IBM Power Systems p
AIX/UNIX/Linux
Virtualization Basics

Key Terminology:
Practical Implementations
of Virtualization
& Live Partition Mobility
Server Workload Considerations

- **Traditional Server Capacity**
  - Workloads fluctuate differently throughout the day and year
  - Workloads can fluctuate greatly moment to moment
  - Servers are purchased to handle individual unknown peaks
  - Unused resources can not be moved between servers
LPAR (Logical Partitions) – Separate O/S Images

AIX Workload Manager (WLM)

Multiple Applications (processes) controlled (fenced off) with a single AIX image

- Improved administrative efficiency by reducing the number of AIX images to maintain
IBM System p5

AIX Workload Partitions

Separate regions of application space within a single AIX image

- Improved administrative efficiency by reducing the number of AIX images to maintain

- **Software partitioned system capacity**
  - Each Workload Partition obtains a regulated share of system resources
  - Each Workload Partition can have unique network, filesystems and security

- **Two types of Workload Partitions**
  - System Partitions
  - Application Partitions

- **Separate administrative control**
  - Each System Workload partition is a separate administrative and security domain

- **Shared system resources**
  - Operating System, I/O, Processor, Memory

---

AIX 6.1

- Workload Partition Test
- Workload Partition Billing
- Workload Partition Web Server
- Workload Partition Application Server
- Workload Partition BI

---

meridian IT
PowerVM AIX Virtualization Continuum

- AIX Workload Partitions
- Complement Logical Partitions

LPAR / Micropartitions
AIX V5.3 on POWER5 or later

Workload Partitions
AIX 6 on POWER4 or later

AIX Workload Manager
AIXV4.3.36 on POWER3 or later

Resource Flexibility

Workload Isolation
Two WPAR AIX Offerings…

- **AIX 6**
  - Workload Partitions (WPAR) included in AIX 6
  - Element (single system) WPAR Management

- **Workload Partitions Manager™**
  - Enablement for Live Application Mobility
  - Cross System Management for Workload Partitions
  - Automated, Policy-based Application Mobility
  - Part of IBM System Director Family
Workload Partitions Manager

- Management of WPARS across multiple systems
- Lifecycle operations
- Single Console for:
  - Graphical Interface
  - Create & Remove
  - Start & Stop
  - Checkpoint & Restart
  - Monitoring & Reporting
  - Manual Relocation
  - Automated Relocation
  - Policy Driven Change
- Infrastructure Optimization
- Load Balancing
Graphical WPAR Manager & Application Mobility

Workload Partition Manager
AIX Live Application Mobility

Move a running Workload Partition from one server to another for outage avoidance and multi-system workload balancing.

Works on any hardware supported by AIX 6, including POWER5 and POWER4.
IBM System  Simultaneous multithreading

POWER5 & POWER6 (simultaneous multithreading)

- FX0
- FX1
- LSO
- LS1
- FP0
- FP1
- BRZ
- CRL

<table>
<thead>
<tr>
<th>Thread0 active</th>
<th>No thread active</th>
<th>Thread1 active</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

System throughput

- ST
- SMT

- Utilizes unused execution unit cycles
- Presents symmetric multiprocessing (SMP) programming model to software
- Natural fit with superscalar out-of-order execution core
- Dispatch two threads per processor: “It’s like doubling the number of processors.”
- Net result:
  - Better performance
  - Better processor utilization

Appears as four CPUs per chip to the operating system (AIX 5L V5.3 and Linux)
IBM partitioning innovations help solve this problem

*In 2001:* Logical Partitioning enabled consolidation of multiple application workloads

*In 2002:* Dynamic Logical Partitioning, enabled dynamic reassignment of workloads
Meeting Demand: Utilizing Available CPU Cycles

- Policy-based, automatic partition resource tuning
- Dynamically adjust CPU and memory allocation

**Before resource tuning**

Unbalanced resource allocation

<table>
<thead>
<tr>
<th>Test LPAR</th>
<th>CRM LPAR</th>
<th>Finance LPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 CPUs</td>
<td>5 CPUs</td>
<td>6 CPUs</td>
</tr>
</tbody>
</table>

**After resource tuning**

Adjust resource allocation based on business priority

<table>
<thead>
<tr>
<th>Test LPAR</th>
<th>CRM LPAR</th>
<th>Finance LPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CPU</td>
<td>3 CPUs</td>
<td>10 CPUs</td>
</tr>
</tbody>
</table>
### Micro-Partitioning technology

**AIX 5.3/AIX 6/Linux/i5 & iOS**

<table>
<thead>
<tr>
<th>Dynamic LPARs</th>
<th>Micro-partitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Processors</td>
<td>Pool of 6 CPUs</td>
</tr>
</tbody>
</table>

- **Micro-partitions: Up to 254***
- **Dynamic LPARs: Up to 32***
- Combination of both

**Configured via the HMC**

- **Number of logical processors**
  - Minimum/maximum

**Entitled capacity**

- In units of 1/100 of a CPU
- Minimum 1/10 of a CPU

**Variable weight**

- % share (priority) of surplus capacity

**Capped or uncapped partitions**

* on p5-590 and p5-595
** on p5-570, p5-590, and p5-595
More Efficiency: **CPU & I/O Sharing**

**Included features:**

- **Enablement for Micro-Partitioning**
  - Share processors across multiple partitions
  - Minimum partition 1/10th processor
  - AIX 5L V5.3 or Linux*

- **Virtual I/O server**
  - Shared Ethernet
  - Shared SCSI and Fibre Channel-attached disk subsystems
  - Supports AIX 5L V5.3 and Linux*

- **Partition Load Manager**
  - Both AIX 5L V5.2 and AIX 5L V5.3 supported
  - Balances processor and memory request

* SLES 9 or RHEL AS 3
POWER 5 & 6 / AIX 5.3 LPAR DISK Sharing

Using LVM mirroring for Availability
One physical drives appears to be multiple logical drives
- LUNs appears as individuals logical drives
Minimizes the numbers adapters
SCSI and Fibre supported
POWER5 / AIX 5.3 LPAR Virtual Ethernet

- Configured like a standard Ethernet
- IP forwarding provided by I/O Server partition
- Can have multiple connections per partition
- Inter-LPAR communications: IPv4 & IPv6 plus large MTUs
Virtual networking

- Virtual Ethernet
  - Partition to partition communication
  - Requires AIX 5L V5.3 and POWER5
- Shared Ethernet Adapter
  - Provides access to outside world
  - Uses Physical Adapter in the Virtual I/O Server
- VLAN – Virtual LAN
  - Provides ability for one adapter to be on multiple subnets
  - Provides isolation of communication to VLAN members
  - Allows a single adapter to support multiple subnets

Virtual Ethernet helps reduce hardware costs by sharing LAN adapters

Available via optional Advance POWER Virtualization or POWER Hypervisor and VIOS features.
Virtual SCSI

External Storage

Micro-partitions

POWER5 Server

Shared Fiber Channel Adapter

VIOS

AIX 5L V5.3

Linux

AIX 5L V5.3

Linux

Virtual SCSI

LVM based storage on VIO Server

Physical Storage can be SCSI or FC

Local or remote

VIOS owns physical disk resources

Micro-partition sees disks as vSCSI (Virtual SCSI) devices

Multiple LPARs can use same or different physical disk

Configure as logical volume on VIOS

Appear a hdisk on the micro-partition

Can assign entire hdisk to a single client

Available via optional Advance POWER Virtualization or POWER Hypervisor and VIOS features.
On Demand Capabilities: Keeps on Expanding

- CoD Offerings (Capacity on Demand)
  - Trial
  - Reserve
  - On/Off
  - Backup

- RAS (Reliability Availability Serviceability)
  - Processor and Memory Sparing

- Building Block Architecture
  - Pay as you grow
Capacity on Demand

**Capacity Upgrade on Demand**
- Upgrade system with processors and/or memory
- No special contracts, no required monitoring (no ability to turn off the capacity)
- Purchase agreement

**On/Off Capacity on Demand**
- Temporary use of requested number of processors or amount of memory
- Client selects the capacity and activates the resource (registered system)
- Capacity can be turned on and off by the client
- Information captured by IBM (or reported to IBM)
- Rental agreement

**Reserve Capacity on Demand**
- Processor resources only (processor days)
- Capacity can be turned on and off by the client
- Prepaid debit agreement
- Requires AIX 5L V5.3 and APV

**Trial Capacity on Demand**
- Allow clients to test the effects of additional processors and/or memory
- Partial or total activation of processors and memory
- Resources available for fixed time
- No formal commitment required

**Dynamic Processor Sparing**
- Automated replacement of de-allocated processors
- Unassigned or inactive processors
IBM System p5

First Failure Data Capture

- Specialized hardware designed to capture failure data at the time of failure
- Repair based on root-cause analysis
- Direct correlation to original problem
- Engineering focus on built-in error detection and capture
- Service action plan driven by captured failure information
- Testcases used only to verify operation

Competitive Strategy - Failure Re-create
- Run diagnostic testcases during service call
- Repair based on testcase symptom
- Questionable correlation to original customer problem
- Development focus on improved testcases
- Open service action plan if failure not re-created
- Testcases used to isolate failures and verify correct operation

First Failure Data Capture
- Specialized hardware designed to capture failure data at the time of failure
- Repair based on root-cause analysis
- Direct correlation to original problem
- Engineering focus on built-in error detection and capture
- Service action plan driven by captured failure information
- Testcases used only to verify operation

Reliably identify failing component reducing costly downtime

CPU
L1 Cache
L2 Cache
Memory
Service Processor
Fault Isolation Register (FIR)
Error Checkers
Disk
Nonvolatile RAM
Log Error
POWER RAS additions...

Primary POWER RAS features

- First Failure Data Capture
- DDR Chipkill™ memory
- Bit-steering/redundant memory
- Memory soft scrubbing
- Redundant power, fans
- Dynamic Processor Deallocation
- Dynamic processor sparing
- ECC memory
- Persistent memory deallocation
- Hot-plug PCI slots, fans, power
- Internal light path diagnostics
- Hot-swappable disk bays
- I/O error handling extended beyond base PCI adapter
- ECC extended to inter-chip connections for the fabric/processor buses

2005 RAS Enhancements

- **Service Processor Failover:**
  - p5-570, p5-590 and p5-595
  - New service processor option for p5-570
    - Required for “Failover” support

- **Dynamic Firmware Maintenance:**
  - Apply firmware fixes without system disruption
  - Fixes only, not new functionality

- **Hot I/O Drawer Add**
  - Install remote I/O drawers without system disruption

HMC required to enable these functions.
Enable high system availability – Building on world-class hardware RAS

<table>
<thead>
<tr>
<th>Summary of key pSeries RAS features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core System Design</strong></td>
</tr>
<tr>
<td>• High quality parts</td>
</tr>
<tr>
<td>• Fewer parts = Fewer failures</td>
</tr>
<tr>
<td>• Designed for low power consumption (less heat = fewer failures)</td>
</tr>
<tr>
<td>• Manufacturing methods, packaging, cooling</td>
</tr>
<tr>
<td>• Continuous System and Commodity Quality Actions</td>
</tr>
<tr>
<td>• Integrated RAS features</td>
</tr>
<tr>
<td>• Failure Avoidance Methodology</td>
</tr>
<tr>
<td>• Designed for Ease of Service</td>
</tr>
<tr>
<td><strong>Fault Resilience</strong></td>
</tr>
<tr>
<td>• N+1 Power Supplies, regulators, power cords</td>
</tr>
<tr>
<td>• Dual redundant fans</td>
</tr>
<tr>
<td>• Dynamic Processor Deallocation and sparing</td>
</tr>
<tr>
<td>• &quot;Chipkill&quot; Technology</td>
</tr>
<tr>
<td>• Predictive Failure Analysis</td>
</tr>
<tr>
<td>• Auto Path Reassignment - data paths, power</td>
</tr>
<tr>
<td><strong>System Restore</strong></td>
</tr>
<tr>
<td>• Deferred Repair</td>
</tr>
<tr>
<td>• Concurrent Repair</td>
</tr>
<tr>
<td>• LED Service Identification</td>
</tr>
<tr>
<td>• Service Consoles</td>
</tr>
<tr>
<td>• Migration to Guided Maintenance</td>
</tr>
<tr>
<td><strong>Fault Isolation &amp; Diagnosis</strong></td>
</tr>
<tr>
<td>• First Failure Data Capture</td>
</tr>
<tr>
<td>• Run Time Self Diagnostics</td>
</tr>
<tr>
<td>• Service Processor</td>
</tr>
<tr>
<td>• Rifle-shot repairs (no &quot;plug and pray&quot; parts replacement approach)</td>
</tr>
</tbody>
</table>
Summary

Why IBM System p AIX/LINUX solutions

– IBM has a **Proven Technology**
– IBM has a **Consistent** Product line
– IBM has a **Clear Direction**
  – A good hockey player plays where the puck is. A great hockey player plays where the puck is going to be.
  – **Wayne Gretzky**

– IBM wants to **EARN** your business