Exploiting Virtualisation on IBM Power Systems with PowerVM

Nigel Griffiths
IBM Power Systems
Advanced Technology Support
EMEA

Processor Virtualisation (CPU Sharing)

Old Style
Separate Systems
Pre-2000
### Processor Virtualisation (CPU Sharing)

<table>
<thead>
<tr>
<th>Old Style</th>
<th>LPAR</th>
<th>DLPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate Systems</td>
<td>Server</td>
<td>Dynamic</td>
</tr>
<tr>
<td></td>
<td>Consolidation</td>
<td>live boundary</td>
</tr>
<tr>
<td></td>
<td>LPAR size</td>
<td>changes</td>
</tr>
<tr>
<td></td>
<td>via start time</td>
<td>Manual or</td>
</tr>
<tr>
<td></td>
<td>boundaries</td>
<td>scripts</td>
</tr>
<tr>
<td>Pre-2000</td>
<td>~2001</td>
<td>~2002</td>
</tr>
</tbody>
</table>
### Processor Virtualisation (CPU Sharing)

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Old Style</th>
<th>LPAR</th>
<th>DLPAR</th>
<th>SPLPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-2000</td>
<td>Separate Systems</td>
<td>Server Consolidation</td>
<td>Dynamic live boundary changes</td>
<td>Shared Processor automatic adjusts at millisecond level by Hypervisor</td>
</tr>
<tr>
<td>~2000</td>
<td>~2001</td>
<td>~2002</td>
<td>~2005</td>
<td></td>
</tr>
</tbody>
</table>

- **Old Style:** Separate Systems
- **LPAR:** Server Consolidation, LPAR size via start time boundaries
- **DLPAR:** Dynamic live boundary changes, Manual or scripts
- **SPLPAR:** Shared Processor automatic adjusts at millisecond level by Hypervisor

~2005
Processor Virtualisation (CPU Sharing)

SPLPAR
Shared Processor
automatic adjusts
at millisecond level
by Hypervisor

~2005

Processor Virtualisation (CPU Sharing)

SPLPAR
Shared Processor
automatic adjusts
at millisecond level
by Hypervisor

Harvesting
“Spare” capacity
ready for adding
more workloads at
zero hardware cost

~2005  ~2006
Processor Virtualisation (CPU Sharing)

<table>
<thead>
<tr>
<th>SPLPAR</th>
<th>Harvesting</th>
<th>Partition Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared Processor</td>
<td>“Spare” capacity</td>
<td>Make a cluster of your machines &amp; flow your workload between them</td>
</tr>
<tr>
<td>automatic adjusts</td>
<td>ready for adding more workloads at zero hardware cost</td>
<td></td>
</tr>
<tr>
<td>at millisecond level by Hypervisor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- ~2005
- ~2006
- ~2008

Virtualisation - Value Proposition
Virtualisation - Value Proposition

Economical I/O Model

Reduced Infrastructure
Virtualisation - Value Proposition

"Your new system will be ready in ..."
"20 Minutes" or "20 Days"

10% Virtualisation - Value Proposition
10% Virtualisation - Value Proposition
10% Virtualisation - Value Proposition
10% Virtualisation - Value Proposition
10% Virtualisation - Value Proposition

Economical I/O Model

Quick Deployment

SAN + Net Cables Switches Ports Power

Reduced Infrastructure

SAN + Net Cables Switches Ports Power

Reduced Infrastructure

SAN + Net Cables Switches Ports Power

Server Consolidation
Virtual Adapters

The I/O centric view of the world:
- CPU used to “modify & feed data” to the networks & disks

Virtual I/O Server (Adapter Sharing)
Virtual I/O Server (Adapter Sharing)

- Production Direct
- The rest Virtual
- Pure Direct
- Virtual except high I/O LPARs
Storage Virtualisation since 2005
Virtual CD/DVD & Virtual Tape

Virtual Tape: Easy to move between LPARs

Virtual: CD & DVD Read-only DVD-RAM Read-Write

Storage Virtualisation with N-Port ID Virtualisation

VIOS Admin in charge

SAN Admin Back in charge
Dual VIOS since 2001 allows concurrent updates 
→ supporting 2 to 200+ Logical Partitions

Noticed: CPUs too fast & memory bloat
What about Virtual Memory!

Virtual Memory available in UNIX to 30+ years
Modern Software need extreme amounts of memory

Power has two mechanisms to increase memory use
- Active Memory Expansion (AIX only)
- Active Memory Sharing
- AIX Workload Partition can also save memory

AME Conceptual Model
Active Memory Expansion

Memory Pages

Dynamically adjusted depending on compression ratio & target
AME - What is your Plan?

Memory Shrinking

to release RAM for other uses

10GB

 Looks like 10GB but is actually 8GB, thanks to AME

2GB released for another LPAR

Memory Growing

for RAM optimisation & performance

10GB

 But want 14GB to improve performance

Actually still using 10 GB but looks like 14 GB, thanks to AME

Classic Virtual Memory (LPAR)

Virtual Memory

Physical Memory

Not Really Here

Tiny 4 KB pages
Active Shared Virtual Memory (LPAR)

Virtual Memory

Logical Memory

Not Really Here

Actually stored or disk

Blue Physical Memory Pages

Loaned Memory To Hypervisor

VIOS

Active Memory Sharing (Virtual Virtual Memory!)

VIO client

SCSI SAS

VIOS

SAN

VIO client

SCSI SAS

VIO client

VIO client

VIO client
Active Memory Sharing (Virtual Virtual Memory!)
Active Memory Sharing (Virtual Virtual Memory!)

Stealing Mode

AME & AMS Comparisons

Active Memory Expansion
- Jan 2010
- AIX6 TL4+ on POWER7
  - Not Linux nor IBM i
- Machine Activation (LPP)
  - 60 day trial
- Pure Virtual LPAR
- Internal to single LPARs
- Assume “spare” CPU cycles for compression
- Simple to setup in LPAR
- Use amepat to predict the compression factor
- Use topas/nmon to monitor

Active Memory Sharing
- May 2009
- POWER6
- AIX6 TL3+, Linux & IBM i 6.1
- PowerVM Enterprise
- Pure Virtual LPAR
- Cooperating group of LPARs
- Assumes loanable RAM
- Pages flow between LPARs at a few MB/s
- More complex to setup on VIOS & LPARs
- Use topas –C to monitor
### Summary so far ....

**Dedicated World**: still available
- Dedicated CPU
- Dedicated Virtual Memory
- Dedicated Disk adapters - SCSI/SAS/SAN
- Dedicated Network adapters
- NFS or NIM !
- CPU load balancing in minutes/hours
- Memory load balancing in minutes/hours

**Virtual World**
- Virtual CPUs
- Shared Virtual Memory (AMS) + AME
- Virtual Disks (vSCSI & NPIV)
- Virtual Networks
- Virtual CD/DVD & Tape & NFS & NIM
- CPU load balancing in milli-seconds
- Active Memory Sharing between LPARs in seconds

**Live Partition Mobility**

---

**Summary so far ....**

**Dedicated World**: still available
- Dedicated CPU
- Dedicated Virtual Memory
- Dedicated Disk adapters - SCSI/SAS/SAN
- Dedicated Network adapters
- NFS or NIM !
- CPU load balancing in minutes/hours
- Memory load balancing in minutes/hours

**Virtual World**
- Virtual CPUs
- Shared Virtual Memory (AMS) + AME
- Virtual Disks (vSCSI & NPIV)
- Virtual Networks
- Virtual CD/DVD & Tape & NFS & NIM
- CPU load balancing in milli-seconds
- Active Memory Sharing between LPARs in seconds

**Live Partition Mobility**

---

**Expensive = “Old school”, Cheap & Flexible**
Live Partition Mobility (LPM)

High levels of efficient Virtualisation means
• Hypervisor support (system firmware)
• Virtual I/O Server (VIOS)
• HMC for management
• Operating System support
For new features & fixes … updates are required
LPM allows for zero downtime updates
Also allows
1. Workload balancing across machine room
2. New machine use at day 1
3. Repair actions & upgrades
Live Partition Mobility (2 of 6)

Live Partition Mobility (3 of 6)
Live Partition Mobility (4 of 6)

Live Partition Mobility (5 of 6)
Live Partition Mobility (6 of 6)

Suspend and Resume

- Resource balancing: suspend low-priority for more urgent processes
- Simplified maintenance: suspend the whole machine system updates
- Debug/forensics: a workload can be temporarily suspended, a copy made for offline analysis for security or performance purposes
Resume

4) Recreate the LPAR

5) Copy memory from the VIOS disk

6) Unfreeze the LPAR
PowerVM Editions offer a unified virtualization solution for all Power workloads.

- **PowerVM Express Edition**
  - Evaluations, pilots, PoCs
  - Single-server projects

- **PowerVM Standard Edition**
  - Production deployments
  - Server consolidation

- **PowerVM Enterprise Edition**
  - Multi-server deployments
  - Advanced Functions

### PowerVM Editions

<table>
<thead>
<tr>
<th>PowerVM Editions</th>
<th>Express</th>
<th>Standard</th>
<th>Enterprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent VMs</td>
<td>VIOS + 2 per VMs</td>
<td>10 per core (up to 1000)</td>
<td>10 per core (up to 1000)</td>
</tr>
<tr>
<td>Virtualization Management</td>
<td>IVM</td>
<td>IVM, HMC</td>
<td>IVM, HMC</td>
</tr>
<tr>
<td>Virtual I/O Server</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PowerVM Lx86</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Suspend/Resume</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shared Processor Pools</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shared Storage Pools</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Thin Provisioning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Live Partition Mobility</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Active Memory Sharing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- **IBM i** supports shared storage and Suspend & Resume, but does not support LPM.

### Virtual Disks 5 type - Reminder

1. Logical Volume from VG Pool
2. Whole Local Disk
3. SAN LUN
4. File-Backed from a file system (pool) based on local or SAN disk(s)
5. NPIV LUN

Virtual I/O Server: vSCSI, NPIV over vFC

Client VM: 3 SAN LUN, 5 NPIV LUN
Now one more

1 Logical Volume from VG Pool
2 Whole Local Disk
4 File-Backed from a file system (pool) based on local or SAN disk(s)

Shared Storage Pool
distributed cluster
co-operating VIOS’s

Virtual I/O Server
Virtual I/O Server
Virtual I/O Server

vSCSI
vSCSI
vSCSI

6 SAN LUN Disks
3 SAN LUN
5 NPIV LUN

Phase 1 Single VIOS Thin Provisioning
NPIV over vFC

Cluster Repository
Pools Disks

Phase 2

Client VM
Client VM
Client VM

1. Thin & Thick Provisioning
2. Disk online to VIOS only
3. Very simple LPM setup
4. Data Migration
5. Snapshot and Cloning disks
6. Linked Clones
7. GUI via Systems Director

Power Systems
Since September 2010