors to prevent the application from being terminated in case of a system error.

Figure 8 shows how SAP HANA uses the Machine Check Architecture recovery.

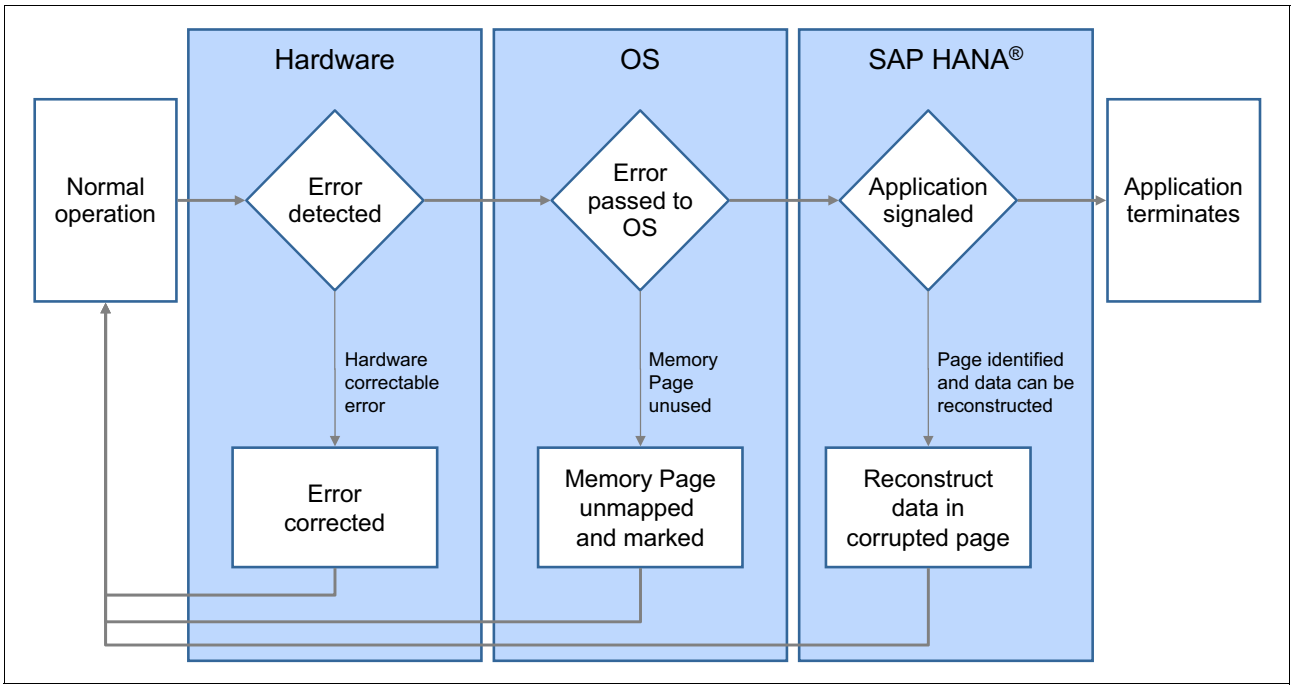

Figure 8 Intel Machine Check Architecture recovery with SAP HANA

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, such as SUSE Linux Enterprise Server (used in the SAP HANA appliance), 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 SAP HANA being MCA-aware, the operating system can signal the error situation to SAP HANA, giving it the chance to try to repair the effects of the memory error.

Using the knowledge of its internal data structures, SAP HANA can decide what course of action to take. If the corrupted memory space is occupied by one of the SAP in-memory tables, SAP HANA reloads the associated tables. In addition, it analyzes the failure and checks if it affects other stored or committed data, in which case it uses save points and database logs to reconstruct the committed data in a new, unaffected memory location.

Using MCA recovery, SAP HANA 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.

Common IBM System x RAS features

In this section, we describe Integrated Management Module, Predictive Failure Analysis, light path diagnostics, and Dynamic System Analysis.

Integrated Management Module

The Integrated Management Module (IMM) consolidates the service processor functionality, Super I/O, video controller, and remote presence capabilities into a single chip on the server system board. The IMM replaces the baseboard management controller (BMC) and Remote Supervisor Adapter II in IBM System x servers.

The IMM offers several improvements over the combined functionality of the BMC and the Remote Supervisor Adapter II:

- A dedicated Ethernet connection.
- One IP address for both the Intelligent Platform Management Interface (IPMI) and the service processor interface.
- Embedded Dynamic System Analysis (DSA).
- The ability to locally or remotely update other entities without requiring a server restart to initiate the update process.
- Remote configuration by using the Advanced Settings Utility (ASU).
- Capability for applications and tools to access the IMM either in-band or out-of-band.
- Enhanced remote-presence capabilities.

The IMM provides the following functions:

- Around-the-clock remote access and management of your server
- Remote management independent of the status of the managed server
- Remote control of hardware and operating systems
- Web-based management with standard web browsers

There are two types of IMM functionality:

- IMM Standard
- IMM Premium

IBM System x3690 X5 and x3850 X5 have premium functionality as a standard feature. Premium upgrade is optional for x3650 M4.

IMM Standard has the following features:

- Access to critical server settings
- Access to server vital product data (VPD)
- Advanced Predictive Failure Analysis (PFA) support
- Automatic notification and alerts
- Continuous health monitoring and control
- A choice of a dedicated or shared Ethernet connection
- Domain Name System (DNS) server support
- Dynamic Host Configuration Protocol (DHCP) support
- Email alerts
- Embedded Dynamic System Analysis (DSA)
- Enhanced user authority levels
- LAN over USB for in-band communications to the IMM
- Event logs that are time stamped, saved on the IMM, and can be attached to email alerts
- Industry-standard interfaces and protocols
- OS watchdogs
- Remote configuration through Advanced Settings Utility (ASU)
- Remote firmware updating
- Remote power control

- Seamless remote accelerated graphics
- Secure web server user interface
- Serial over LAN
- Server console redirection
- Simple Network Management Protocol (SNMP) support
- User authentication by using a secure connection to a Lightweight Directory Access Protocol (LDAP) server

In addition, IMM Premium has the following features (these features are standard on x3690 X5 and x3850 X5 servers, and optional on x3650 M4 servers):

- Remote presence, including remote control of a server
- Operating-system failure screen capture and display through the web interface
- Remote disk, which enables the attachment of a diskette drive, CD/DVD drive, USB flash drive, or disk image to a server

IMM is a part of the RAS strategy. It supports error prevention with 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).

Predictive Failure Analysis

PFA for IBM 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.

Providing peace of mind, IBM stands behind its server products with a 3 year, on-site limited warranty (standard with many servers, although entry systems come with a 1 year warranty that can be optionally extended to 3 years). Labor and IBM parts are covered for the full duration of the warranty period, including parts identified during PFA.

PFA uses a dedicated systems management processor (IMM or Integrated Management Module II) that runs independently of the system processor and provides the intelligence for remote management, system monitoring, alert notification, error logging, and environmental monitoring.

PFA features on the System x 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

Light path diagnostics allows systems engineers and administrators to easily and quickly diagnose hardware problems on the IBM 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. A pop-out light path diagnostics panel has an LED next to the notation for the failed subsystem. This LED directs the engineer or administrator directly to the failed component, which also shows an illuminated LED near it.

For example, memory DIMM failures can be difficult to locate. Under normal conditions, after the error is diagnosed as a memory chip failure, each individual DIMM needs to be isolated and tested or a system error log need to be examined. This process can take a significant amount of time, especially when the server loses power.

With light path diagnostics, diagnosing the memory error is simple, as the defective DIMM is immediately identified by the lit LED next to it. The memory can then be removed and replaced and the server restarted quickly. In addition, the light path diagnostics panel indicates the failing component even if the server loses power.

The following components are covered by light path diagnostics:

- Disk drives
- Memory
- Power supplies
- Processors
- Voltage Regulator Modules
- PCI Buses
- Service processors
- Temperature alerts
- Fans
- Non-masked interrupts

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.

Figure 9 shows the light path diagnostics panel of the x3850 X5 server.

Figure 9 Light path diagnostics panel of the x3850 X5 server

IBM Dynamic System Analysis

IBM Dynamic System Analysis (DSA) collects and analyzes system information to aid in diagnosing system problems. With DSA, troubleshooting and repair time can be minimized, thus improving overall system availability.

DSA collects information about the following aspects of a system:

- System configuration
- Installed applications and test fixes
- Device drivers and system services
- Network interfaces and settings
- Performance data and running process details
- Hardware inventory, including PCI information
- Vital product data and firmware information
- SCSI device sense data
- ServeRAID configuration
- Application, system, security, ServeRAID, and service processor system event logs

Additionally, DSA creates a merged log that allows users to easily identify cause-and-effect relationships from different log sources in the system.

Availability evaluation

In this section, we describe what you can expect from the advanced RAS features of the IBM eX5 portfolio in terms of server infrastructure availability and how it compares to the traditional scale out building block approach.

As we described earlier, the main RAS goal is to minimize system downtime to ensure that the business processes are not interrupted. This situation can be expressed in terms of RA, IRA, UIRA, and HIO (see “RAS: Reliability, availability, and serviceability” on page 3). For our purposes, we estimate the relative number of different types of repair actions that are expected to be performed for both scale up and scale out server farms based on picked up sample failure rates of server components. We assume and pick up sample failure rates close to the estimated averages of component failure rates found in various field studies[3, 4, 5].

Assume that our large-scale infrastructure is composed of 80,000 processing cores and 640 TB of memory, and that the operating system is installed on local hard disk drives configured in a RAID 1 array. We use an 8-way x3850 X5 serve as a scale up building block and a 2-way x3650 M4 serve as a scale out building block.

The intent of the comparison is to provide a qualitative estimation of the availability we can expect with two different approaches by using our generic example.

[3] *A Large-Scale Study of Failures in High-Performance Computing Systems*, found at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4775906&isnumber=5624749

[4] *Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?*, found at: http://static.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html

[5] *DRAM Errors in the Wild: A Large-Scale Field Study*, found at: http://research.google.com/pubs/pub35162.html

Table 4 provides characteristics of the scale out and scale up scenarios.

Table 4 Scale up and scale out building block characteristics

Characteristic	Scale out building block	eX5 scale up building block
Intel Xeon processor product family	E5-2600	E7-8800
Number of servers	5,000	1,000
Number of processors	10,000 (2 per server)	8,000 (8 per server)
Number of processor cores	80,000 (16 per server)	80,000 (80 per server)
Number of DIMMs used	80,000 (16 per server)	80,000 (80 per server)
DIMM capacity	8 GB	8 GB
Memory capacity	640 TB (128 GB per server)	640 TB (640 GB per server)
Number of system boards	5,000 (1 per server)	2,000 (2 per server)
Number of HDDs	10,000 (2 per server)	2,000 (2 per server)
Number of system fans	20,000 (4 per server)	8,000 (8 per server)
Number of power supplies	10,000 (2 per server)	4,000 (4 per server)
Key RAS features		
IBM Chipkill	Yes	Yes
Redundant Bit Steering	No	Yes
Processor-to-I/O hub link redundancy	No	Yes
Hot swap HDDs, power supplies, and fans	Yes	Yes
Automatic Server Restart	Yes	Yes
Predictive Failure Analysis	Yes	Yes

Hard disk drives, system fans, and power supplies are hot-swap and can be replaced concurrently (or without shutting down the system). Replacement of other components require a system shutdown.

The sample model we use to evaluate the availability of scalable solution is simple and based on the following characteristics:

- We use relative average failure rates for processors, memory DIMMs, hard disk drives, system fans, power supplies, and system boards to estimate the number of failures expected in our scalable infrastructure during one year of operation.
- Then, each failure is accessed as:
  - Transparent (the hardware corrected the error without system downtime, for example, RBS activated the spare DRAM because of a memory failure).
  - Concurrent repair, where the component can be replaced without shutting down the system (for example, replacing a hot-swap drive or a hot swap system fan).
  - IRA, where a planned shutdown is required to replace the failing or failed component (for example, a PFA alert is generated by the processor, or the system is restarted in degraded mode with the failed component isolated).

– UIRA, where the system isolated the failed component and automatically restarted itself (Automatic Server Restart) in degraded mode (for example, recovery from an uncorrectable memory error, where the system disables the memory module and restarts itself using the remaining amount of memory). After UIRA occurs, scheduled maintenance is required.

– HIO, where a fault is catastrophic and a service technician is required to perform a repair (for example, the failure of a system board).

Figure 10 shows the results of our estimation expressed in relative failure rates that require repair actions and types of these actions for both scale up and scale out building blocks.

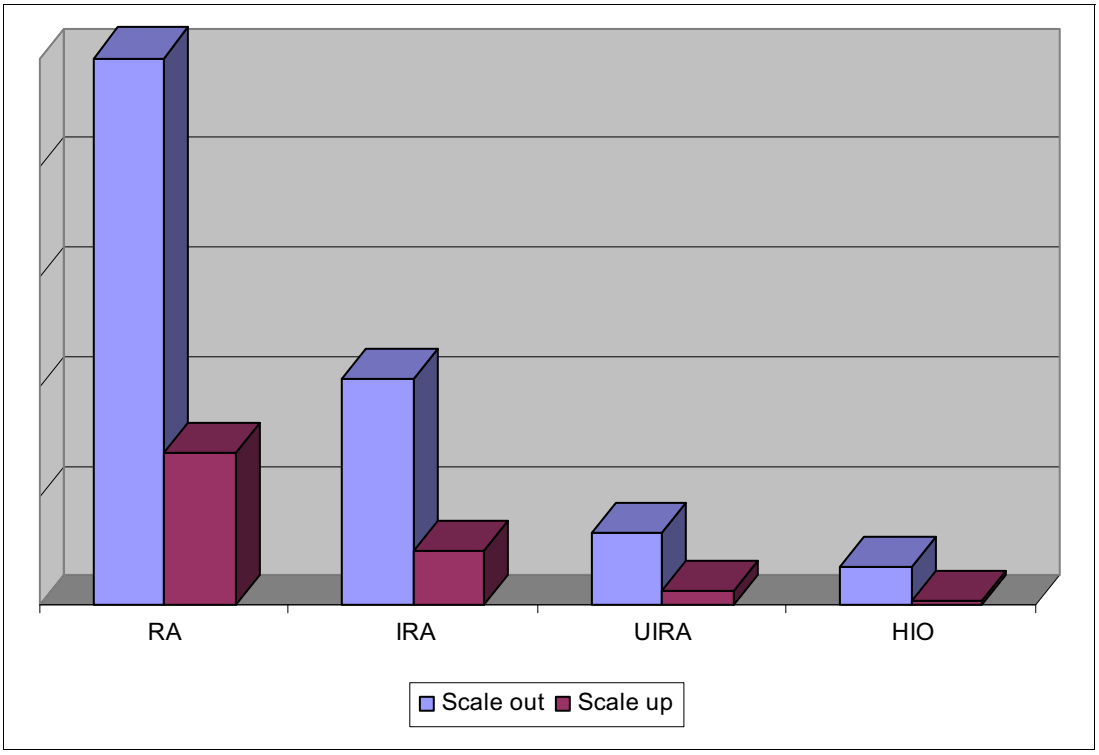

Figure 10 Relative number of repair actions for scale out and scale up building blocks

As we expected, advanced RAS features make a difference. An x3850 X5 serve requires more than three times less repair actions (RA) than a comparable scale out infrastructure. Among them, the eX5 scale up approach has four times less scheduled repair actions (IRA), performs five times less self-recovering actions by using degraded mode (UIRA), and encounters more than six times less unexpected outage that requires manual intervention to repair (HIO).

What makes that difference? Key points to highlight include:

- There are fewer components that are used to build the scale up infrastructure:
  – There are 20% fewer processors.
  – There are 2.5 times fewer system boards, fans, and power supplies.
  – There are 5 times fewer HDDs.

  The lower the number of components, the lower the number of overall failures that might happen in the infrastructure.

- More extensive memory protection techniques.

  Redundant bit steering increases the effectiveness of Chipkill.

  Memory failures are the most common cause of system downtime, and the RBS effectively doubles the number of Chipkill actions that are sustainable per server.

- Redundant processor-to-I/O hub connections.

  Ability to self-recover from processor failure.

  If the primary processor (the processor used for booting the operating system) fails, then an eX5 server can use a secondary processor to boot the OS, as the server still has access to the integrated I/O devices because of the redundant links between the processors and I/O hubs.

- Two interconnected 4-way nodes create an 8-way building block.

  Self-healing from a single-node failure.

  Two interconnected nodes form a resilient 8-way configuration. If there is a single-node failure, the system can be restarted in degraded mode, thus eliminating HIO.

In addition, the scale up infrastructure often requires less network and fabric ports as well as connections to the electrical power infrastructure, therefore simplifying cabling and requiring fewer networking infrastructure building blocks.

Conclusion

Achieving IT infrastructure availability is one of the key goals of most companies so that they can run their critical business services 24x7. At the same time, server hardware failures are one of the top causes of overall system downtime. Therefore, choosing the platform that satisfies key business, performance, scalability, efficiency, and RAS requirements becomes an important task.

RAS features can be split into two categories: standard and advanced. Standard RAS features found in IBM high volume systems provide sufficient capabilities to run most business-critical applications, but when mission-critical capabilities are required, then advanced RAS features are used.

IBM eX5 systems using the Intel Xeon processor E7 family or processors 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, IBM eX5 systems can help increase application availability and reduce downtime by enabling 24x7 mission-critical capabilities to run your core business services.

In scalable implementations, IBM eX5 systems can help achieve more than three times better overall availability (which is expressed in the number of repair actions compared to standard systems), reducing the number of system outages related to hardware failures and lowering maintenance and support costs.

Related publications

The publications listed in this section are considered suitable for a more detailed discussion of the topics covered in this paper.

- *IBM eX5 Portfolio Overview: IBM System x3850 X5, x3950 X5, x3690 X5, and BladeCenter HX5*, REDP-4650
- Dell, Timothy, *A White Paper on the Benefits of Chipkill-Correct ECC for PC Server Main Memory*, IBM Microelectronics, 1997
- *A Large-Scale Study of Failures in High-Performance Computing Systems*, found at

  http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4775906&isnumber=5624749
- *Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?*, found at:

  http://static.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html
- *DRAM Errors in the Wild: A Large-Scale Field Study*, found at:

  http://research.google.com/pubs/pub35162.html
- IBM Power Platform Reliability, Availability, and Serviceability (RAS), found at:

  http://www.ibm.com/systems/power/hardware/whitepapers/ras.html

The author who wrote this paper

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

Ilya Krutov is a Project Leader at the ITSO Center in Raleigh and has been with IBM since 1998. Before joining the ITSO, Ilya served as a Run Rate Team Leader, Portfolio Manager, Brand Manager, Technical Sales Specialist, and Certified Instructor for IBM. Ilya has expertise in IBM System x and BladeCenter products, server operating systems, and networking solutions. He has a Bachelor's degree in Computer Engineering from the Moscow Engineering and Physics Institute.

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

Kevin Barnes, Tamikia Barrow, Mary Comianos, Shari Deiana, Cheryl Gera, Linda Robinson, Wade Wallace, David Watts
International Technical Support Organization, Raleigh Center

Michelle Brunk and Randall Lundin
IBM

Brad Buland
Intel

Now you can become a published author, too!

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

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

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:

  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806
- Explore new IBM Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm
- Stay current on recent Redbooks publications with RSS Feeds:

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

Notices

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

This document REDP-4864-00 was created or updated on May 10, 2012.

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

Trademarks