

# Planning for Virtualization

Jaqui Lynch  
UKCMG

Jaqui.lynch@mainline.com  
<http://www.circle4.com/papers/ukvirtual.pdf>

## Agenda

- Partitioning Concepts
- Virtualization
- Planning
- Hints and Tips
- References

## Partitioning Concepts

- Logical Partitions
- Full System Partition
- Managed Systems
  - Partitions
- Profiles
  - Partition Profiles
    - Info on assigned resources for partitions
    - Activating this activates an LPAR
  - System Profiles
    - Collection of partition profiles to be activated at the same time

## Dynamic LPAR

- Allows moving of resources between LPARs without a reboot
- Can move:
  - Memory
  - Processors
  - I/O cards

## Dynamic LPAR

- Add processors to partition
- Move processors between partitions
- Remove processors from a partition
- Add memory to a partition
- Move memory from one partition to another
- Remove memory from a partition
- Add a PCI adapter
- Move a PCI adapter
- Remove a PCI adapter

## Reasons to Partition

- Consolidation
- Production and Test on same hardware
- Multiple Operating Systems
- Consolidate Applications on different time zones
- Complying with license agreements

## Role of the HMC

- Required to partition any box
- Can use HMC to manage systems
- Provides a console to manage hardware
- Detecting, reporting and storing changes in hardware
- Service focal point (requires Ethernet)
- Vterms to partitions
- COD
- Inventory and microcode management
- Clustering
- Remote power control
- Virtualization
- Creating in memory channels for Virtual Ethernet
- 7310 is POWER5 only
  - Uses USB ports – no PS/2
  - No serial – has ethernet 10/100/1000
- Some 7315 models can be upgraded to support POWER5
  - They cannot then go back to supporting POWER4

## Supported Operating Systems

- AIX 5.3
  - Enables Virtualization when on Power5
- AIX 5.2
  - Minimum of ML4 required for Power5
- AIX 5.1
  - Will not run on Power5 systems
  - Does not support:
    - Dynamic LPAR
    - Memory Capacity Upgrade on Demand
    - Dynamic Processor Sparing
    - Dynamic CPU Guard
- Suse Linux, United Linux 1.0, Redhat EL AS3, Turbolinux and Conectiva Linux
- No version of AIX prior to v5 will work
- Check required ML levels for each box
- Check required microcode levels on HMC, pSeries boxes and cards, especially fiber cards

## Software

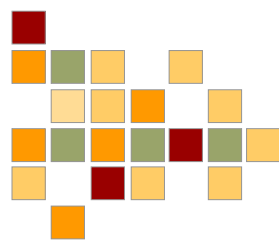
- Make sure HMC and all boxes are at the latest microcode level
- pSeries Microcode can be found at:
  - <http://techsupport.services.ibm.com/server/mdownload>
- HMC Corrective Service can be found at:
- <https://techsupport.services.ibm.com/server/hmc/power4>
  - <https://techsupport.services.ibm.com/server/hmc/power5>
- Latest HMC Software version is
  - Power 4/4+ - v3v3.1 as of July 28, 2004
  - Power5 – v4R1.2 as of July 5, 2004
- Don't forget BIOS updates which are at the HMC locations above
- As of March 2004 HMC maintenance is now a customer responsibility.

## Planning – Power4/4+

- Each Power4/4+ LPAR must have the following
  - 1 processor
  - 256mb memory
  - LMBs are 256mb
  - 1 boot disk
  - 1 adapter to access the disk
  - 1 Ethernet adapter to access the HMC
  - An installation method such as NIM
  - A means of running diagnostics
- The above also applies to Power5 partitions running AIX v5.2 ML4 or earlier versions of RHAS and SLES

## Planning – Power5

- Each Power5 LPAR running AIX v5.3 with APV must have the following
  - 1/10 processor
  - 128mb memory
  - LMBs are 16mb
  - 1 boot disk (virtual or real)
  - 1 adapter to access the disk (virtual or real)
  - 1 Ethernet adapter to access the HMC (virtual or real)
  - An installation method such as NIM
  - A means of running diagnostics



Virtualization

## Terminology

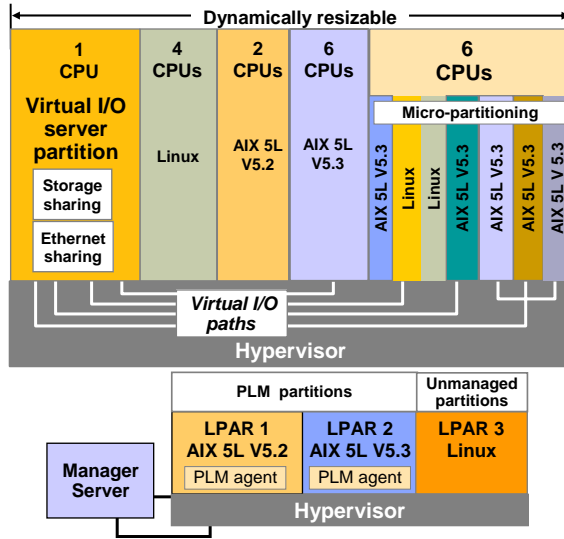
- Hypervisor
- MicroPartitioning
  - Shared Processor Pool
  - Capped
  - Uncapped
  - Virtual Processors
  - Entitled Capacity
- Virtual I/O Server
- Virtual Ethernet
- Shared Ethernet Adapter (SEA)
- Virtual SCSI Server

## APV

### Advanced Power Virtualization Feature

- Included on 590 and 595
- Optional on all other p5 boxes
- Enables:
  - MicroPartitioning
  - Virtual I/O Server
  - Shared Ethernet Adapter
  - Virtual SCSI Server
  - Partition Load Manager
- OPV (Openpower PV)
  - Equivalent to APV for Openpower
  - Also required to enable partitioning on OP

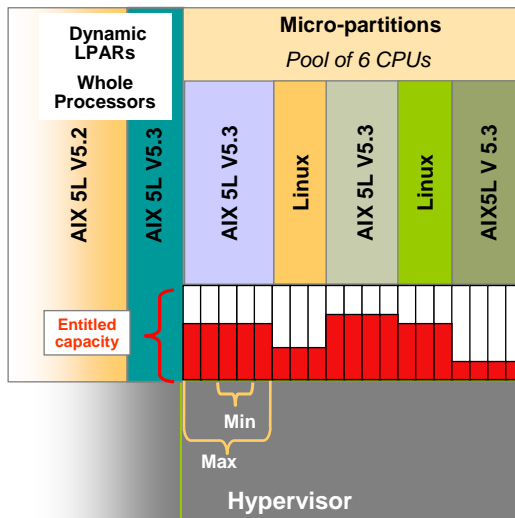
# p5 advanced virtualization option



- Virtual I/O server**
  - Shared Ethernet
  - Shared SCSI and Fibre Channel-attached disk subsystems
  - Supports AIX 5L V5.3 and Linux\* partitions
- Micro-Partitioning**
  - Share processors across multiple partitions
  - Minimum partition 1/10<sup>th</sup> processor
  - AIX 5L V5.3 or Linux\*
- Partition Load Manager**
  - Both AIX 5L V5.2 and AIX 5L V5.3 supported
  - Balances processor and memory request
- Managed via HMC**

\*SLES 9 or RHEL AS 3

# Micro-Partitioning



- Increased number of LPARs**
  - Micro-Partitions: 160\*
  - Dynamic LPARs: 16\*
- Configured via the HMC**
- Number of logical processors**
  - Minimum/maximum
- Entitied 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**

\*\*Capped with weight=0 is functionally identical to uncapped

\*\*Processors, by default, are available to the SPP unless they are being used by a dedicated partition

\*on p5-570

## Defining Processors

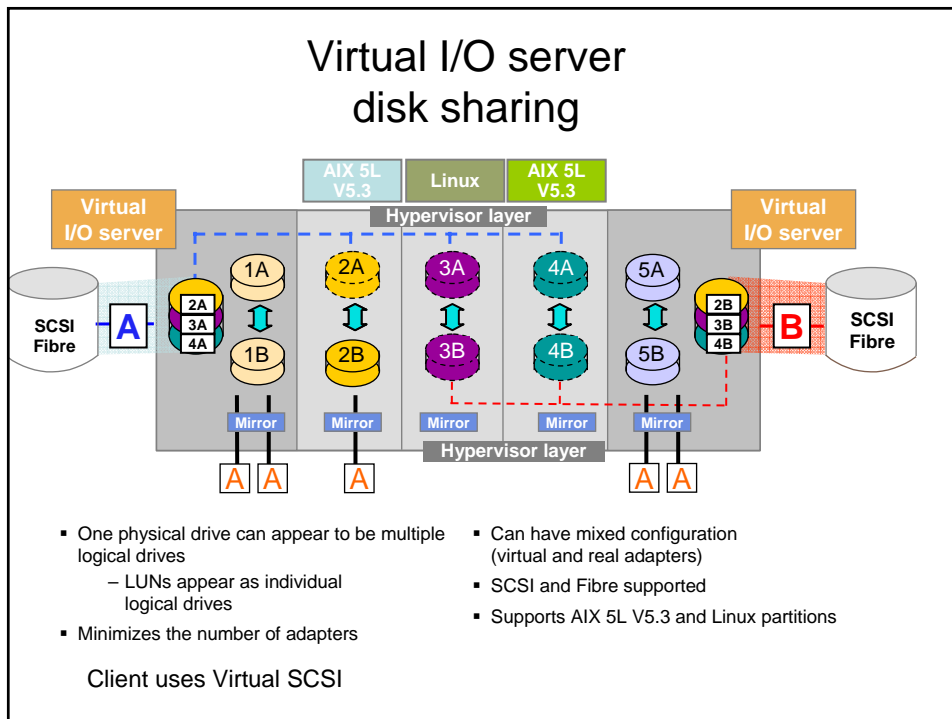
- Minimum, desired, maximum
- Shared or dedicated
- For shared:
  - Capped
  - Uncapped
    - Variable capacity weight (0-255 – 128 is default)
    - Weight of 0 is capped
  - Minimum, desired and maximum Virtual Processors

## Virtual Processors

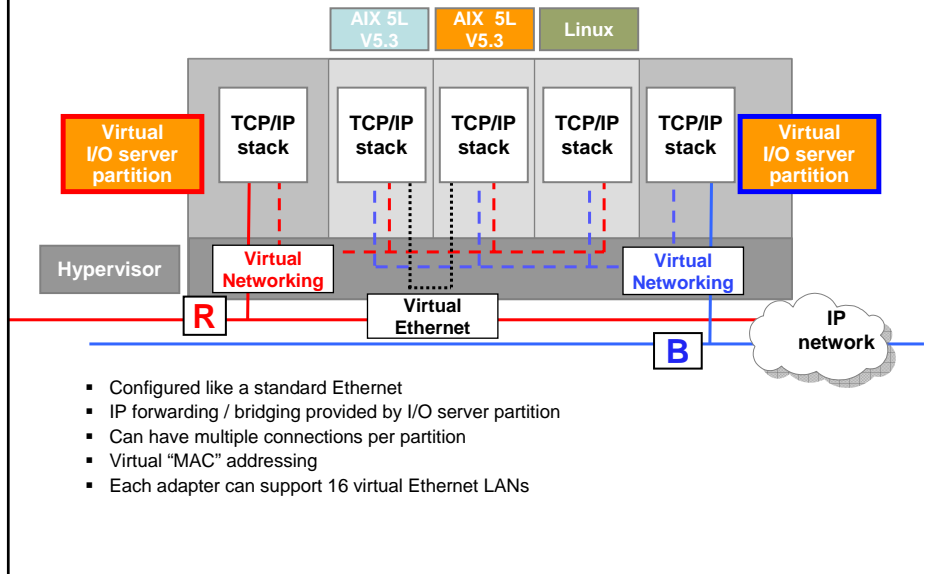
- Partitions are assigned Pus (process units)
- VPs are the whole number of concurrent operations
  - i.e. how many processors do I want these Pus to map to
- Maximum of 64 VPs
- VPs round up from the PU by default
  - .5 Pus will be 1 VP
  - 2.25 Pus will be 3 VPs
  - You can define more and may want to
- VPs put a cap on the partition if not used correctly
  - i.e. define .5 PU and 1 VP you can never have more than one PU even if you are uncapped

# Virtual I/O Server

- Custom AIX v5.3 partition
- Provides services for:
  - Shared Ethernet Adapter
    - Built on Virtual Ethernet
  - Virtual SCSI Server
- Owns the physical resources
- Run 2 if in production
- Can use SDD for multipath I/O
- Can do Etherchannels
- Maximum of 65535 virtual I/O slots
- Max of 256 VIO slots per partition
- Mirroring done in client



## Virtual I/O server Ethernet sharing

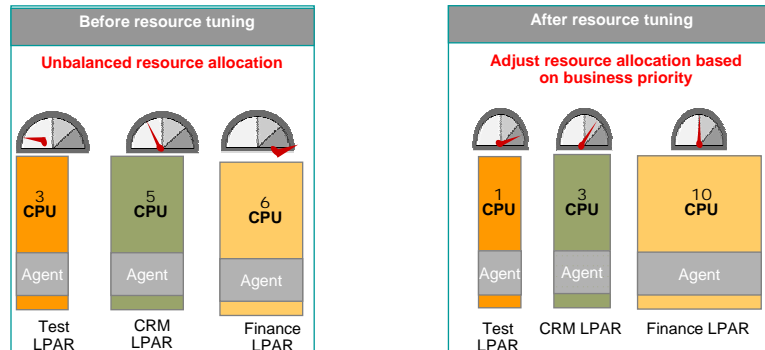


## Workload Manager

- Around since AIX v4.3.3
- Used to control resources in one O/S instance
- Controls:
  - CPU
  - Memory
  - I/O Bandwidth
- References:
  - SG24-5977 AIX 5I Workload manager
  - <http://www.eservercomputing.com/>
    - Then search on Jaqui AND Lynch

## Partition Load Manager for AIX 5L p5 systems

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



Requires AIX v5.2 ML4 or AIX v5.3  
HMC must be at least v3.2.6

## Capacity on Demand for p5 systems

### 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

## Power4 Memory

- Only have a Hypervisor if running partitioned
- With Hypervisor some memory is reserved for LPAR use
  - Hypervisor - 256mb
  - Power4 - TCE (Translation Control Entry) – 256mb to 1gb
    - Used to translate I/O addresses to system memory addresses
    - Always 256mb on a p630
  - Page Table Entries
    - min 256mb on Power4
  - So overhead for the first 256mb partition is at least 768mb
- For 2 or more LPARS expect overhead to be at least 2gb memory

## Power5 Memory

- In POWER5 you always have a Hypervisor
- Some memory is reserved for LPAR use
  - Hypervisor and Global Openfirmware - 256mb starting at 0
  - HPT (Hypervisor Page Table) Entries
    - 1 per partition
    - Reserves 1/64 of maximum memory setting
  - For 2 or more LPARS expect overhead to be at least 1gb memory
- LVT tool used to get estimates for Power5
  - <http://www-1.ibm.com/servers/eserver/series/lpar/systemdesign.htm>

# LVT Tool

**Memory Specifications**

System Model: 9113\_550  
 Processor/Package Feature: 5237  
 System Memory (GB): 64.0  
 Total Processors: 4

System Memory(MB): 65536  
 Configured Memory(MB): 60764  
 Hypervisor Memory(MB): 768  
 Unallocated Memory(MB): 4004

Partition	OS Version	Memory	Max Memory	Virtual Slots	Virtual Ethernet	Virtual Serial	Server SCSI	Client SCSI
P1	I/O_Virtual_Ser...	32764	32764	6	0	2	4	0
P2	ADX_53	8000	0	4	0	2	0	1
P3	ADX_53	8000	0	4	0	2	0	1
P4	ADX_53	12000	0	4	0	2	0	1

OS/400 License(s) Required: 0.0  
 ADX License(s) Required: 4.0  
 Linux License(s) Required: 0.0

< Back Finish Cancel

# LVT Tool

**Memory Specifications**

System Model: 9113\_550  
 Processor/Package Feature: 5237  
 System Memory (GB): 64.0  
 Total Processors: 4

System Memory(MB): 65536  
 Configured Memory(MB): 57344  
 Hypervisor Memory(MB): 1280  
 Unallocated Memory(MB): 6912

Partition	OS Version	Memory	Max Memory	Virtual Slots	Virtual Ethernet	Virtual Serial	Server SCSI	Client SCSI
P1	I/O_Virtual_Ser...	32768	32768	6	0	2	4	0
P2	ADX_53	8192	8192	4	0	2	0	1
P3	ADX_53	8192	8192	4	0	2	0	1
P4	ADX_53	8192	16384	4	0	2	0	1

OS/400 License(s) Required: 0.0  
 ADX License(s) Required: 4.0  
 Linux License(s) Required: 0.0

< Back Finish Cancel

# LVT Tool

```
LPARValidator Report Viewer
P1 Partition OS Level.....: I/O_Virtual_Server
System Model.....: pSeries Model 550
Processor/Package Feature...: 5237 - Model 9113-550 0/2way 1.65 Ghz
Interactive Feature.....: N/A
Console Type.....: -
System Memory (GB).....: 64.0
Total Processors.....: 4
Dedicated Processors.....: 0
Shared Processors.....: 3.0
Batch CPW.....: 0
Interactive CPW.....: 0
Total Partitions.....: 4
OS/400 Licenses Required....: 0
AIX Licenses Required.....: 2
Linux Licenses Required.....: 2

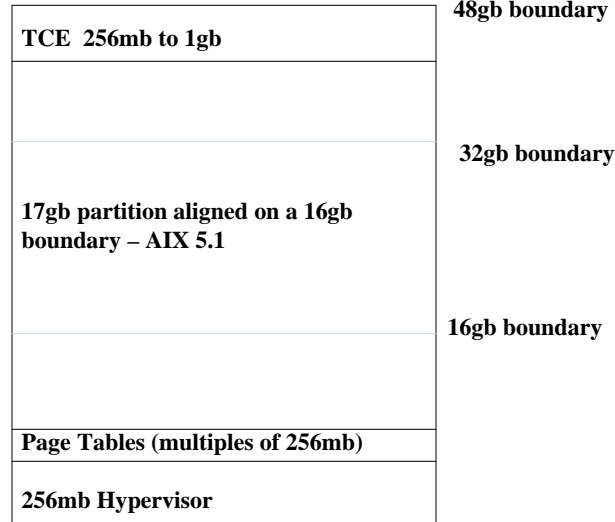
-- Partition Specifications --
Partition  OS Ver  Shared # Procs # Max Procs  Batch CPW  Unc Max
-----
P1         I/O_Virtual_S  N    1.00    0    0    0    0
P2         AIX_Virtual_C    Y    0.50    0    0    0    0
P3         AIX_Virtual_C    Y    0.50    0    0    0    0
P4         Linux_Virtual    Y    2.00    0    0    0    0
-----
Totals         -----  ---    4.00    0    0    0    0

-- Memory Specifications --
Partition  OS Ver  Memory  Max Memory  V Slots  V Ethernet  V Serial  C SCSI  S SCSI
-----
P1         I/O_Virt  32768.00  32768.00    6.00    0.00    2.00    0.00  4.00
P2         AIX_Virt  8192.00   8192.00    3.00    0.00    2.00    1.00  0.00
P3         AIX_Virt  8192.00   8192.00    3.00    0.00    2.00    1.00  0.00
P4         Linux_Vi  8192.00   8192.00    3.00    0.00    2.00    1.00  0.00
-----
Totals         -----  57344.00  57344.00   15.00    0.00    8.00    3.00  4.00
```

## Real Mode Address Region (RMO)

- Small Real Mode Address Region
  - Allows you to use managed system memory more efficiently
  - Only valid for Linux and AIX 5.2 or higher
  - Avoids the memory boundary constraints
  - AIX 5.1 may not boot if you turn this on
- Large Real Mode Address Region (DEFAULT)
  - Assigns memory on 256mb, 1gb and 16gb boundaries (contiguous real mode memory)
  - Partition <= 16gb gets 1gb plus the rest in 256mb increments

## Power4 Memory



## Hints and Tips

- Which LPAR is your service LPAR?
- How will you do installs
  - Allocate cd?
  - NIM?
- Backup Methodology?
- If using virtualization planning is more critical than ever
- Ensure Inventory scout is working on all LPARs and that VPD is being uploaded to IBM
- Create a partition layout in advance
  - Include devices, etc
- I/O devices are allocated at the slot level
- Which planar is in the I/O drawer
  - Affects the number of high-speed adapters
- Boot disks –
  - I/O drawer or 2104, Raid, Fiber
- 32bit kernel versus 64bit kernel
  - 32 bit supports up to 96gb memory
  - Need 64bit kernel to have more than 96gb in an LPAR
  - Need 64bit kernel for more than 16 processors in an LPAR

# Tools

`lparstat -h`  
percentage spent in Hypervisor and number of Hcalls

`lparstat -i`  
Info on entitled capacity, setup info, etc

`mpstat -s`  
SMT info

`mpstat -d`  
Detailed affinity and migration statistics

`sar -P ALL`

`topas -L`

# lparstat

`lparstat -h`

System Configuration: type=shared mode=Uncapped smt=On lcpu=4 mem=512 ent=5.0

**%user %sys %wait %idle** physc %entc lbusy app vcsw phint **%hypv hcalls**

0.0 0.5 0.0 99.5 0.00 1.0 0.0 - 1524 0 0.5 1542

16.0 76.3 0.0 7.7 0.30 100.0 90.5 - 321 1 0.9 259

Physc – Number of physical processors consumed

%entc – Percentage of entitled capacity consumed

Lbusy – logical processor utilization for system and user

Phint – phantom interrupts to other partitions (targeted to another shared partition in this pool)

%hypv – Percentage of time spent in Hypervisor

Hcalls – Number of Hypervisor calls

`lparstat -H`

Gives info per Hypervisor call type as follows:

Number of calls, Time spent on this types of calls, Hypervisor time spent on this type of call,

Average call time, Max call time

<http://publib.boulder.ibm.com/infocenter/pseries/index.jsp?topic=/com.ibm.aix.doc/cmds/aixcmds3/lparstat.htm>



## Maximums

- 254 partitions per server or  $10 * \#$  processors (whichever is smaller)
- 64 Virtual processors per partitions
- 256 Virtual Ethernet adapters per partition
- 21 VLANs per VE adapter
- 16 VEs per physical adapter (SEA) with 21 VLANs per

## References

- IBM Redbooks
  - SG24-7940 – Advanced Power Virtualization on IBM p5 servers – Introduction and Basic Configuration
  - SG24-5768 - Advanced Power Virtualization on IBM p5 servers – Architecture and Performance Considerations
  - The Complete Partitioning Guide for IBM eServer pSeries Servers
  - pSeries – LPAR Planning Redpiece
  - Logical Partition Security in the IBM eServer pSeries 690
  - Technical Overview Redbooks for p520, p550 and p570, etc
  - SG24-7039 - Partitioning Implementation on p5 and Openpower Servers
  - Redpiece 9115 - Virtualization and the On Demand Business
- eServer Magazine
  - <http://www.eservercomputing.com/>
  - Click on the UNIX server icon
    - \* Feb 2005 focussed on Virtualization
  - AixTra – Sign up at: <http://eservercomputing.com/>
  - <http://www.circle4.com/jaqui/eserver.html>
- Find more on Mainline at:
  - <http://mainline.com/ebrochure>
- Power5 Information Center
  - [http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en\\_US/index.htm](http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm)

# Questions

