

NEC Express5800/1000 Technology Guide Vol.1

Powered by the Dual-Core Intel® Itanium® Processor

NEC Express5800/1000 Series

Reliability and Performance through
the fusion of the NEC “A³” chipset and
the Dual-Core Intel® Itanium® processor



1320Xf/1160Xf



1080Rf



In today's fast-paced business environment, all enterprises, from the world's largest companies to the smallest depend on IT. Enterprise resource planning (ERP), customer relationship management (CRM), and business intelligence (BI) all require that transactions are quickly processed and that the resulting data is reliable as to meet the requirements of the rapidly changing business environment. The need for higher performance and better reliability is growing exponentially in enterprise IT platforms.

People no longer consider mainframe systems and vector supercomputers as open enterprise IT platforms. However if one were able to have supercomputer performance and mainframe reliability for the cost of an open server in a datacenter, many may reconsider.

Next generation enterprise IT platform NEC Enterprise Server Express5800/1000 series

Leveraging NEC's vector supercomputer and mainframe technology, Express5800/1000 series is designed to meet the requirement of today's mission critical enterprises.

With the new Dual-Core Intel® Itanium® processor 9000 series and the NEC designed third generation chipset "A3", from chipset, board to system-level design, NEC has never compromised to realize mainframe-class reliability and supercomputer-class performance.

Express5800/1000 series is the perfect IT platform for the most demanding mission critical enterprises.



Supercomputer-class Performance

- **High processing power by the Dual-Core Intel® Itanium® processor:**
Dual-Core, massive L3 cache and EPIC (Explicitly Parallel Instruction Computing) architecture
- **Very Large Cache (VLC) Architecture:**
High-speed / low latency Intra-Cell cache-to-cache data transfer
- **Dedicated Cache Coherency Interface (CCI):**
High-speed / low latency Inter-Cell cache-to-cache data transfer
- **Crossbar-less configuration (Available only on 1080Rf):**
Improved data transfer latency through direct attached Cell configuration

Flexibility and Operability

- **Resource virtualization through Floating IO:**
Flexible resource management allows for robust server virtualization
- **Multi-OS Support / Rich application lineup:**
Supports Windows® and Linux operating systems
- **Superior standard chassis configuration:** Small footprint and highly scalable IO



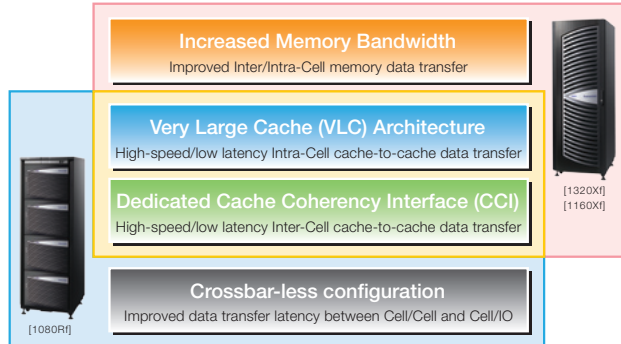
Features for performance improvement

Dual-Core Intel® Itanium® processor and high-speed inter/intra Cell cache-to-cache data transfer

At the heart of the Express5800/1000 series server is the 64-bit Dual-Core Intel® Itanium® processor, redesigned for even faster processing of larger data sets.

The system has been equipped with the NEC designed chipset, "A³", in order to improve performance by utilizing, to its full extent, the massive 24MB of cache memory that has been built into the Dual-Core Intel® Itanium® processor

Technologies to increase cache-to-cache data transfer, such as the VLC architecture and CCI, have been implemented to maximize the performance for enterprise mission critical computing.



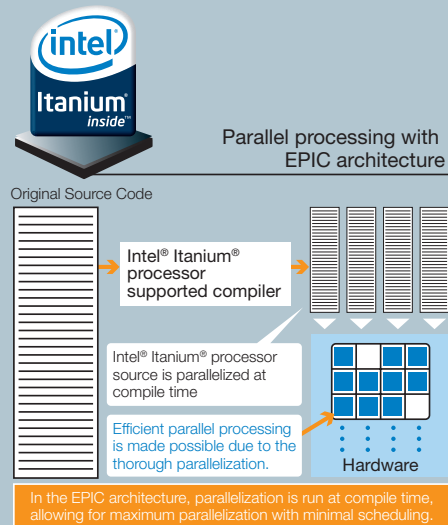
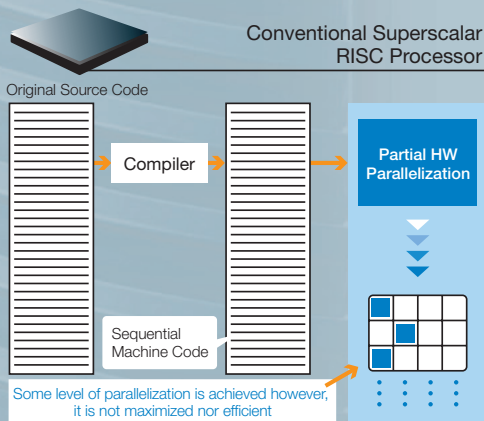
High processing power of the Dual-Core Intel® Itanium® processor

Dual-Core, massive L3 cache and EPIC (Explicitly Parallel Instruction Computing) architecture

The Dual-Core Intel® Itanium® processor is Intel's first production in the Itanium® processor family with two complete 64-bit cores on one processor and also the first member of the Intel® Itanium® processor family to include Hyper-Threading Technology, which provides four times the number of application threads provided by earlier single-core implementations.

With a maximum of 24MB of On-Die L3 cache, the Dual-Core Intel® Itanium® processor excels at high volume data transactions.

EPIC architecture provides a variety of advanced implementations of parallelism, predication, and speculation, resulting in superior Instruction-Level Parallelism (ILP) to help address the current and future requirements of high-end enterprise and technical workloads.

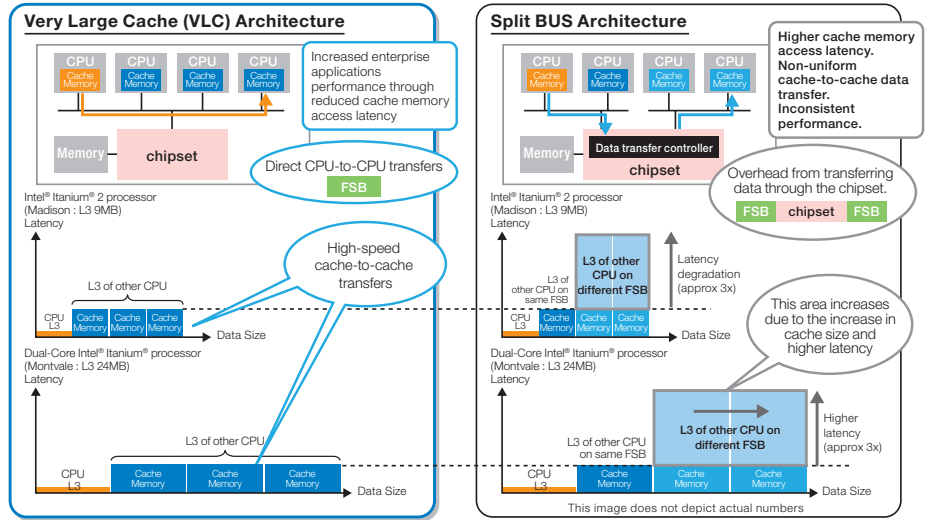


VLC Architecture

High-speed / low latency Intra-Cell cache-to-cache data transfer

The Express5800/1000 series server implements the VLC architecture, which allows for low latency cache-to-cache data transfer between multiple CPUs within a cell.

In a split BUS architecture, for a cache-to-cache data transfer to take place, the data must be passed through a chipset. However, in the VLC architecture, data within the cache memory can be accessed directly by one another, bypassing the chipset. This allows for lower latency between the cache memory, which results in faster data transfers.



Dedicated Cache Coherency Interface (CCI)

High-speed / low latency Inter-Cell cache-to-cache data transfer

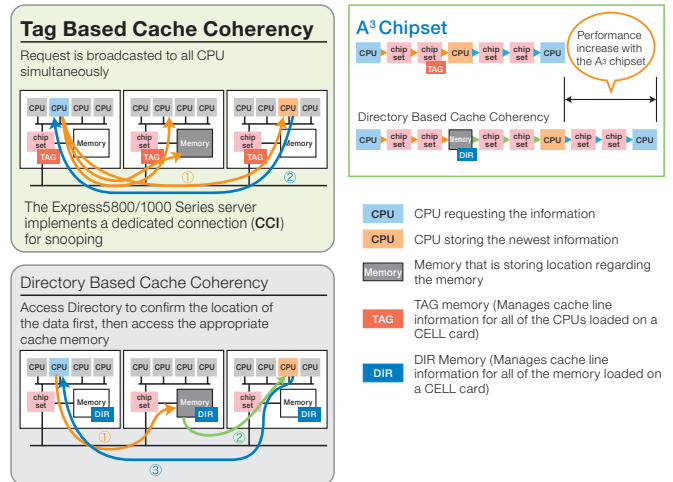
Another technology implemented in the Express5800/1000 series server to improve cache-to-cache data transfer is the Cache Coherency Interface (CCI). CCI, the inter-Cell counterpart of the VLC architecture, allows for a lower latency cache-to-cache data transfer between Cells.

Information containing the location and state of cached data is required for the CPU to access the specific data stored in cache memory. By accessing the cache memory according to this information, the CPU is able to retrieve the desired data.

Two main mechanisms exist for cache-to-cache data transfer between Cells, directory based and TAG based cache coherency. The cache information, described above, is stored in external memory (DIR memory) for the directory based, and within the chipset for the TAG based mechanisms.

In a directory based system, the requestor CPU will first access the external memory to confirm the location of the cached data, and then will access the appropriate cache memory. On the other hand, in a TAG based system, the requestor CPU broadcasts a request to all other cache simultaneously via TAG.

The benefit of the TAG based mechanism, thus implemented in the Express5800/1000 series server, is that by accessing the TAG, unnecessary inquiries to the cache memory are filtered for a smoother transfer of data. Furthermore, the Express5800/1000 series server includes a dedicated high-speed cache coherency interface (CCI) which is used to connect the Cells directly to one another without using a crossbar. This interface is used for broadcasting and other cache coherency transactions to allow for even faster cache-to-cache data transfer.



Crossbar-less configuration

Improved data transfer latency through direct attached Cell configuration

Within the Express5800/1000 series server lineup, the 1080Rf has been able to lower the data transfer latency by removing the crossbar and directly connecting Cell to Cell, and Cell to PCI box.

Even with the crossbar-less configuration, virtualization of the Cell card and I/O box has been retained as not to diminish computing and I/O resources.

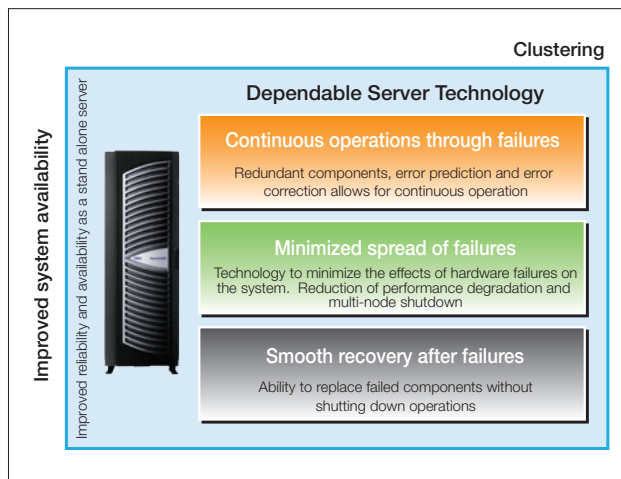
RAS Design Philosophy

Realization of a mainframe-class continuous operation through the pursuit of reliability and availability in a single server construct

Generally, in order to achieve reliability and availability on an open server, clustering would be implemented. However, clustering comes with a price tag. To keep costs at a minimum, the Express5800/1000 series servers were designed to achieve a high level of reliability and availability, but within a single server.

The Express5800/1000 series server's powerful RAS features were developed through the pursuit of dependable server technology.

Continuous operations throughout failures; minimize the spread of failures; and smooth recovery after failures were goals set forth which lead to implementation of technologies such as memory mirroring, increased redundancy of intricate components, and modularization. Through these technologies a mainframe level of continuous operation was achieved.



	Reliability	Availability	Serviceability
Mainframe Level			
Center plane	No chipset on the center plane		
Chipset	ECC protection of main data paths, intricate error detection of the high-speed interconnects	Partial chipset degradation/ Dynamic recovery	Hot Pluggable ⁴
Clock		Duplexed ¹ 16 processor domain segmentation ²	Hot Pluggable ⁴
Core I/O		Core I/O Relief	Hot Pluggable ⁴
PCI card			Hot Pluggable ⁴
Memory	ECC protection SDCC Memory	Memory Mirroring ³	
CPU L3 cache	Intel® Cache Safe Technology ³		
Power		N+1 Redundant Two independent power sources	Hot Pluggable ⁴
HDD		Software RAID Hardware RAID	Hot Pluggable ⁴
Conventional open server Level			
PC Server Level			

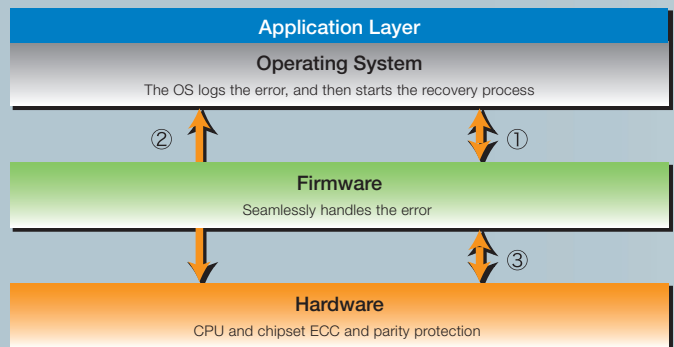
¹ Available only on the 1320X/1160X
² Available only on the 1320X
³ Intel® technology designed to avoid cache based failures
⁴ Replacement of failed component without shutting down other partitions.

The Dual-Core Intel® Itanium® processor MCA (Machine Check Architecture)

The framework for hardware, firmware and OS error handling

The Dual-Core Intel® Itanium® processor, designed for high-end enterprise servers, not only excels in performance, but is also abundant in RAS features. At the core of the processor's RAS feature set, is the error handling framework, called MCA.

MCA provides a 3 stage error handling mechanism – hardware, firmware, and operating system. In the first stage, the CPU and chipset attempt to handle errors through ECC (Error Correcting Code) and parity protection. If the error can not be handled by the hardware, it is then passed to the second stage, where the firmware attempts to resolve the issue. In the third stage, if the error can not be handled by the first two stages, the operating system runs recovery procedures based on the error report and error log that was received. In the event of a critical error, the system will automatically reset, to significantly reduce the possibility of a system failure.



- ① The Firmware and OS aid in the correction of complex platform errors to restore the system
- ② Error details are logged, and then a report flow is defined for the OS
- ③ Detects and corrects a wide range of hardware errors for main data structures

Memory Mirroring

Continuous operation even in the event of a non-correctable memory error

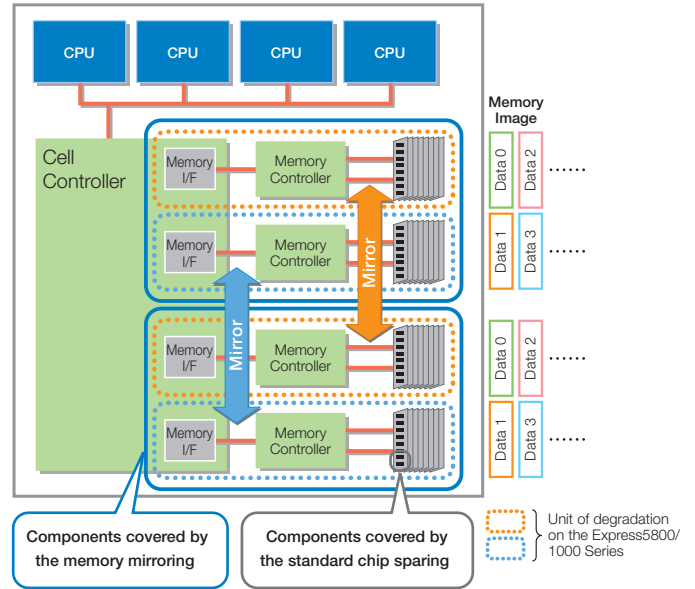
The Express5800/1000 series server supports high-level memory RAS features to ensure that the server can rapidly detect memory errors, reduce multi-bit errors and continually operate even in the event of memory chip or memory controller failures. Memory scan, memory chip sparing (SDDC*) and memory scrubbing are examples of those features.

A memory scan is run on all loaded memory modules at each OS boot. If the system detects a memory failure, the failed component is immediately isolated and detached from the system preventing possible downtime during business operations.

Chip sparing (SDDC*) memory is a memory system loaded with several DRAM chips that can correct errors at the chip level. If a failure were to occur in the memory, the error can be corrected immediately to allow for continuous operation.

Memory scrubbing checks memory content regularly (every few milliseconds) during operation without affecting performance. When an error is detected, it is corrected and then reported. The scrubbing function is effective in detecting errors in a timely manner which ultimately results in the reduction of multi-bit errors.

Memory mirroring takes place continuously, where the same data is written onto 2 separate memory blocks instead of 1 (available only on the 1160Xf and 1320Xf). In the event of a non-correctable error, due to the fact that the data exists on two independent blocks, operations are able to continue without interruption.



This construct allows for continuous operation through all non-correctable memory errors, not limited to the memory themselves, but also in the memory interfaces and the in memory controllers.

* Single Device Data Correction

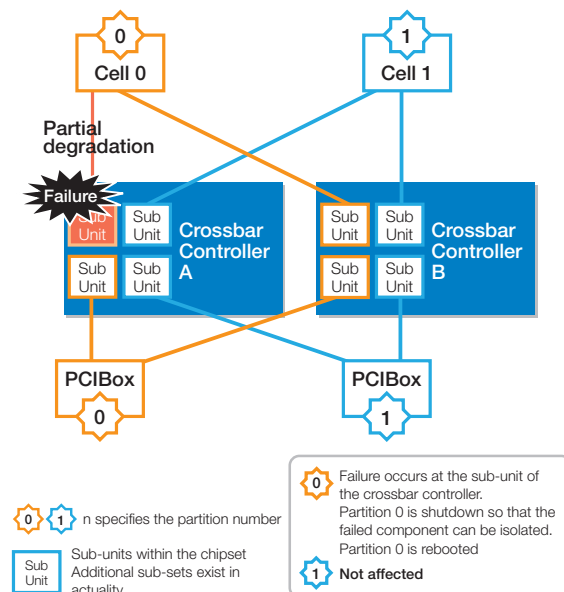
Partial Chipset degradation

Avoid multi-partition shutdowns resulting from chipset failures

In certain instances when multiple server partitions share a common crossbar controller, effects of a single partition failure may result in a multi-partition shutdown. To resolve this issue, the Express5800/1000 series servers have been designed to allow for the partial degradation of chipsets.

Within each of the LSI chips, which make up the chipset, multiple LSI sub-units exist. These sub-units are connected to other sub-units located on separate LSI chips. The combined sub-units together make up single partition. If an error were to occur on an LSI sub-unit, that sub-unit alone can be degraded to isolate the failure to a single partition, thus preventing the failure to spread to other partitions.

Furthermore, the downed partition can automatically reboot itself, after isolating the failed subsystem, to resume operations in a degraded mode without the intervention of a system administrator. This is made possible, on the Express5800/1000 series servers, by the redundant paths between the Cells and the IO.



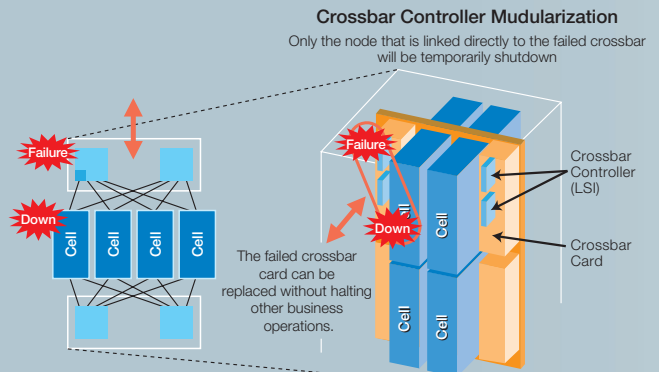
Highly Available Center Plane

System restoration after the replacement of a failed crossbar card no longer requires a planned system downtime

The Express5800/1000 series server has separated and modularized the crossbar controller which ordinarily would reside on the system center plane. By moving the crossbar controller off of the center plane, a reduction in center plane failures has been realized.

In the unlikely event of a crossbar failure, only the partition that is linked to the crossbar will be temporarily shutdown, allowing for the other partitions to continue operations uninterrupted, including during the replacement of the crossbar card.

(The 1080Rf has a crossbar-less configuration.)

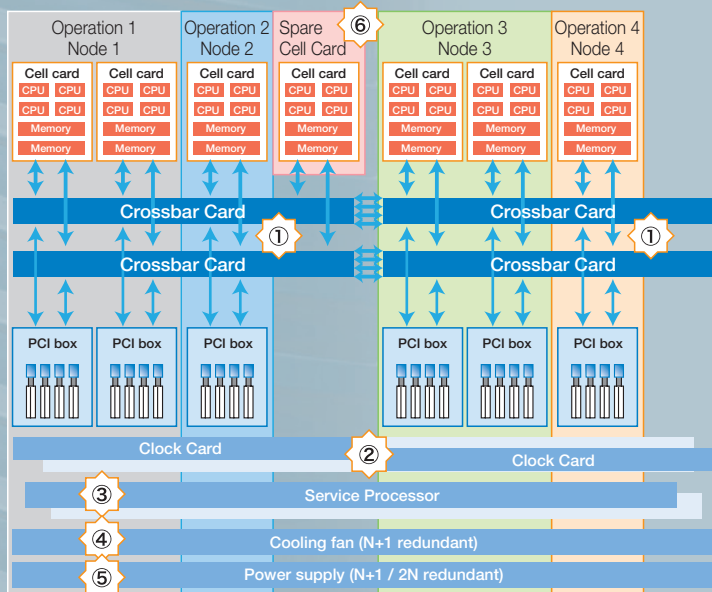
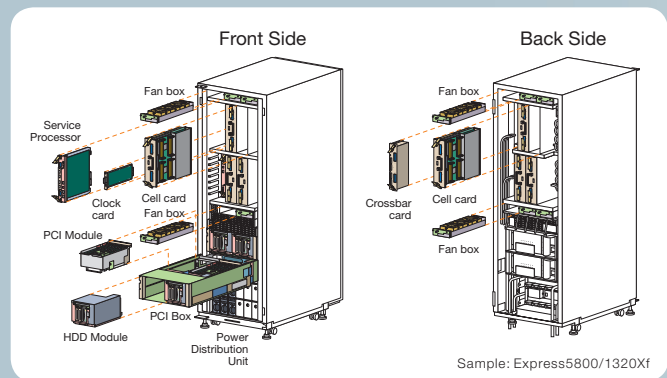


Complete modularization and redundancy

Improvements in fault resilience, continuous operation and serviceability

Major components of the Express5800/1000 series servers have been modularized, allowing for better serviceability and easy replacement in the event of a component failure.

Furthermore, to minimize the existence of single point of failure, many of these modules have redundancy, allowing for continuous operations (fault resilience).



- ① Redundant Crossbar
- ② Redundant Clock Module (Redundancy or Segmentation)
- ③ Redundant service processor
- ④ N+1 redundant cooling fan
- ⑤ N+1 redundant power supply
- ⑥ Quick recovery is possible with a spare CELL card

① 1080Rf is crossbar-less
 ② Full redundancy is available on the 1320Xf/1160Xf. Segregation is available on the 1320Xf
 ③ Available on the 1320Xf/1160Xf
 ⑤ 2N is included in the 1320Xf, and is offered as an option on the 1160Xf/1080Rf
 * This picture illustrates a 1320Xf

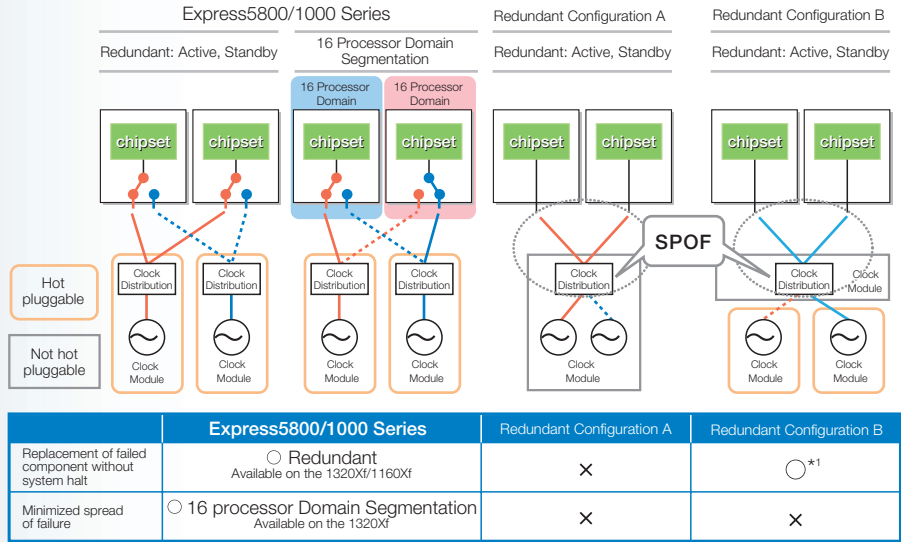
Modularization, redundancy and domain segmentation of the system clock

Minimizes downtime, and avoids multi-partition shutdown due to clock failure

Through modularization and redundancy, system downtime, due to clock failures, have been minimized. The Express5800/1000 series server has taken it one step further. In many cases, when a system is said to have a redundant clock, in actuality, only the oscillator is redundant. Integral clock distribution mechanisms such as the clock driver or the amplifier are, many times, not redundant. Such a construct leads to the existence of system single point of failures. The Express5800/1000 series servers have redundancy in not only

the oscillator, but also in the clock distribution mechanisms so that system downtime can be minimized.

The 1320Xf system allows for the division of the system into two 16 processor segments, where one segment utilizes one system clock, and the other 16 processor segment utilizes the remaining system clock. A failure in a system clock therefore, will not result in shutdown of the entire system.



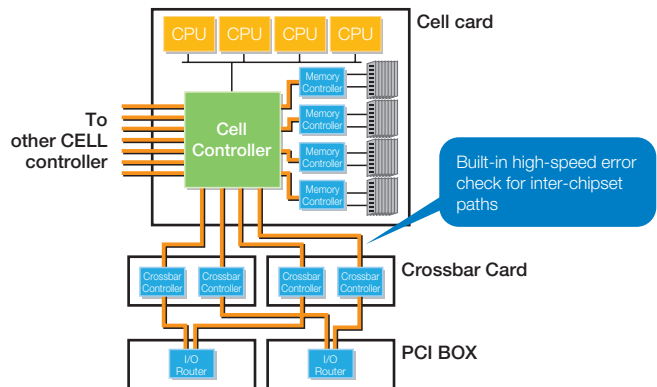
*1: Hot plugging of the redundant oscillator is possible, however the hot plugging of the single clock driver is not possible

Diagnostics of the error detection circuits

Substantial strengthening of data integrity

Main data paths of the A³ chipset on the Express5800/1000 series servers have been protected by ECC. When a single bit error is detected, a hardware error correction is carried out. Furthermore, paths between the A³ chipset interfaces support multi-bit error detection, and resending of errored data.

In addition to maintaining data integrity through these RAS features, the Express5800/1000 series server has the ability to run diagnostics on its own error detection circuits. During every system boot, all error detection circuits are diagnosed for possible failures. Without this feature, a failure in these circuits could result in the inability to detect errors during system operation.

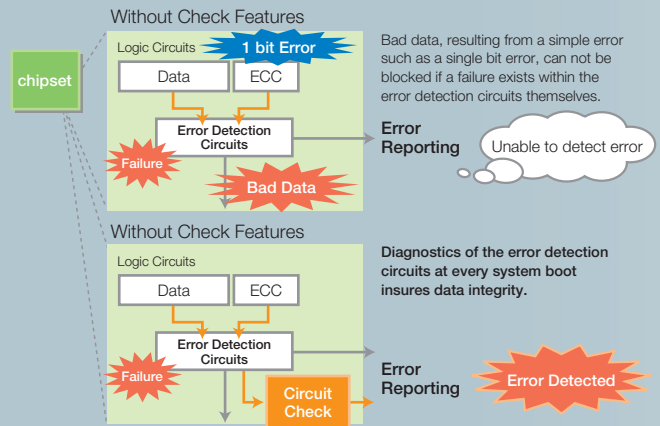


Enhanced error detection of the high-speed interconnect

Intricate error handling through multi-bit error detection and resending of errored data

Since higher speed interconnects are implemented to increase system performance, there are higher probabilities that interference noise will cause errors occurring along these interconnects. One method of handling these interconnect errors would be to disable the errored interconnect and operate in a degraded mode.

In addition to above method, the Express5800/1000 series servers have implemented a methodology prevalent in supercomputers, where by intricate multi-bit error detection is carried out, and errored data is resent upon detection of an error. This allows the Express5800/1000 series servers to handle the intermittent errors which occur along the high-speed interconnects, without impacting the system performance.



Two independent power sources

Avoid system shutdown due to failures of the power distribution units

The previous 32 processor and the 16 processor models supported having two independent power supplies, where the 8 processor model did not. This feature is now available on the new 8 processor system (1080Rf) so that the system can continue operations even in the event of a failure with in the power distribution unit.

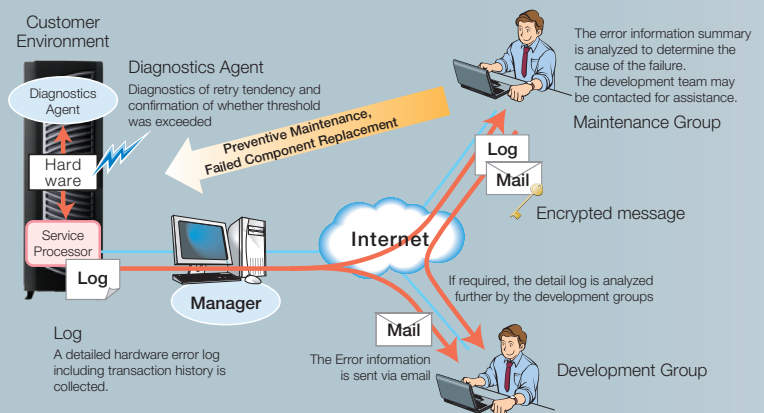
Implementation of an Uninterruptible Power Supply (UPS) can further increase availability. The two independent power source feature is a standard feature on the 1320Xf and is available as an optional feature for 1160Xf and 1080Rf.

Autonomic reporting of error logs with pinpoint prognosis of failed components

Realization of a mainframe-class platform serviceability

The Express5800/1000 series servers are equipped with a service processor which process server management and platform error handling. The service processor can be considered the core component which supports the RAS features of the system. One feature of the service processor is its ability to analyze detail logs (BID: built-in diagnosis) which are collected by the chipset in the event of an error. The BID is able to diagnose the location of the error, and will pinpoint the required FRU (Field Replaceable Unit) so that the time required to replace the component and recover the system, can be minimized.

In the event of a failure, the Express5800/1000 series servers also have the capability to automatically send detailed error logs to maintenance personnel, enabling us to further lessen the time required to resolve a system error. Furthermore, to minimize the possibility of a critical error, the diagnostics engine is able to proactively predict errors rather than just react to errors.





Pursuit of flexibility and operability in a system

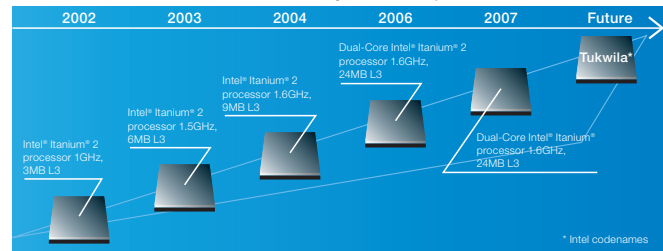
– Flexible resource virtualization using floating I/O for improved operability

Investment Protection

Smooth migration to future processors

The Express5800/1000 series servers now support the Dual-Core Intel® Itanium® processors with two complete 64-bit cores on each processor. From the beginning of development, state-of-the-art technologies have been built into the Itanium® processors to answer to the stringent levels of throughput, scalability, reliability, and availability that are required by the server platforms, and also provided top-level performance. With the deployment of the present day Dual-Core system, a smooth migration to future multi-core systems can be assured.

■ Intel® Itanium® Processor Family Roadmap

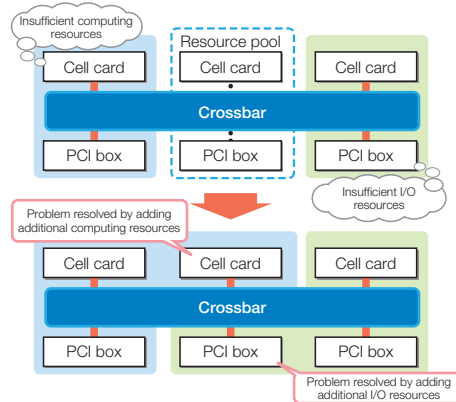


Resource virtualization through floating I/O

Flexible resource management allows for robust server virtualization

The Express5800/1000 series employ floating I/O to allow for the flexible combination of Cell cards and PCI boxes (I/O). The computational and I/O resources can be virtualized, allowing for the flexibility to reallocate system resources into the most optimal configuration according to operation or load.

Furthermore, with the existence of a spare Cell card, the system can swap the failed Cell card with the spare in the event of a failure, and reboot the system so that business operation can resume without losing valuable computational resources.



Multi OS support / Rich application lineup

Windows® operating system and Linux operating systems supported

Along with the industry's prevalent Microsoft® Windows® operating system, the Express5800/1000 series servers also support the Linux operating system. By dividing the system into multiple partitions, it is possible to support multiple operating systems within a single server.

With the inception of the Itanium® Solutions Alliance (ISA), whose main objective is to promote the advancement of Itanium®-based solutions, applications streamlined to perform on the Itanium®-based servers, such as the Express5800/1000 series servers, have increased considerably.

Superior standard chassis configuration



Small footprint and a highly scalable I/O

With the ability to load 32 Dual-Core Intel® Itanium® processors (1320Xf) into an industry standard 19-inch rack footprint, the Express5800/1000 series server has proved to have the industries highest level of performance per unit area. Because additional space is not required in the datacenter in order to accommodate

the Express5800/1000 series server, it is an ideal candidate for replacement or consolidation of older systems.

The 1080Rf is a very compact 8U model which can support up to 8 internal 3.5 inch HDD and 16 PCI cards.

■ NEC Express5800/1000 series Specifications

Model										
	1080Rf			1160Xf			1320Xf			
CPU	Processor	Dual Core Intel® Itanium® processor								
	Intel® Processor Number	9120N	9140N	9150N	9120N	9140N	9150N	9120N	9140N	9150N
	Clock frequency	1.42GHz	1.60GHz	1.60GHz	1.42GHz	1.60GHz	1.60GHz	1.42GHz	1.60GHz	1.60GHz
Maximum Number of CPU(core)		8 (16)			16 (32)			32 (64)		
On-chip cache	L1 Cache/core	16KB (I) / 16KB (D)								
	L2 Cache/core	1MB (I) / 256KB (D)								
	L3 Cache/core	6MB	9MB	12MB	6MB	9MB	12MB	6MB	9MB	12MB
	L3 Cache/CPU	12MB	18MB	24MB	12MB	18MB	24MB	12MB	18MB	24MB
Maximum Memory Capacity		128GB			512GB			1TB		
Maximum Number of I/O slots		16			32			32/64		
Internal Disk Drives	Disk Bay	8			16			16/32		
	Maximum Capacity	2,400GB (300GB * 8)			4,800GB (300GB * 16)			9,600GB (300GB * 32)		
LAN Interface		10/100Base-T (For Management console)								
Cabinet Type		Rack mount (8U)			Standalone (37U)					
Dimension (W * D * H)		441 x 857 x 351 mm			600 x 1070 x 1800 mm					
Weight		110kg			464kg			563.4kg		
Power Supply		AC 200-240V / 50Hz-60Hz								
ES Temperature/Humidity		5 - 35 degree C / 20 - 80 % RH (operation), 5 - 45 degree C / 8 - 80 % RH (non-operation) without condensation								
Supported OS		Microsoft® Windows Server® 2008 for Itanium-based Systems Microsoft® Windows Server® 2003 Enterprise Edition / Datacenter Edition Red Hat Enterprise Linux								

* NEC is a registered trademark and Empowered by Innovation a trademark of NEC Corporation and/or one or more of its subsidiaries. All are used under license. * Intel, Intel logo, Itanium and Itanium inside are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. * Microsoft and Windows are registered trademarks or trademarks of the US Microsoft Corporation in the United States and other countries. * Red Hat and Shadow Man logos are registered trademarks or trademarks of Red Hat Inc. in the United States and other countries. * Linux is a trademark or registered trademark of Linus Torvalds in the United States and other countries. * All other trademarks and registered trademarks are the property of their respective owners.

Safety notes

Please read carefully before use and observe the cautions and prohibitions in the instruction, installation, planning, operations and other manuals. Incorrect usage may cause fire, electric shock, or injury.

Company names and product names used in this catalogue are trademarks or registered trademarks of the respective companies.

If this product (including the software) comes under the regulations of Foreign Exchange and Foreign Trade Law as a regulated article or other item, observe the procedures (such as application for export permission) required by the Japanese government when taking the product out of Japan.

The colors of the products in this catalogue may be slightly different from the actual colors. Specifications are subject to change without prior notice for the purpose of improving the product.

© 2008 NEC Corporation. All rights reserved.

Information in this document is subject to change without notice.