NonStop Advanced Architecture, System Configuration and Site Planning

PF-06-HP

Bob Kossler
Product Management, HP
May, 2005

© 2004 Hewlett-Packard Development Company, L.P.
The information contained herein is subject to change without notice
NonStop Advanced Architecture

A new way for the NonStop Enterprise Division to build fault tolerant servers for the world’s most critical applications.

We leverage components from other parts of HP and

We put NonStop development resources into our unique value-add components
Microprocessor Comparison Becomes a Challenge

• Replicating and comparing microprocessors has become more difficult with each new generation.
  – Non-deterministic operation.
  – Variable frequency clocks.
  – Very high speeds with multiple clock domains.
  – Higher soft (transient) error rates.
  – Multi-core designs.
NonStop Advanced Architecture

• Use modified HP Integrity Itanium-based Servers in a cluster of uniprocessors.

• Loose Synchronization.
  - Each server runs on its own clock.
  - Each can perform soft error corrections without causing a miscompare.

• Choice: Triple or Double Redundancy.
  - Customer selects the level of availability they want.

• Compare results only on output operations.
  • All I/O is through the ServerNet.
  • Voter logic connects the ServerNet to the replicated 4-way servers.
Inside a NonStop server Blade Element

Memory reintegration link in

Memory reintegration link out

Mem. Copy function for Blade Element Reintegration

Microprocessors, I/O bridge, memory controller

NonStop server value-add (not just a standard product)
## Terminology

- **Logical Processor** - a single processor in the original NSK sense

- **Processing Element (PE)** - a single microprocessor or microprocessor core capable of executing a single instruction stream

- **Blade Element** - 4 PEs each with memory

- **Logical Synchronization Unit (LSU)** - a voting unit plus a ServerNet Interface

- **Blade Complex** - 4 logical processors composed of 2 to 3 blades plus their associated LSUs
Four Logical Processors built from Triplex of Blade Elements

BE A
PE_{A0}
PE_{A1}
PE_{A2}
PE_{A3}

BE B
PE_{B0}
PE_{B1}
PE_{B2}
PE_{B3}

BE C
PE_{C0}
PE_{C1}
PE_{C2}
PE_{C3}

"Blade Element (BE)" (the Itanium-based Server)

"Processor Element" (e.g. an Itanium μP)

Triplex Logical Processor

"Blade Complex"
Blade Complex - Duplex

memory reintegration link

CPU 0
CPU 1
CPU 2
CPU 3
memory

BE A
BE B
LSU 0
LSU 1
LSU 2
LSU 3
ServerNet
ServerNet
ServerNet
ServerNet
Blade Complex – Duplex
(highlighting memory partitioning)
Blade Complex - Triplex multiple fault tolerance

memory reintegration link

CPU 0
CPU 1
CPU 2
CPU 3

memory

BE A

BE B

BE C

LSU 0
LSU 1
LSU 2
LSU 3

ServerNet
ServerNet
ServerNet
ServerNet
Three Key Mechanisms in the Architecture

- **Voting (Comparison)**
  - LSUs compare the (I/O) output streams
  - Supports simplex, duplex, and triplex models

- **Reintegration (Memory Copy)**
  - When a blade element (BE) has diverged or has been replaced (repaired), copy memory from one slice to another and resume
  - This is an on-line operation; think “disk revive” for memory

- **Rendezvous (synchronization)**
  - A distributed consensus algorithm with hardware assist
  - Periodic checking and resynchronization (can optionally remove time skew)
    - Time-of-Day
    - Interrupts/Dispatches
    - I/O Initiation
Hardware Configuration Options
Modular Fabric Topology

Cluster Switch (X Fabric) → P-Switch (X Fabric) → IOAME

Cluster Switch (Y Fabric) → P-Switch (Y Fabric) → IOAME

Maximum of 10 ServerNet Adapters per IOAME
Modularity and industry standards

- Allow independent technology update
- Modular CPU blade elements allow ease of upgrade to future generations of Intel Itanium processors
- Modular Input/Output adapters are industry-standard Fibre Channel and Gigabit Ethernet products
- Modular cabinet is an industry-standard 19-inch rack
- The Uninterrupted Power Supplies and Extended Runtime Module are standard components available from HP
The modular cabinet features

- Industry-standard 19-inch rack
  - 42U height
  - Power distribution unit (PDU)
  - Two PDUs per cabinet
  - 14 outlets per PDU
  - Two fuses per outlet
- Ceiling or floor power input
- Vertical Cable Management System
- Other separate, optional components:
  - Side panels
  - Baying kit
NonStop Blade Element

• Functional Characteristics
  – Advanced architecture Itanium processors
  – Duplex or triplex
  – 4, 8 and 16GB of main memory per CPU
  – Dual ServerNet Ports at 2.5 Gbit serial

• Self Contained module- has own power supplies and cooling

• External cables for
  – Two ServerNet fabrics
  – Maintenance LAN
  – 5U height (1U = 1.75 inches)
  – 19 inch rack mountable
  – 112 Lbs fully populated
Logical Synchronization Unit

• Logical synchronization Unit
  - 4U height (1U = 1.75 inches) with 8 slots
  - 19 inch rack mountable
  - 95 Lbs fully populated
  - Dual AC input
  - Up to 8 LSU modules
    • Voter Logic Board
    • Voter Optics Board
Modular cabinet
Power distribution unit

PDU
- 2 PDUs per cabinet
- 14 outlets per PDU
  - 2 fuses per outlet
- Ceiling or floor power input

PDUs
NED designed PDU outlet
# Modular cabinet

## PDU power options

<table>
<thead>
<tr>
<th>Geography</th>
<th>PDU power configuration</th>
<th>Product ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S., JPN</td>
<td>208 VAC Three Phase Delta, 60A RMS, 4 wires</td>
<td>M8311-01</td>
</tr>
<tr>
<td>WW</td>
<td>200 to 250 VAC Single Phase, 63A RMS, 3 wires</td>
<td>M8311-02</td>
</tr>
<tr>
<td>EMEA</td>
<td>230/400 VAC Three Phase Wye, 63A RMS, 5 wires</td>
<td>M8311-03</td>
</tr>
<tr>
<td>EMEA</td>
<td>230/400 VAC Three Phase Wye, 63A RMS, 5 wires, Harmonized</td>
<td>M8311-04</td>
</tr>
<tr>
<td>WW</td>
<td>200 to 250 VAC Single Phase, 63A RMS, 3 wires, Harmonized</td>
<td>M8311-05</td>
</tr>
</tbody>
</table>
Uninterruptible power supplies

- **Uninterruptible power supply (UPS)**
  - For NonStop customers without a site UPS
  - Provides ride-through in the event of a power failure
    - Estimated ride-through is 5–60 minutes (depending on load)
  - **Product ID**
    - M8330-6 for North America and Japan
    - M8330-6W for all other international locations

- **Extended Runtime Module (ERM)**
  - Provides additional ride-through in the event of a power failure
    - Estimated ride-through is 20–170 minutes (depending on load)
  - **Product ID** is M8330-ERM
Enclosures as building blocks

- IO Adapter
- Module
- Enclosure

- Disk Enclosure
- Disk Enclosure

- Blade Element
- Blade Element
- LSU

- Switch layer
- Switch layer
- Switch layer

- Blade Element
- Blade Element
- Blade Element
- LSU
- Cluster Switch
- Complex

- Blade Element
- Blade Element
- Blade Element
- LSU
- Cluster Switch
- Complex

- P-Switch
- P-Switch
- P-Switch
- Cluster Switch
- Complex

- 5.5kva UPS

- Modular Tape
- SCSI Router

- Console

- Maintenance LAN

- UPS

- Battery
Example 2 or 4 P DMR System

- Single footprint
  - Up to 4 Processors
  - Up to 14 Mirrored disk volumes
  - SCSI Router for connection to tape (for example, table top DAT)
  - Desktop console not shown
Example DMR 16P System
with Enterprise Storage

IOAME
P-Switch-X
P-Switch-Y
LSU-0
Element 1B
Element 1A
PDU Keep-out

DM 1 Backup
DM 1 Primary
LSU-1
Element 3B
Element 3A
Element 2B
Element 2 A
PDU Keep-out

Modular Tape
IOAME
Element 4B
Element 4A
PDU Keep-out

XP Enterprise Storage
Example TMR 16P System
optional tape and console shown

• Disk Mirroring example with cabinet FT for disk
System Expansion Issue

8P

DM 1 Backup
DM 1 Primary
IO enclosure
P-Switch-X
P-Switch-Y
LSU-0
Element 1 B
Element 1 A
PDU Keep-out
DMR 4P Base

DM 4 Backup
DM 4 Primary
DM 3 Backup
DM 3 Primary
DM 2 Backup
DM 2 Primary
Element 2 B
Element 2 A
PDU Keep-out
DMR 4P Add-on

12P

DM 4 Primary
DM 4 Backup
DM 1 Backup
DM 1 Primary
IO enclosure
P-Switch-X
P-Switch-Y
LSU-0
Element 1 B
Element 1 A
PDU Keep-out
DMR 4P Base

DM 4 Backup
DM 3 Backup
DM 3 Primary
DM 2 Backup
DM 2 Primary
Element 3 B
Element 3 A
Element 2 B
Element 2 A
PDU Keep-out
DMR 4P Add-on
IO Configuration Considerations
Single IOAME, 2 Adapter FCDM configuration

- **Minimum configuration**
  - Lowest fault redundancy
Single IOAME, 4 Adapter FCDM Configuration

- **Better configuration:**
  - Adapters are paired such that primary and mirror disks do not share IO adapter resources.
Two IOAME IO configuration

- **Best Configuration**
  - When two IOAMEs are ordered, IO adapter pairs are spread across both IOAMEs
S-series enclosures attached to Itanium-based NonStop Server

- Maximum 24 Legacy IO Enclosures
- For Every connection that would have attached to an IOAME, 4 S-series IO enclosures can be attached

New Integrity NS-Server stack

- Itanium-based CPU n
- Itanium-based CPU 0
- IOMF2
- ServerNet cable connections from original S-series enclosures
- P-Switch
Convert System Enclosures to IO Enclosures

- Any ServerNet attachment to a Cluster must be shifted to the P-Switch.
- Any Modular I/O Adapter attachment must be shifted directly to the P-Switch.
- Any Fox-Gateway connections must be eliminated. If needed, an S-Series node may be kept in a ServerNet Cluster and used to connect to K-Series nodes.
- All MSEBs may be removed.
- If desired, disks and ServerNet Adapters from partially populated system enclosures may be consolidated into fewer enclosures and the empty enclosures may be removed from the system.
  - Up to 4 ServerNet Adapters per enclosure
  - Up to 16 disks per enclosure
- For all system enclosures that are to remain in the system
  - All PMF CRUs must be exchanged for IOMF-2 with MMF PICs
  - Any S7000 enclosure must be upgrade to include a power shelf.
Rack Mounted Console

Rack Mounted Server
- Optional OSM/TSM Console
- Proliant DL320 1 U Server
- Provide console services for up to 5 Nonstop systems

Rack Mounted KVM
- Integrated drawer with Keyboard, Flat panel display, and rollerball mouse
- Consumes 1U of rack space
A Word From Our Sponsors!
# HP Integrity NonStop Server

## Education & Training Options

<table>
<thead>
<tr>
<th>Title</th>
<th>Audience</th>
<th>Availability</th>
<th>Location/Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>- HP Integrity NonStop NS-Series Ops &amp; Mgmt Differences U8526S</td>
<td>Operations Management</td>
<td>Now</td>
<td>United States Education Centers</td>
</tr>
<tr>
<td>- NonStop H-Series Operating Systems Application Migration U8616S</td>
<td>Programmer Analysts</td>
<td></td>
<td>Lecture/Lab, RAIL</td>
</tr>
<tr>
<td>- NonStop Kernel Architecture TNS/E U8609S</td>
<td>System Administrators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- NonStop H-Series Operating Systems Application Migration U8616S</td>
<td>System Managers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overviews, Presentations, Technical Update Presentations</td>
<td>Application developers</td>
<td>Now</td>
<td></td>
</tr>
<tr>
<td></td>
<td>System designers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### News from Nonstop Education & Training

- **Free Apple iPod Mini promotion**: Take any two NonStop classes and earn a Free Apple iPod.
- **Private training price reduction**: Up to 60% reduction on any 2 NonStop classes at your site.

## Your NonStop Education & Training Contacts

- **Private Classes at your site**: Francine Barr
  - francine.barr@hp.com
  - (703) 803-2931
- **Instructor-Led Live classes over Internet**: Ben Wood
  - ben.wood@hp.com
  - (408) 285-9662
- **Scheduled classes at Education Centers**: Ben Wood
  - ben.wood@hp.com
  - (408) 285-9662
- **Customer Conference Call enrollment**: Dan Porter
  - porter@hp.com
  - (412) 303-5213
- **Nonstop University Subscriptions**: Phyllis Longbons
  - phyllis.longbons@hp.com
  - (408) 285-9131

---

<table>
<thead>
<tr>
<th>Enroll</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-800-472-5277 (US) 1-800 563-5089 (Canada)</td>
</tr>
<tr>
<td><a href="http://www.hp.com/education/">http://www.hp.com/education/</a></td>
</tr>
</tbody>
</table>

---

This is an HP copyrighted work that may not be reproduced without the written permission of HP.
End of Presentation
Next month = “___”

Thank You!