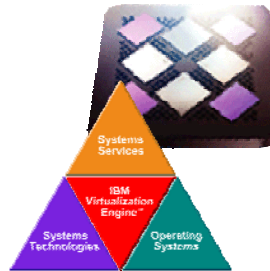


# Planning and Sizing Virtualization



**Jaqui Lynch**

**Architect, Systems Engineer**

**Mainline Information Systems**

<http://www.circle4.com/papers/cmgvirt.pdf>

**Mainline:** solutions you need  
from people you trust 1

## What is Virtualization?

- Being able to dynamically move resources
- Being able to share resources
- Making better use of the resources
- Driving utilization up
- Etc etc etc
  
- Some think it is 42



**Mainline:** solutions you need  
from people you trust 2

## Reasons to Partition

- Consolidation
- Hosting
- Production and Test on same hardware
- Multiple Operating Systems
- Consolidate Applications on different time zones
- Complying with license agreements
- Flexibility and scalability
- Optimization of resources

Mainline: solutions you need  
from people you trust 3

## Planning Power4/4+ or AIX 5.2 or dedicated

- Each Power4/4+ LPAR must have the following
  - 1 processor
  - 256mb memory
  - 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 all partitions running AIX v5.2 ML4 or earlier versions of RHAS and SLES
- Also applies to dedicated LPARs, even v5.3

Mainline: solutions you need  
from people you trust 4

## Planning – Power5

- Each Power5 LPAR running AIX v5.3 (or higher) with APV (advanced power virtualization feature) installed must have the following:
  - 1/10 processor
  - 128mb memory (really needs more than this to run)
  - 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

Mainline: solutions you need  
from people you trust 5

## Power5 Memory

- In POWER5 and higher you always have a Hypervisor
- Some memory is reserved for LPAR use
  - Hypervisor - 256mb
  - HPT (Hypervisor Page Table) Entries
    - 1 per partition
    - Reserves 1/64 of **maximum** memory setting and rounds up to nearest LMB (usually 256)
  - For 2 or more LPARS expect overhead to be at least 1gb memory
  - Use 8% of memory as an estimate of overhead
    - Do not forget it!!!!
- SPT replaces LVT tool used to get estimates for Power5 and Power6
  - <http://www-1.ibm.com/servers/eserver/series/lpar/systemdesign.htm>
  - Latest version is v2.07.229

Mainline: solutions you need  
from people you trust 6

## Planning – Power6

- To go to v7 on HMC you need to order feature code 0962
- Similar requirements to Power5
  - 1/10 of a processor, etc
  - Need APV to enable micro-partitioning and virtualization features
  - Similar memory overhead
- PCI-E and PCI-X not interchangeable so you must know where a card is going
  - Each CEC has 4 x PCI-E and 2 x PCI-X
  - I/O drawers are PCI-X
- Other p6-570 notes
  - all six disks on one adapter
  - One media bay per CEC
    - Is on same adapter as disk
    - Need to look at virtualizing DVD
  - Give up a PCI-E slot if adding RIO drawers and must add a RIO card as default is Infiniband

Mainline: solutions you need  
from people you trust 7

## 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
- CuOD
- Virtualization
- New interface for v7 and v7 required for power6
- The Overseer:



Mainline: solutions you need  
from people you trust 8

Hardware Management Console

Contents of: p5-570

Select	Name	ID	Status	Processing Units	Memory (GB)	Active Profile	Environment	Reference Code
<input type="checkbox"/>	p5aix52	2	Not Activated	0	0	default	AIX or Linux	00000000
<input type="checkbox"/>	p5aix6a	11	Running	0.5	1	default	AIX or Linux	
<input type="checkbox"/>	p5aix6b	12	Running	0.5	1	default	AIX or Linux	
<input type="checkbox"/>	p5aix63	4	Not Activated	0.5	0.5	default	AIX or Linux	00000000
<input type="checkbox"/>	p5aix65	5	Not Activated	0.5	0.8125	default	AIX or Linux	00000000
<input type="checkbox"/>	p5nim	1	Running	0.5	1	default	AIX or Linux	
<input type="checkbox"/>	p5orarc1	9	Not Activated	0	0	default	AIX or Linux	00000000
<input type="checkbox"/>	p5orarc2	10	Not Activated	0	0	default	AIX or Linux	00000000
<input type="checkbox"/>	p5ihat1	7	Not Activated	0.2	0.5	default	AIX or Linux	00000000
<input type="checkbox"/>	p5iaes1	6	Not Activated	0.2	0.5	default	AIX or Linux	00000000
<input type="checkbox"/>	p5vios	3	Running	0.5	1	default	Virtual I/O Server	
<input type="checkbox"/>	p5vios2	8	Running	0.5	1	default	Virtual I/O Server	
				Total: 12	Filtered: 12	Selected: 0		

Tasks: p5-570 [ Expand All | Collapse All ]

- Properties
- Operations
- Configuration
- Connections
- Hardware Information
- Updates
- Servicability
- Capacity On Demand (CoD)

Status: Attentions and Events

# Memory Usage

p5-570

General Processors **Memory** I/O Power-On Parameters Capabilities

Details of the managed system's memory are listed below.

Installed memory: 8192MB  
 Deconfigured memory: 0MB  
 Available memory: 0MB  
 Configurable memory: 8192MB  
 Memory region size: 32MB  
 Current memory available for partition usage : 7488MB  
 System firmware current memory: 704MB Note firmware use

Partition memory usage

Partition name	Amount of memory (MB)
p5vios	1024
p5aix52	0
p5nim	1024
p5aix6a	1024
p5aix6b	1024

OK Cancel Help

## HMC versus IVM

- Integrated Virtualization Manager
  - An LPAR that acts as the HMC
  - Provides single system partitioning without a Hardware Management Console (HMC)
  - Creates LPARs
  - Manages virtual storage and virtual Ethernet
  - Eliminates the need to purchase dedicated hardware console
  - Included at **no additional charge** with purchase of optional Advanced POWER Virtualization feature or POWER Hypervisor and VIOS features.
- Limitations
  - Must own all resources in the box
  - All resources must therefore be controlled by VIOS
  - Aimed at 550 and below

Mainline: solutions you need  
from people you trust 11

## 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



Mainline: solutions you need  
from people you trust 12

## Micro-Partitioning

- Requires Advanced Power Virtualization Feature (APV)
- Mainframe inspired technology
- Virtualized resources shared by multiple partitions
- Benefits
  - Finer grained resource allocation
  - More partitions (Up to 254)
  - Higher resource utilization
- New partitioning model
  - POWER Hypervisor
  - Virtual processors
  - Fractional processor capacity partitions
  - Operating system optimized for Micro-Partitioning exploitation
  - Virtual I/O



**Mainline:** solutions you need  
from people you trust 13

## Shared processor partitions

- Micro-Partitioning allows for multiple partitions to share one physical processor
- Up to 10 partitions per physical processor
- Up to 254 partitions active at the same time
- One shared processor pool
- Dedicated processors are in the pool by default if their LPAR is powered off
- Partition's resource definition
  - Minimum, desired, and maximum values for each resource
  - Processor capacity (processor units)
  - Virtual processors
  - Capped or uncapped
    - Capacity weight
    - Uncapped can exceed entitled capacity up to number of virtual processors (VPs) or the size of the pool whichever is smaller
  - Dedicated memory
    - Minimum of 128 MB and 16 MB increments
  - Physical or virtual I/O resources

**Mainline:** solutions you need  
from people you trust 14

## Math 101 and Consolidation

- Consolidation Issues
- Math 101
  - 4 workloads
    - A 6.03
    - B 2.27
    - C 2.48
    - D 4.87
    - Total = 15.65
  - The proposed 8way is rated at 16.88
  - LPARs use dedicated processors
  - Is it big enough to run these workloads in 4 separate dedicated LPARs?
  - NO



Mainline: solutions you need from people you trust 15

## Why micropartitioning is important

- 8w 1.45g p650 is 16.88 rperf
- 2w 1.45g p650 is 4.43 rperf
- So 1w is probably 2.21
- Now back to Math 101

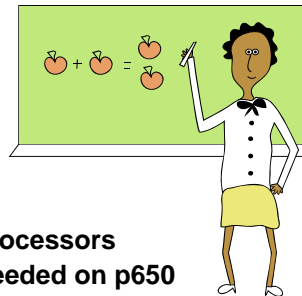
• Wkld Rperf

- A 6.03
- B 2.27
- C 2.48
- D 4.87
- Total = 15.65

Processors Needed on p650

- 3 (6.64)
- 2 (4.42 - 2.27 is > 2.21)
- 2 (4.42 - 2.48 is > 2.21)
- 3 (6.64 - 4.87 is > 4.42)
- 10 (22.12)

- Watch for granularity of workload



Mainline: solutions you need from people you trust 16

## On Micropartitioned p5 with no other Virtualization

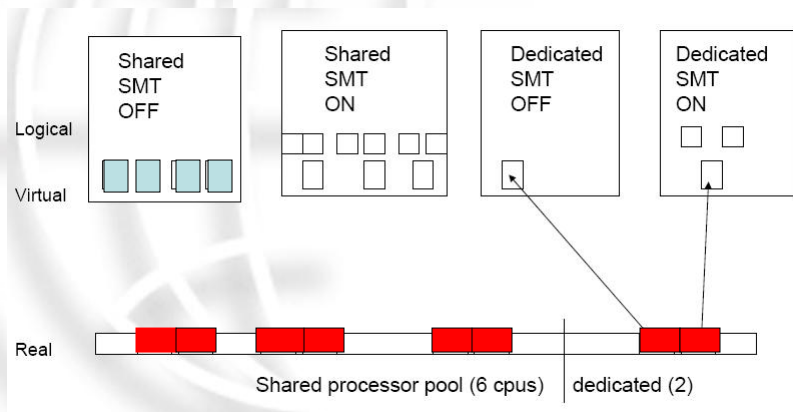
- 8w 1.45g p650 was 16.88 rperf
- 4w 1.65g p550Q is 20.25 rperf
- So 1w on 550Q is probably 5.06
  - BUT we can use 1/10 of a processor and 1/100 increments
- Now back to Math 101

Wkld	Rperf	Processors 650	Processors 550Q
A	6.03	3	1.2
B	2.27	2	.45
C	2.48	2	.49
D	4.87	3	.97
<b>Total =</b>	<b>15.65</b>	<b>10</b>	<b>3.11</b>

- Watch for granularity of workload
- On the p5 we use fewer processors and we fit!
- p6 is even better!

Mainline: solutions you need  
from people you trust 17

## Logical, Virtual or Real?

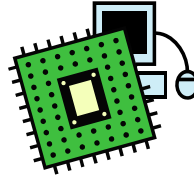


In shared world there is no one to one relationship between real and virtual processors  
The dispatch unit becomes the VP

Mainline: solutions you need  
from people you trust 18

## Defining Processors

- Minimum, desired, maximum
- Maximum is used for DLPAR
- Shared or dedicated
- For shared:
  - Capped
  - Uncapped
    - Variable capacity weight (0-255 – 128 is default)
    - Weight of 0 is capped
    - Weight is share based
    - Can exceed entitled capacity (desired PUs)
    - Cannot exceed desired VPs without a DR operation
  - Minimum, desired and maximum Virtual Processors



Mainline: solutions you need  
from people you trust 19

## Virtual Processors

- Partitions are assigned PUs (processor units)
- VPs are the whole number of concurrent operations
  - Do I want my .5 as one big processor or 5 x .1 (can run 5 threads then)?
- 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
  - Basically, how many physical processors do you want to spread your allocation across?
- 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
- Cannot exceed 10x entitlement
- VPs are dispatched to real processors
- Dispatch latency – minimum is 1 millisecond and max is 18 milliseconds
- VP Folding
- Maximum is used by DLPAR
- Use commonsense when setting max VPs!!!

Mainline: solutions you need  
from people you trust 20

## Planning for Virtual

- Virtual Ethernet
  - Included in all Power5 systems
  - Allows for in-memory connections between partitions
  - Requires AIX v5.3 or Linux
  - Up to 256 per partition
  - Does NOT require APV
- All virtual services below require the Advanced Power Virtualization feature which is included with the p590 and p595
- Virtual I/O Server (latest is v1.4)
  - Link to article on setting up a VIOS
    - <http://www-128.ibm.com/developerworks/aix/library/au-aix-vioserver-v2/>
  - Provides virtual I/O sharing for shared Ethernet and virtual SCSI partitions
  - Owns the real resources
- Shared Ethernet Adapter
- Virtual SCSI
- Excellent Redbook draft – Introduction to Advanced Power Virtualization SG24-7940

Mainline: solutions you need  
from people you trust 21

## Virtual I/O Server

- Custom AIX v5.3 partition
  - Commands have been customized
  - oem\_setup\_server
  - viostat
  - nmon works as does topas
- Provides services for:
  - Shared Ethernet Adapter
    - Built on Virtual Ethernet
  - Virtual SCSI Server
- Owns the physical resources
- Run 2 or 4 if in production
  - Maximum is around 10
- Can use multipath I/O
- Can do Etherchannels
- Maximum of 65535 virtual I/O slots
- Max of 256 VIO slots per partition



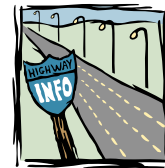
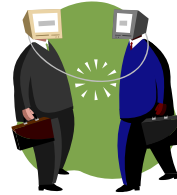
Can you have your cake and eat it?

**Do not run other workloads here**  
**But do set tunables**

Mainline: solutions you need  
from people you trust 22

# Virtual Ethernet

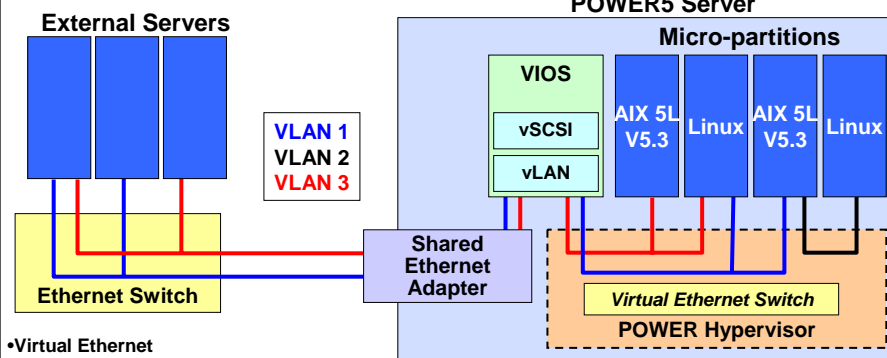
- **New Integrated Virtual Ethernet Adapter**
- Enables inter-partition communication.
  - In-memory point to point connections
- Physical network adapters are not needed.
- Similar to high-bandwidth Ethernet connections.
- Supports multiple protocols (IPv4, IPv6, and ICMP).
- No Advanced POWER Virtualization feature required.
  - POWER5 or 6 Systems
  - >= AIX 5L V5.3 or appropriate Linux level
  - Hardware management console (HMC)
- Hypervisor acts as a VE switch
- SEA is built on virtual ethernet
- Shared Ethernet Adapter
  - Used to allow partitions to share a real Ethernet adapter
  - Saves on dedicated I/O slots and adapters



**Mainline:** solutions you need from people you trust 23

# Virtual networking

Virtual Ethernet helps reduce hardware costs by sharing LAN adapters



- **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**
  - Provide ability for one adapter to be on multiple subnets
  - Provide isolation of communication to VLAN members
  - Allows a single adapter to support multiple subnets

- **IEEE VLANS**
    - Up to 4096 VLANS
    - Up to 65533 vENET adapters
    - 21 VLANS per vENET adapter
- Don't forget to plan out usage on p6 of Integrated Virtual Ethernet Adapters**

SOURCE: IBM

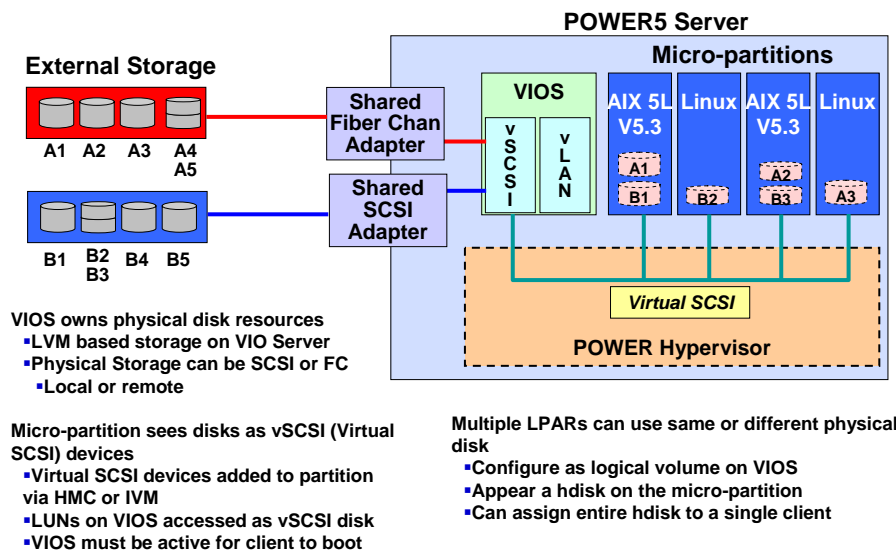
# Virtual SCSI

- Many protocols supported including: fibre channel, SCSI, SAS, iSCSI, SATA .....
- Allows sharing of storage devices
- Vital for shared processor partitions
  - Overcomes potential limit of adapter slots due to Micro-Partitioning
  - Allows the creation of logical partitions without the need for additional physical resources
  - Can save significant money due to reduction of I/O drawers
- Allows attachment of previously unsupported storage solutions
- Use for boot disks
  - With 2 x VIOS servers you now have full path redundancy during boot, something you don't get with internal physical disks
  - Disks for boot can be internal or SAN attached to the VIOS
  - It works really well and is very reliable
- Virtual SCSI Server
  - Disks on I/O server are defined as logical volumes and exported to client systems or the full volume is handed over
  - Typical rootvg size is between 20 and 30gb
  - Client systems see these as SCSI disks

**Mainline: solutions you need from people you trust** 25

## Virtual SCSI

Virtual I/O helps reduce hardware costs by sharing disk drives



SOURCE: IBM

29

## Virtualizing the DVD

- In Power6 the DVD is assigned to the same LPAR as the 6-pack in that module
- With VIOS
  - Can use mkvdev to assign DVD to an LPAR
- Without VIOS
  - 1. NFS.
  - 2. External SCSI DVD with a SCSI adapter.

Mainline: solutions you need  
from people you trust 27

## Hints and Tips

- Which LPAR is your service LPAR?
- How will you do installs
  - Allocate cd?
  - NIM?
- Backup Methodology? Especially for HMC!!
- If using virtualization planning is more critical than ever
- Change control
- Create sysplans at the HMC and save them as documentation BEFORE you need them
- Ensure Inventory scout is working on all LPARs and that VPD is being uploaded to IBM – can now use Web SSL instead of dial or VPN
- Create a partition layout in advance
  - Include devices, etc
- I/O devices are allocated at the slot level
- Are cards going in CEC or I/O drawer? Which I/O drawer?
  - Affects the number of PCI-E and PCI-X adapters
  - Boot disks –
    - I/O drawer or 2104, Raid, Fiber
    - Have VIOS control all boot disks?
    - P6 CEC – one LPAR owns all 6 disks and DVD
    - P5 CEC – 6 disks are split into 2 x 3 packs and each trio can be owned by a different LPAR
  - Check the number of LPARs per boot disk
- 64bit kernel
  - Moot point as no 32 bit in AIX v6.1 anyway

Mainline: solutions you need  
from people you trust 28

PLANNING SHEET										
Note: NIM server will use 2 onboard ethernet							Dual Port			
Instance	3.5ghz cpu	73GB Boot Disks	146GB Memory GB	Single fiber	Copper GB ether	SPP	VIO CPUS	VIO BOOT	VIO ETHER	VIO SAN
lp1	6	2	96	2	2	N	N	N	N	N
lp2	1	2	16	0		N	N	Y	Y	Y
lp3	0.9		16	0		Y	Y	Y	Y	Y
lp4	0.9		24	0		Y	Y	Y	Y	Y
NIM	0.5	2	4	1	1	Y	N	Y	Y	Y
VIO Server 1	0.5	2	4	2	2	Y	N	N	N	N
VIO Server 2	0.5	2	4	2	2	Y	N	N	N	N
OVERHEAD			4							
<b>Totals</b>	<b>10.3</b>	<b>10</b>	<b>4</b>	<b>168</b>	<b>7</b>	<b>7</b>				
<b>Total # of Adapters</b>		<b>Disks</b>	<b>14</b>		<b>14</b>					

Then have an additional spreadsheet that maps each lpar to adapters  
 And I/O drawers, etc  
 First disk on each VIO server is the VIO server rootvg, not for servicing client LPARs  
 Documentation is critical – do it ahead of time

## Planning for Memory

### PLANNING SHEET Memory

#### Overhead Calculation

	Mem	Max Mem	Mem Overhead	Divide by 256	Round Up	New Overhead
lp1	98304	102400	1600	6.25	7	1792
lp2	16384	20480	320	1.25	2	512
lp3	16384	20480	320	1.25	2	512
lp4	24576	28672	448	1.75	2	512
NIM	4096	8192	128	0.5	1	256
VIO Server 1	4096	8192	128	0.5	1	256
VIO Server 2	4096	8192	128	0.5	1	256
Hypervisor						256
TCEs for drawers, etc?						512
<b>TOTAL Overhead</b>						<b>4864</b>

This gives a rough estimate  
 Assumes LMB size is 256

**Mainline:** solutions you need  
 from people you trust

### More Sizing - original

Server	LPAR	CPUs	Memory GB	rPerf	AIX	KVA	Watts	BTU/hr
p630 1.45ghz 4 way	LPAR1	2	4	4.41	5.2	0.78	750	2450
	LPAR2	2	4	4.41	5.2			
p630 1.45ghz 4 way	LPAR3	2	4	4.41	5.2	0.78	750	2450
	LPAR4	2	4	4.41	5.2			
p630 1.45ghz 4 way	LPAR5	2	8	4.41	5.2	0.78	750	2450
	LPAR6	2	8	4.41	5.2			
p630 1.45ghz 4 way	LPAR7	2	4	4.41	5.2	0.78	750	2450
	LPAR8	2	4	4.41	5.2			
p650 1.45ghz 8 way	LPAR9	2	4	4.47	5.3	1.7	1600	5461
	LPAR10	2	4	4.47	5.3			
	LPAR11	2	4	4.47	5.3			
	LPAR12	2	4	4.47	5.2			
p650 1.45ghz 8 way	LPAR13	4	16	9.12	5.3	1.7	1600	5461
	LPAR14	4	16	9.12	5.2			
7 x I/O drawers						2.52	2380	8127
		<b>32</b>	<b>88</b>	<b>71.4</b>		<b>9.04</b>	<b>8580</b>	<b>28849</b>

rPerf taken from rsperfs021406.doc from IBM

### Micropartioning Only

So, let's assume we can migrate everything to AIX v5.3 and take advantage of micro-partitioning. Let's also assume we have policies in place as follows:

Per LPAR

2 x fiber cards

2 x internal boot disks

2 x 10/100/1000 Ethernet cards

4gb memory per processor if database

2gb memory per processor for other

In the unconsolidated environment above we would have the following:

Disks 28  
 Fiber cards 28  
 Ethernet cards 28  
 Memory 88GB + 8% overhead = 96GB  
 Processors 32 x 1.65ghz  
 rPerf 71.4

If we moved this to a purely AIX v5.3 environment (no VIOS) on the new p5+-570 with 1.9ghz processors using shared processors we would require the following:

Disks 28  
 Fiber cards 28  
 Ethernet cards 28  
 Memory 88GB + 8% overhead = 96GB  
 Processors 14 x 1.9ghz  
 rPerf 75.88

NB rPerf for p5 assumes SMT is turned on. If running 5.2 or without SMT then divide the number by 1.3

Given the architecture of the p5-570 we would have the ability to boot 8 LPARs of our total 14 from 3-packs in the CECs. So we would need additional I/O drawers for the other 6. The 570 also provides only 24 PCI-X slots total so an additional 32 slots would be needed in I/O drawers. A 7311-D20 I/O drawer provides 7 PCI-X slots and 2 x 6-packs for disk. So we would need a total of 5 x I/O drawers to support the necessary cards. 3 of the I/O drawers would also each have 4 disks (2 per 6-pack) to support the final 6 lpar boot disks

Environmentals on the new configuration would look like this:

Servers	KVA	Watts	BTU/HR
OLD	9.04	8,580	28,849
570 16way 1.9g	5.472	5,200	17,748
5x7311-D20	1.8	1,700	5,805
<b>TOTAL</b>	<b>7.272</b>	<b>6,900</b>	<b>23,553</b>
<b>SAVINGS</b>	<b>1.768</b>	<b>1,680</b>	<b>5,296</b>

### Adding 2 VIOS

LPAR	Fiber	Ethernet	Disks	Uses	Disk	Fiber	Ethernet
	OLD	Old		VIO?	NEW	NEW	NEW
LPAR1		2	2	2 y		0	0
LPAR2		2	2	2 y		0	0
LPAR3		2	2	2 y		0	0
LPAR4		2	2	2 y		0	0
LPAR5		2	2	2 y		0	0
LPAR6		2	2	2 y		0	0
LPAR7		2	2	2 y		0	0
LPAR8		2	2	2 y		0	0
LPAR9		2	2	2 n		2	2
LPAR10		2	2	2 n		2	2
LPAR11		2	2	2 n		2	2
LPAR12		2	2	2 n		2	2
LPAR13		2	2	2 n		2	2
LPAR14		2	2	2 n		2	2
2 x VIOS (uses 300gb)		0	0	0 y		4	6
		<b>28</b>	<b>28</b>	<b>28</b>		<b>16</b>	<b>18</b>
<b>Savings over old</b>		<b>12</b>	<b>10</b>	<b>10</b>		Saves 20 slots and can Reduce by 2 I/O drawers	

Note assumes 2 VIO servers booting from internal disk and providing client boot disks from SAN  
Some LPARs are still booting from their own disks and have their own devices

## General Server Sizing thoughts

- Correct amount of processor power
- Balanced memory, processor and I/O
- Min, desired and max settings and their effect on system overhead
- Memory overhead for page tables, TCE, etc
- Shared or dedicated processors
- Capped or uncapped
- If uncapped – number of virtual processors
- Expect to safely support 3 LPARs booting from a 146gb disk through a VIO server
- Don't forget to add disk for LPAR data for clients

Mainline: solutions you need  
from people you trust 35

## VIOS Sizing thoughts

- Correct amount of processor power and memory
- Shared uncapped processors
- Capped or uncapped
- Number of virtual processors
- Higher weight than other LPARs
- Expect to safely support 3 LPARs booting from a 146gb disk through a VIO server
- Don't forget to add disk for LPAR data for clients
- Should I run 2 or 4 x VIOS?
  - Max is somewhere around 10
- Virtual I/O Server Sizing Guidelines Whitepaper
  - <http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/perf.html>
  - Covers for ethernet:
    - Proper sizing of the Virtual I/O server
    - Threading or non-threading of the Shared Ethernet
    - Separate micro-partitions for the Virtual I/O server

Mainline: solutions you need  
from people you trust 36

# Ethernet

- [http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en\\_US/info/iphb1/iphb1\\_vios\\_planning\\_network.htm](http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphb1/iphb1_vios_planning_network.htm)
  - for information on VLAN benchmark results with different processing power and mtu sizes
- Short summary
  - MTU=1500 needs 1 x 1.65ghz processor per gigabit adapter to drive the adapter fully
  - MTU=9000 needs about half a processor
  - Has calculators in there for sizing for SEA
  - #cpus = (Throughput (bytes) X MTU cycles/byte) / cpu speed
    - i.e.  $200 \times 1024 \times 1024 \times 11.2 = 2,348,810,240$  cycles /  $1,650,000,000 = 1.42$  processors
    - This is for 200 MB streaming throughput on a 1.65ghz processor with a 1500 MTU
  - 512mb memory is usually fine if only SEA as buffers and data structures are fairly small
  - Dedicated receive buffers
    - Physical ethernet adapter uses 4MB for MTU 1500 and 16MB for MTU 9000
    - Virtual Ethernet uses around 6MB
  - I usually allocate 2gb if it is SEA only just to be safe
- Other Memory Needs (P6)
  - Each active IVE ethernet port 102MB
  - So quadport on base is 408MB overhead

Mainline: solutions you need  
from people you trust 37

# Virtual SCSI

- [http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en\\_US/info/iphb1/iphb1\\_vios\\_planning\\_vscsi.htm](http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphb1/iphb1_vios_planning_vscsi.htm)
- Short summary
  - A little extra latency using virtual scsi (.03 to .06 millisec per I/O) regardless of LV or physical disks being assigned to the client
  - Latency is dependent on blocksize
  - Memory allocation is critical
    - For large I/O configurations and high data rates VIOS may need 1gb
  - Processor allocation is critical
    - Based on IOPS, Blocksize and processor speed
    - i.e. using IBM's example in the page above:
      - 2 clients, each with full disks allocated to them rather than LVs
      - One has 10,000 IOPS size 32KB (CPU cycles for 32KB are 58000)
      - One has 5,0000 IOPS size 64KB (CPU cycles for 64KB are 81000)
      - Processor is 1.65GHZ
      - $CPUS = (\#IOPS \times Cycles) / CPU \text{ speed}$
      - $((10,000 * 58,000) + (5,000 \times 81,000)) / 1,650,000,000 = 0.60$  CPUs

Mainline: solutions you need  
from people you trust 38

Approximate cycles per second on a 1.65 Ghz partition

Disk type	4 KB	8 KB	32 KB	64 KB	128 KB
<i>Physical disk</i>	45,000	47,000	58,000	81,000	120,000
<i>Logical volume</i>	49,000	51,000	59,000	74,000	105,000

SOURCE: IBM – see previous slides for link

Mainline: solutions you need  
from people you trust 39

## Virtual SCSI Bandwidth

Table 1. Physical and Virtual SCSI bandwidth comparison (in MB/s)

I/O type	4 K	8 K	32 K	64 K	128 K
<i>Virtual</i>	20.3	35.4	82.6	106.8	124.5
<i>Physical</i>	24.3	41.7	90.6	114.6	132.6

In the tests, a single thread operates sequentially on a constant file that is 256 MB in size with a Virtual I/O Server running in a dedicated partition.

SOURCE: IBM – see previous slides for link

Mainline: solutions you need  
from people you trust 40

# Sysplans and SPT

- System Planning Tool
  - <http://www-03.ibm.com/servers/eserver/support/tools/systemplanningtool/>
- Sysplans on HMC
  - Can generate a sysplan on the HMC
  - Print it to PDF and you are now fully documented
  - Cannot read this into SPT for some reason
  - Bug in v7 MH01045 causes error on creation attempt for Sysplan
    - Still the case with SP1 if trying to get a sysplan for an LPAR on a system with a VIOS (checked this Tuesday night)
- Solutions Assurances
  - These protect you from errors in the configurations
  - They are required for the p5-570 and above
  - Request one even for smaller boxes to ensure no power, etc problems

# SPT

The screenshot displays the IBM System Planning Tool (SPT) interface. At the top, it shows the system plan as 'p6 570 plan' and the system as 'System 01 (IBM System p 9117-MMA)'. The main window is titled 'Partitions' and shows a 'Memory' tab. The memory configuration is as follows:

Memory	
System memory (MB):	65536
Configured memory (MB):	57088
Hypervisor memory (MB):	3072
Unassigned memory (MB):	5376
Logical memory block size (MB):	256

Memory for Partitions					
Name	ID	Operating System	Min	Desired	Max
LPAR1	1	AIX_53	2048	16128	19968
LPAR2	2	AIX_53	2048	8192	16128
LPAR3	3	AIX_53	2048	8192	9984
LPAR4	4	AIX_53	2048	8192	9984
LPAR5	5	AIX_53	2048	8192	9984
LPAR6	6	AIX_53	2048	4096	8192
VIOSA	7	Virtual I/O Server	1024	2048	3072
VIOSB	8	Virtual I/O Server	1024	2048	3072

At the bottom of the window, there are buttons for 'OK', 'Apply', 'Cancel', 'Report', and 'Help'.

# SPT

Partitions Summary										
Processors										
Partition	OS Version	Shared	Processor Minimum	Processor Desired	Processor Maximum	Virtual Processor Minimum	Virtual Processor Desired	Virtual Processor Maximum	Uncapped	Weight
LPAR1	5.3	Y	0.5	4.0	10.0	1	8	10	Y	128
LPAR2	5.3	Y	0.5	3.0	6.0	1	6	6	Y	128
LPAR3	5.3	Y	0.5	2.0	4.0	1	4	4	Y	128
LPAR4	5.3	Y	0.5	2.0	4.0	1	4	4	Y	128
LPAR5	5.3	Y	0.5	2.0	4.0	1	4	4	Y	128
LPAR6	5.3	Y	0.5	1.0	2.0	1	2	2	Y	128
VIOSA		Y	0.1	1.0	2.0	1	2	2	Y	128
VIOSE		Y	0.1	1.0	2.0	1	2	2	Y	128

Memory										
Partition	OS Version	Virtual Memory Minimum	Virtual Memory Desired	Virtual Memory Maximum	Virtual Serial	Virtual LAN	Client SCSI	Client Server	Total	
LPAR1	5.3	2048	16128	19968	2	2	2	0	16	
LPAR2	5.3	2048	8192	16128	2	2	2	0	16	
LPAR3	5.3	2048	8192	9984	2	2	2	0	16	
LPAR4	5.3	2048	8192	9984	2	2	2	0	16	
LPAR5	5.3	2048	8192	9984	2	2	2	0	16	
LPAR6	5.3	2048	4096	8192	2	2	2	0	16	
VIOSA		1024	2048	3072	2	1	0	6	25	
VIOSE		1024	2048	3072	2	1	0	6	25	

OS Details					
Partition	OS Version	Interactive Percent	Primary Console	Alternate Console	Load Source
LPAR1	5.3				
LPAR2	5.3				
LPAR3	5.3				
LPAR4	5.3				
LPAR5	5.3				
LPAR6	5.3				
VIOSA					
VIOSE					

## Maximums

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



**Mainline:** solutions you need from people you trust

## LPAR Notes Power5

- Max LPARS = 10x #processors
- Only Power5 or 5+ on the same HMC if at v6
- Current 7310 HMC can be upgraded to v7 to add Power6 support
- All use 19" racks but very specific power needs
- p5-520 and p5-550
  - Only 7311-D20 I/O drawer supported
  - Comes with 1 x 4pack internal – another can be added
- P5-560Q
  - Similar to p5-570 but no I/O drawers, uses quadcore and only 2 modules
- p5-570 (each 4-way)
  - 2 x 3 packs of disk internal (on separate SCSI buses)
  - Supports 2 x LPARs without extra SCSI cards
  - 7311-D20 I/O drawer has 2 x 6 packs – SCSI cards needed and must be in slots 4 and 7
- P5s Require AIX v5.2 ML4 at a minimum

Mainline: solutions you need  
from people you trust 45

## LPAR Notes Power6

- Max LPARS = 10x #processors
- Power5, 5+ and 6 on the same HMC (7310 v7 or 7402)
  - 7315 HMC not supported
  - Requires HMC code v7
  - Can use new Web interface
- All use 19" racks but very specific power needs
- p6-570 (each 4-way)
  - 1 x 6 packs of disk internal (on one SCSI bus per six-pack)
  - 1 x media bay per
  - Supports 1 x LPAR without extra SCSI cards
  - Supports 7311-D20, 7311-D11 RIO drawers but you have to give up a PCI-E slot
    - 7311-D20 I/O drawer has 2 x 6 packs – SCSI cards needed and must be in slots 4 and 7
  - Supports new G03 drawer without using a PCI-E slot
  - To get redundancy you must have 3 x FC 5625 power regulators per CEC
- P6 requires AIX v5.2 TL10 or AIX v5.3 TL06 or an enabled version of Linux

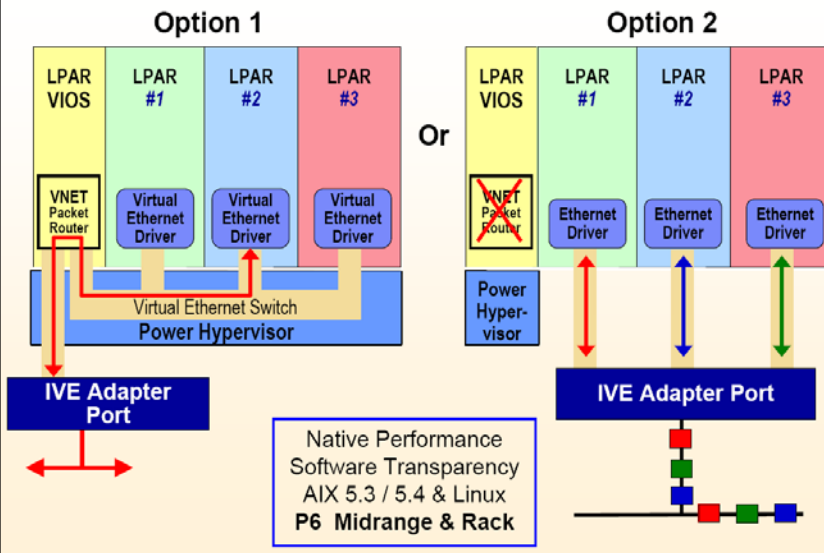
Mainline: solutions you need  
from people you trust 46

# Upgrading p5 to p6 - gotchas

- Each P5-570 module has 6 x PCI-X slots. Each p6-570 module has 4 x PCI-E and 2 x PCI-X slots.
  - PCI-X cards may have to go in an I/O drawer
  - Can replace them with PCI-E adapters
- 3 I/O drawers supported
  - 7314-G30 (new drawer with 6 x PCI-X slots)
  - 7311-D11 (6 x PCI-X slots per side, requires RIO card)
  - 7311-D20 (7 x PCI-X slots and 2 x disk 6-packs, requires RIO)
- P6-570 has 2 x GX slots but no integrated RIO
- Disks in p6 are SAS so SCSI disks will need to move to an I/O drawer
- Not all adapters are supported in a drawer when attached to a p6-570
- All processors must migrate and DDR1 memory does NOT migrate
- P6-570 has 1 6-pack per module, not 2 x 3-packs like the p5-570
- DVD is on same adapter as 6-pack
- Use IVE to reduce number of ethernet adapters
- Must upgrade 7310 HMC to v7 (via v6) to support Power6
  - V7 is ordered as a feature code on the HMC (0962)
- Specific microcode, VIOS, HMC and AIX levels needed
- PLM not included in APV on Power6
- Redundant service processor in the second CEC but not active till firmware update (SOD 4Q07)

Mainline: solutions you need from people you trust 47

## Integrated Virtual Ethernet How it works.....



SOURCE: IBM

Mainline: solutions you need from people you trust 48

## IVE Notes

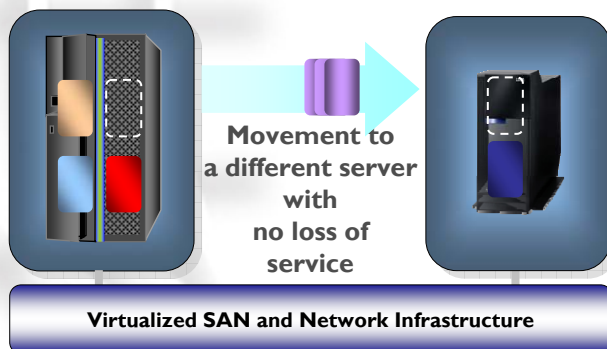
- Which adapters do you want? Each CEC requires one.
- Adapter ties directly into GX Bus
  - No Hot Swap
  - No Swap Out for Different Port Types (10GbE, etc.)
- Not Supported for Partition Mobility, except when assigned to VIOS
  - Option 1 on Previous Chart
- Partition performance is the same as a real adapter
  - No VIOS Overhead
  - Intra-partition performance may be better than using Virtual Ethernet
- Usage of serial ports on IVE
  - Same restrictions as use of serial ports that were on planar on p5
  - Once an HMC is attached these become unusable
- Naming
  - Integrated Virtual Ethernet – Name used by marketing
  - Host Ethernet Adapter (HEA) Name used on user interfaces and documentation

Mainline: solutions you need  
from people you trust 49

## Planned Live Partition Mobility with POWER6

### Allows migration of a running LPAR to another physical server

- ✓ Reduce impact of planned outages
- ✓ Relocate workloads to enable growth
- ✓ Provision new technology with no disruption to service
- ✓ Save energy by moving workloads off underutilized servers



SOURCE: IBM

Mainline: solutions you need  
from people you trust 50

## Partition Mobility Pre-Reqs

- All Systems in a Migration Set must be managed by the same HMC
  - HMC will have orchestration code to control migration function
- All Systems in a Migration Set must be on the same subnet.
- All Systems in a Migration Set must be SAN connected to shared physical disk – no VIOS LVM-based disks.
- ALL I/O must be shared/virtualized. Any dedicated I/O adapters must be deallocated prior to migration.
- Systems must be firmware compatible (within one release)

Mainline: solutions you need  
from people you trust 51

## Partition Mobility – Other Considerations

- Intended Use:
  - Workload Consolidation
  - Workload Balancing
  - Workload Migration to Newer Systems
  - Planned CEC outages for maintenance
  - Unplanned CEC outages where error conditions are picked up ahead of time.
- What it is not:
- A Replacement for HACMP or other clustering.
  - Not automatic
  - LPARs cannot be migrated from failed CECs
  - Failed OS's cannot be migrated
- Long Distance Support Not Available in First Release

Mainline: solutions you need  
from people you trust 52

# 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

nmon & seastat

Mainline: solutions you need  
from people you trust 53

# Traps for Young Players

- Under-sizing VIOS
- Over-committing boot disks
- Forgetting Memory and processor Overhead
- Planning for what should and should not be virtualized
- Misunderstanding needs
- Workload Granularity
- Undersizing memory and overhead
  - Hypervisor
  - I/O drawers, etc
  - VIOS requirements
  - Setting maximums
- Chargeback and capacity planning may need to be changed



Mainline: solutions you need  
from people you trust 54

## References



- IBM Redbooks
  - <http://www.redbooks.ibm.com>
    - These are updated regularly
  - 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
  - SG24-7349 – Virtualization and Clustering Best Practices
  - Red piece 4194 – Advanced Power Virtualization Best Practices
  - Red piece 4224 – APV VIOS Deployment Examples
  - The Complete Partitioning Guide for IBM eServer System P Servers
  - System P – 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
- eServer Magazine
  - <http://www.eservercomputing.com>
    - Several articles focused on Virtualization
- Find more on Mainline at:
  - <http://www.mainline.com>

**Mainline:** solutions you need  
from people you trust 55

## Questions?



**Mainline:** solutions you need  
from people you trust 56

# Backup Slides

**Mainline:** solutions you need  
from people you trust 57

## Supported Operating Systems

- AIX 6.1
  - SOD for 4Q07
- AIX 5.3
  - Enables Virtualization when on Power5
  - TL06 required for Power6 – latest version is TL06-03
- AIX 5.2
  - Minimum of ML4 required for Power5
  - TL10 required for Power6 – latest version is TL10-02
- AIX 5.1
  - Will not run on Power5 or Power6 systems
- Suse Linux, Redhat EL AS
  - The Linux 2.6 kernel versions provide full virtualization support except for DLPAR memory
- Check required ML levels for each box
- Check required microcode levels on HMC, System P boxes and cards, especially fiber cards
- HMC must be at v7.3.1.0 – latest fixpack is MH01045 which seems to break sysplan
  - NOTE new web interface at v7

**Mainline:** solutions you need  
from people you trust 58

## POWER6 and AIX 6 new function

Feature	Licensed Via		Supported OS			Supported Hardware			GA Date
	APV	AIX v6.1	AIX v5.3	AIX v6.1	Linux	POWER4	POWER5	POWER6	
Dedicated processor sharing	✓		✓	✓	✓			✓	6/07
Hardware Decimal FP			✓	✓	✓			✓	6/07
Integrated Virtual Ethernet			✓	✓	✓			✓	6/07
Storage keys - application			✓	✓				✓	6/07
Storage keys – kernel				✓				✓	4Q07
Live Partition Mobility	✓		✓	✓	✓			✓	4Q07
Multiple virtual shared pools	✓		✓	✓	✓			✓	4Q07
WPARs		✓		✓		✓	✓	✓	4Q07
Live Application Mobility		✓		✓		✓	✓	✓	4Q07

SOURCE: IBM

## Software

- Make sure HMC and all boxes are at the latest microcode level
  - Double-check supported HMC, VIOS, AIX and firmware matrix
  - <http://www14.software.ibm.com/webapp/set2/sas/f/power5cm/supportedcode.html>
- System P Microcode can be found at:
  - <http://www14.software.ibm.com/webapp/set2/firmware/gjsn>
  - Latest p5 microcode is SF240-320 as of May 29, 2007
  - P6 microcode level will be announced shortly
- HMC Corrective Service can be found at:
  - <https://www14.software.ibm.com/webapp/set2/sas/f/hmc/home.html>
- Latest HMC Software version (as of September 3, 2007) is
  - Power 4/4+ - v3v3.7 U810911
  - Power5
    - V4.R5.0 MH00929
    - v5.R2.1 MH00891
    - v6.R1.2 MH01031
    - V7.3.1.0 MH01045
  - Power6 (and 5 and 5+)
    - V7.3.1.0 MH01045
      - With v7 you can use a direct web interface instead of WebSM
      - NO support for 7315
- Don't forget BIOS updates which are at the HMC locations above
- As of March 2004 HMC maintenance is now a customer responsibility.
- Planning
  - Need to keep HMC software and BIOS up to date
  - Backups of the HMC
  - Server and card microcode/firmware updates
  - AIX maintenance planning
- All links valid as of 5/29/07

## POWER6 Rollout – Other Random Items

- AIX Version 6.1 Open Beta Starts July/August 2007
  - Open to everyone
- P5 / p5+ Upgrades to POWER6 have a lot of rules/regulations and varying dates of availability.
- Note new Service Strategies for both AIX (5.3 TL06) and Firmware
  - Starting with TL06 – support will last for 2 years on single TL.
- Service Agent is **required** on POWER6 systems and will be part of CE Installation unless the option to not do so is chosen as part of the configuration
- Look at Tivoli and IBM Director for Comprehensive Systems Management
  - Performance and Capacity – ITM
  - Keeping Track of Everything – TADDM / CCMDB

**Mainline:** solutions you need  
from people you trust 61