

Focus On... Dynamic Hardware Partitioning - NEC, Microsoft, Intel

Abstract

The collaboration between NEC, Microsoft and Intel is bearing fruit, offering high-end capabilities often associated with the mainframe environment, but in the world of Windows Servers on Intel Dual-Core Intel® Itanium® 2 processors. The partnering of these technologies is delivering a high availability capability, called Dynamic Hardware Partitioning, through the NEC Express5800/1000 Series, the Intel Itanium platform with its Machine Check Architecture (MCA), and Microsoft Windows Server 2008 (formerly codename Longhorn). The combination provides mainframe-class reliability, availability and serviceability (RAS) by allowing processors, memory, and I/O to be dynamically allocated to a partitioned operating environment (hot add) during peak processing times, without having to reboot the operating system. Going one step further, this Dynamic Hardware Partitioning allows faulting processors or memory to be replaced, with no interruption to the critical running OS or applications (hot replace).

Business Drivers

Tightening financial concerns and increasing 24x7 business demands within corporations are driving more and more companies to look for cost-effective, yet highly available alternatives for running their critical business applications. Companies cannot afford to have their data center servers down for hours or even minutes at a time. This is forcing stricter RAS requirements on their infrastructure. IT is being pushed to deliver mainframe-class RAS, but without a mainframe budget.

This need is driving users to evaluate options such as Intel Itanium-based servers running Microsoft Windows Server, as the building blocks for their business-critical applications in their data centers.

Delivering Through Partnership

The partnering between NEC, Intel, and Microsoft brings together the experiences and strengths of each to this solution. NEC brings a long history developing some of the most powerful mainframes in the world. Intel offers its high-end Dual-Core Intel® Itanium® 2 processors with the ability to automatically handle 99.99% of memory and processor errors. Microsoft is using its OS experience to deliver the latest and most sophisticated, full-featured release of Windows Server. Together, these vendors' solutions offer the promise of mainframe-class RAS without the high costs.

Solution: Dynamic Partitioning

The NEC Express5800/1000 Series includes the capability for hardware partitioning of the processors, memory and I/O within the server, into physically isolated partitions, each running its own Operating System environment. The **Dynamic Hardware Partitioning** offered by the combination of these three vendors' technologies, at its highest level, allows the automation of a hot add or hot replace of processors and their associated memory, based on thresholds and policies, to a running OS partitioned on the server, with no disruption to applications on the critical OS.

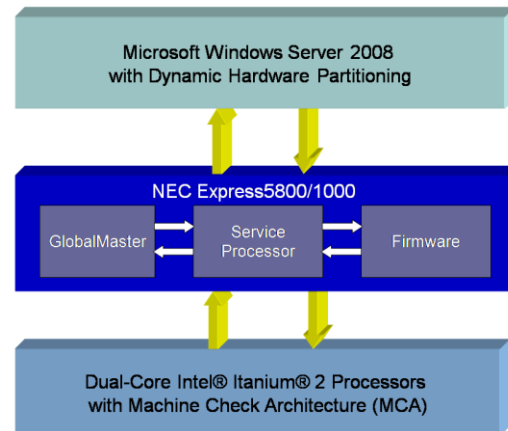


Figure 1: High Level Architectural Diagram

Implementing this type of Dynamic Hardware Partitioning requires the hardware, firmware, and operating system to work together. The operating system must understand that its processor and memory space has changed, before it can use the additional resources. Additionally, if processors and memory are being replaced with different components, perhaps because errors are exceeding an acceptable limit, the application and operating system must be quiesced, then the entire state must be moved to the new processors and memory, within milliseconds, without disrupting the operation.

The NEC Express5800/1000 Series of servers, utilizes the Intel Itanium Machine Check Architecture (for error handling), and adds another layer of intelligence through a service processor and related software called the GlobalMaster (managed via either a GUI or CLI), as shown in Figure 1. The GlobalMaster also communicates with Microsoft Windows Server 2008 to convey to the operating system that a hot add (of processors, memory, or I/O) or hot replace (of processors and memory) is taking place.

Focus On... Dynamic Hardware Partitioning - NEC, Microsoft, Intel

The NEC Express5800/1000 Series system is broken down into processing and memory “cells,” each with four processor sockets and the associated memory. Each of the cells is connected to one another and the I/O host bridges through a crossbar. These cells can be partitioned at the hardware level to create multiple physically isolated servers, depending on what each environment needs. As shown in Figure 2, the NEC Express5800/1320, which supports 32 sockets or up to 8 cells, could be partitioned with four cells allocated to one server performing business critical applications, another two cells running customer support applications and the final two cells allocated to development activities. In this example, there would be three separate servers running in this one system.

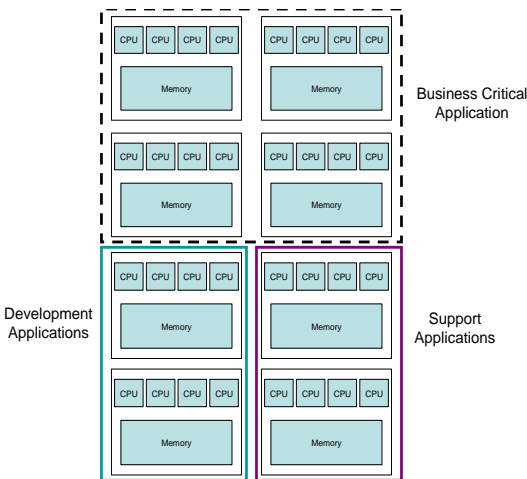


Figure 2: Hardware Partitioned Cells

The GlobalMaster has the ability to set policies, for example, to automatically reconfigure/repartition the hardware based on thresholds. Policies could include:

- Workload-based repartitioning
- Fault-based repartitioning
- Time-based repartitioning

For workload-based or fault-based repartitioning, a policy threshold is hit which causes the automatic repartitioning to be initiated. In the example above, this might occur if the processors on the business critical server have hit 90% utilization or if one of the cells in the business critical server were to hit a threshold of either memory or processor errors (reported to the service processor through the Itanium MCA – Machine Check Architecture). These would cause the GlobalMaster to notify Windows Server 2008 that a hot add (workload) or hot replace (fault) needs to occur. In the case of a hot replace, the service processor would replace the failing cell with a

cell from a less critical server, for example, the development server. Windows Server’s Dynamic Hardware Partitioning would then “move” the critical applications environment over to the “new” cell, with no disruption to the running applications. (The development server, since it is losing resources, would need to be restarted on its reduced resources.)

Time-based repartitioning, while not quite as dynamic, offers additional flexibility. It is often used when processing workloads change, for example, daytime versus nighttime processing, or regular daily operation versus month-end. In time-based processing, generally all servers are shut down, the cells are reconfigured, and the systems are restarted.

Key Benefits

Dynamic Hardware Partitioning allows partitioning of multiple operating environments on a single physical server, while minimizing critical business application downtime caused by hardware faults. Automated fault handling combined with dynamic reconfiguration of resources offers increased availability with minimal application disruption - Mainframe-class RAS without the high cost.

Focus Assessment

The Dynamic Partitioning capability from these vendors’ collaboration offers a unique and innovative way to leverage leading technologies to increase the RAS of Windows Server resources in the data center. Their joint efforts are bringing a cost-effective solution to an age-old problem.

Analysts

Anne Skamarock, advisory analyst with Focus, has spent nearly 30 years in software engineering and technical marketing, as an end-user, vendor, analyst, and author, with SRI, Sun, Solbourne, StorageTek, and Enterprise Management Associates (EMA), and has been a regular columnist for Network World. Anne is co-author of *Blade Servers and Virtualization: Transforming Enterprise Computing While Cutting Costs*, along with Barb Goldworm.

Barb Goldworm, president and chief analyst of Focus, has spent 30 years in technical, marketing, sales, senior management, and industry analyst positions with IBM, Novell, StorageTek, EMA, and multiple startups. Barb chaired the *Server Blade Summit*, and the Storage track of Interop, and has been a regular expert columnist for TechTarget, ComputerWorld Storage Networking World Online and NetworkWorld.