



# HP NonStop Multi-core Architecture

White paper

## Table of contents

Abstract.....	2
Business challenges.....	3
HP Integrity NonStop NB50000c BladeSystem .....	3
HP Integrity server blade .....	3
HP c-Class BladeSystem enclosure .....	3
HP BladeSystem manageability software .....	4
The HP NonStop Multi-core Architecture (NSMA).....	4
NSMA Process Scheduler.....	4
Why a process scheduler?.....	4
NSMA Process Scheduler affinity classes and sampling .....	5
Scheduling DP2 process groups and interrupt handler processes .....	6
NSMA interoperability and backward compatibility.....	6
Performance measurement software (Measure) exceptions .....	7
New Measure counters .....	7
Modified Measure counters .....	7
Summary .....	9
For more information.....	10

## Abstract

This white paper describes the HP NonStop Multi-core Architecture (NSMA), the performance oriented system software architecture for the HP Integrity NonStop BladeSystem. The NSMA integrates multi-core logical processors with Integrity NonStop hardware and software to deliver nearly twice as much processing power as Integrity NonStop NS-Series systems. The paper also describes one of the most innovative features of the NSMA, a new operating system component called the NSMA Process Scheduler that maximizes the workload on each core in a logical processor. NSMA interoperability and backward compatibility information is also included, in particular about new and modified multi-core processor counters integrated into HP Measure performance measurement software.

## Business challenges

Enterprises today face a continuous struggle to lower costs, reduce infrastructure complexities, conserve valuable floor space, and meet ever-changing business needs. Add spiraling energy costs—for power consumption demands and to cool the data center—and the list continues to grow. Businesses want a flexible, high-performance server platform that can provide a single solution to all these challenges.

Nevertheless, for businesses like stock exchanges, banks, telecommunications providers, and healthcare enterprises, these business technology efficiencies are only part of the solution. Enterprises like these—that handle massive transaction workloads around the clock—must protect valuable services with technologies that ensure the highest availability and reliability.

The new HP Integrity NonStop BladeSystem personifies HP's Blades Everything strategy by using the standard-based HP blades infrastructure together with NonStop value-added technology to deliver a highly available, fault-tolerant, linearly scalable system that also dramatically increases performance, reduces costs and complexities, saves floor space, and uses energy efficiently.

## HP Integrity NonStop NB50000c BladeSystem

Combining the economies of standards-based, modular computing with the trusted high availability and data integrity of NonStop technology, the Integrity NonStop NB50000c BladeSystem has half the footprint of Integrity NonStop NS-series systems yet delivers up to twice as much processing power per unit floor space within the same power envelope. At the same time, it preserves the proven, dependable, and highly available attributes of the Integrity NonStop environment, empowering customers to add new capacity to existing facilities at a lower per-transaction cost and a lower total cost of ownership.

The NonStop NB50000c utilizes the following standards-based, modular HP BladeSystem components.

- The HP Integrity server blade
- The HP c-Class BladeSystem enclosure
- HP BladeSystem manageability software

### **HP Integrity server blade**

The NonStop NB50000c leverages the Integrity server blade—running the NonStop Operating System—in an innovative way to deliver and preserve trusted Integrity NonStop technological capabilities. In the NonStop NB50000c configuration, the Integrity server blade is a full-height blade server featuring Intel® Itanium® 9100 series dual-core 1.66GHz processors supporting up to 48 GB memory (12 DIMM slots). It has been fortified with fault tolerant ServerNet server area networking interconnect technology.

### **HP c-Class BladeSystem enclosure**

The c-Class BladeSystem enclosure has been modified to provide the NonStop NB50000c with the power, cooling, and I/O infrastructure needed to support modular server, interconnect, and storage components. The enclosure is 10U high and holds up to 8 Integrity server blades. Using the c-Class BladeSystem enclosure, the NonStop NB50000c supports a minimum configuration of 2 logical processors with 8 GB of main memory per logical processor to a maximum configuration of 16 logical processors and 768 GB of main memory per system. A logical processor contains two cores.

## HP BladeSystem manageability software

The NonStop NB50000c also comes standard with HP's Systems Insight Manager, the Integrity Lights Out 2 Advanced Pack, and the Onboard Administrator—the same solutions used by other HP BladeSystem environments to provide optimal system manageability.

- HP Systems Insight Manager (SIM) helps IT organizations save time with simple and reliable hardware infrastructure provisioning monitoring and control.
- Integrity Lights Out 2 (iLO 2) Advanced Pack offers unprecedented ease in advanced remote server management and includes virtual Keyboard, Video, Mouse (KVM) and graphical remote console.
- The Onboard Administrator has been designed for both local and remote administration of the HP BladeSystem c-Class, providing wizards for simple and fast setup and configuration.

System managers and operators can use traditional NonStop management tools or these new tools—whichever is most familiar and convenient.

## The HP NonStop Multi-core Architecture (NSMA)

The NB50000c BladeSystem is the first NonStop system to deliver multi-core processing, featuring Intel® Itanium® 9100 series dual-core 1.66GHz processors. To support these powerful multi-core processors, HP introduces the new NonStop Multi-core Architecture (NSMA)—a performance oriented system software architecture.

The NSMA, embodied in the new Integrity NonStop NB50000c BladeSystem, runs relational database and transaction processing software. It also supports advanced storage area network products from the HP StorageWorks product family. A comprehensive suite of services is available that enables customers to fully benefit from the Integrity NonStop NB50000c BladeSystem.

The NSMA and the NonStop Operating System leverage powerful multi-core processing to achieve a significant boost in performance. With the introduction of the NSMA also comes a new release series of the NonStop Operating System, called the J-series, which has been integrated with and optimized for use in a multi-core architecture environment. A new standards-based NonStop I/O infrastructure also improves response time and throughput. Multi-core processing capabilities allow the NonStop NB50000c to scale up, providing nearly twice as much processing power per logical processor at a lower per-transaction cost. Typical with other NonStop systems, the NonStop NB50000c scales out through built-in clustering of logical processors—up to 4,080 logical processors in the maximum number of clustered systems (8,160 cores).

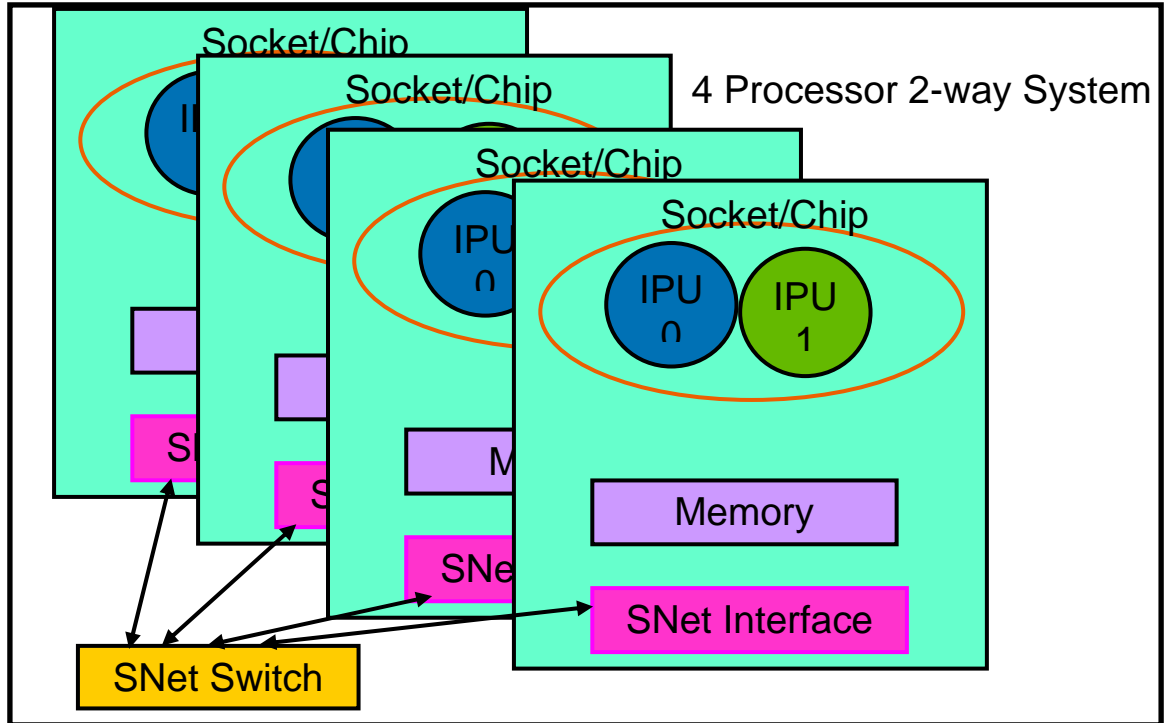
## NSMA Process Scheduler

One of the most innovative features of the NSMA is the Process Scheduler, a new operating system component responsible for maximizing the workload on each core in a logical processor.

### Why a process scheduler?

Each Integrity server blade in the NonStop NB50000c is a logical processor that consists of two or more cores. These cores are called Instruction Processing Unit (IPUs). Figure 1 shows an example of a four processor, two-way (i.e., dual core) system. NonStop NB50000c system performance approaches nearly double the performance of a similarly configured non-NSMA system because both IPUs in the Integrity server blade logical processor can execute two different user processes independently.

Figure 1. Four processor, two-way system



For the NonStop NB50000c to achieve the highest levels of performance, the IPU workload in each logical processor must be managed for optimal performance. To achieve this, HP developed the NSMA Process Scheduler. The NSMA Process Scheduler represents a departure from the traditional NonStop Operating System, which is based on a shared nothing, uni-processor architecture. With the NSMA, the IPUs share not only memory but the operating system image as well. They also run multiple processes within a single logical processor.

#### **NSMA Process Scheduler affinity classes and sampling**

The NSMA Process Scheduler uses affinity classes to decide when and how to manage process load balancing. By taking periodic IPU samples, NSMA Process Scheduler determines IPU and process busy times and can make intelligent decisions.

##### *Affinity classes*

Process schedulers used in multi-core operating systems load balance certain types of processes by moving them across IPUs. However, moving a process from one IPU to another is always an expensive operation; with it, huge amounts of process state information must be moved and caches must be reestablished. For efficient load balancing management, processes are therefore classified as having an "affinity"; that is, a classification for whether or not they can be moved to another IPU. Some processes must, if at all possible, run on a particular IPU and should not be moved. These types of processes are classified as having Hard Affinity with the IPU. A good example of a Hard Affinity class is network interrupt handling process groups, which are often limited to a specific set of processors because they carry with them so much state information.

The NSMA Process Scheduler classifies process types into four affinity classes, each defined by its own set of rules.

- Dynamic Affinity—The system dynamically decides which IPU executes a Dynamic Affinity class process.
- Hard Affinity—A Hard Affinity Class process can only run on a particular IPU.
- Soft Affinity—A Soft Affinity class process can be moved to run on another IPU.
- Group Affinity—An entire process group that can be moved to another IPU.

#### *Process Scheduler sampling*

To determine when to move a process or a process group, the Process Scheduler periodically samples the IPU—roughly once each second—to determine both IPU and process busy times. If Process Scheduler sampling determines that the IPU load isn't balanced, it takes further measures.

Process Scheduler sampling itself uses IPU processing power and can lead to performance degradation if not handled carefully. If the Process Scheduler takes samples too frequently, valuable processing time is used for sampling rather than for user processes. Conversely, if sampling is performed too infrequently, IPU utilization suffers. To keep performance optimal and to avoid either over-sampling or IPU underutilization, Process Scheduler sampling times have been rigorously tuned by HP.

#### **Scheduling DP2 process groups and interrupt handler processes**

The NSMA Process Scheduler restricts dynamic load balancing to DP2 process groups and interrupt handler processes—the two process types that consume the most processing power and are therefore critical to performance optimization.

#### *DP2 process group scheduling*

DP2 process groups represent a large load on the operating system—often using anywhere between one-third to three-quarters of computing power. Because of this, the Process Scheduler classifies them with Group Affinity, i.e., an entire process group that can be moved to another IPU. If the Process Scheduler determines with its sampling that an IPU is underutilized and a DP2 process group is using too many cycles, then all the processes in the DP2 process group will be migrated.

#### *Interrupt handler process scheduling*

The Process Scheduler classifies interrupt handler processes with Dynamic Affinity, i.e., the system dynamically decides which IPU will execute them. All low-level interrupts initially arrive at a specific IPU. This IPU is defined as the monarch IPU—that is, the IPU first detected by the hardware at boot and the one that executes the boot code. Once a low-level interrupt arrives at the monarch IPU, the Process Scheduler quickly determines the interrupt type and then chooses which IPU to execute the interrupt handler process. In this way, the monarch IPU is not heavily taxed by interrupt handling and can efficiently continue executing user processes. Meanwhile, interrupt processing is efficiently distributed among available IPUs.

## NSMA interoperability and backward compatibility

A move to the Integrity NonStop BladeSystem is compelling for enterprise customers with mission-critical applications—multi-core processing provides high performance for power-hungry applications and NonStop technology reliably delivers 24/7 availability. The standards-based modular components used in the Integrity NonStop BladeSystem let businesses pack more computing power in the same data center space and streamline operations while making future growth and reconfiguration easier.

Integrity NonStop applications and binaries are easily migrated, without modification, to the new Integrity NonStop BladeSystem—just move applications to the NSMA system environment and they

are ready to run. Integrating a NSMA based system to a ServerNet cluster is equally simple—add it just as if it were an Integrity NonStop NS-series system.

Migrating NonStop S-series applications to the NSMA environment is identical to migration from NonStop S-series to Integrity NS-series systems. Simply use an HP cross compiler on the S-series code base to create application binaries optimized for the Integrity NonStop platform. See the **HP Integrity NonStop Server Evolution program website** [<http://www.hp.com/products1/evolution/nonstop/>] for more information.

## Performance measurement software (Measure) exceptions

The NSMA is designed to simplify system administration by insulating operators and users from nearly every change resulting from the integration of multi-core processing and multiple IPU. Logical processors look the same—they simply run faster. One exception, however, is the necessary addition and modification of some Measure performance measurement counters.

### New Measure counters

To integrate Measure software with the new multi-core architecture, IPU-related fields have been added to Measure's CPU and PROCESS entities. The IPU field, part of the CPU entity, reports the number of IPU's included in a logical processor. On a non-NSMA system, this field will always be returned as 1.

Three other IPU-related CPU counters have also been added to Measure:

- IPU-BUSY-TIME: The amount of time that an IPU was busy.
- IPU-QTIME: The amount of time processes have spent on an IPU's ready list.
- IPU-DISPATCHES: The number of process dispatches for an IPU.

Another counter, IPU-SWITCHES, has been added to Measure's PROCESS entity. This field indicates the total number of times that a process has switched from one IPU to another while executing. On a non-NSMA system, this field will always be returned as 0.

### Modified Measure counters

Two CPU entity fields, CPU-BUSY-TIME and PROCESH-SAMPLES, have also changed. In the NSMA environment, CPU-BUSY-TIME is the aggregate of the BUSY-TIMEs of the individual IPU's that comprise the logical processor. With the 'RATE ON' option, the maximum value of this field is always 100 percent. However, with 'RATE OFF' option, the field represents the raw aggregate value of the BUSY-TIMEs of all IPU's in the logical processor.

For example, using a measurement duration of 5 seconds, if IPU 0 of CPU 0 is 4 seconds busy and IPU 1 is 3 seconds busy, then the CPU-BUSY-TIME for that CPU is actually 7 seconds. Seven seconds is then the value of this field and also what is displayed with a 'RATE OFF' option. However, with a 'RATE ON' option, the measurement is normalized and reported as 70 percent busy rather than 140 percent busy.

PROCESH-SAMPLES is the aggregate value of all the number of samples on all the IPU's in the logical processor and hence is typically equal to  $n \times \text{sampling frequency on that CPU}$ , where  $n$  is the number of IPU's.

Figure 2. LIST CPU with RATE OFF option

```

Cpu 1 NSE-L          Init Lock Pgs      59          Mem Pages 1048576
Memory MB 16384      PCBs          8086        Pg Size   16384 Bytes
IPUs 2
Format Version: H03 Data Version: H03 Subsystem Version: 2
Local System \BLITUG From 30 Jan 2008, 17:28:18 For 27.7 Seconds
----- Processor -----
Cpu-Busy-Time          709.65 ms   Dispatches          146,422 #
Cpu-Qtime              772.98 ms   Intr-Busy-Time      94.10 ms
Comp-Traps              0.45 ms     Process-Ovhd
Native-Busy-Time      709.65 ms   Accel-Busy-Time
TNS-Busy-Time          34,922 #    PROCESSH-Samples
----- Memory -----
Starting-Free-Mem      1,006,033 #   Ending-Free-Mem    1,006,090 #
...
----- IPUs -----
IPU 0
Ipu-Busy-Time          667.04 ms   Ipu-Dispatches     126,318 #
Ipu-Qtime              726.90 ms
IPU 1
Ipu-Busy-Time          42.61 ms   Ipu-Dispatches     20,104 #
Ipu-Qtime              46.09 ms

```

## Summary

HP NonStop technology specialists have devoted countless development hours to ensure that the NSMA delivers all the necessary underpinnings for the new high performance Integrity NonStop BladeSystem. The design of the NSMA performance oriented system software environment preserves proven, dependable, and highly available Integrity NonStop capabilities while leveraging the power of multi-core processing and the economies of standards-based, modular computing.

The NSMA is designed to simplify system administration by insulating operators and users from nearly every change resulting from the integration of multi-core processing with NonStop technology. Logical processors look the same; they simply run faster. A new operating system component, called the NSMA Process Scheduler, maximizes the workload on each core in a logical processor to further support high performance computing.

With the introduction of the NSMA also comes a new release series of the NonStop Operating System, called the J-series, which has been integrated with and optimized for use in a multi-core architecture environment. A new standards-based NonStop I/O infrastructure has been developed to further improve response time and throughput.

The NSMA makes it possible for Integrity NonStop applications and binaries to be easily migrated without modification—just move applications to the new system and they are ready to run. Integrating a NSMA based system to a ServerNet cluster is equally simple—you can add it just as if it were an Integrity NonStop NS-series system.

A move to the Integrity NonStop BladeSystem is compelling for enterprise customers with mission-critical applications: multi-core processing provides high performance for power-hungry applications and NonStop technology reliably delivers 24/7 availability. The standards-based modular components used in the Integrity NonStop BladeSystem let businesses pack more computing power in the same data center space and streamline operations while making future growth and reconfiguration easier.

To find out more about the Integrity NonStop BladeSystem, refer to the many literature options available at [www.hp.com/go/nonstopblade](http://www.hp.com/go/nonstopblade) or contact your HP representative.

## For more information

[www.hp.com/go/nonstopblade](http://www.hp.com/go/nonstopblade)

© Copyright 2008 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Linux is a U.S. registered trademark of Linus Torvalds. Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. UNIX is a registered trademark of The Open Group.

4AA2-0026ENW, May 2008

