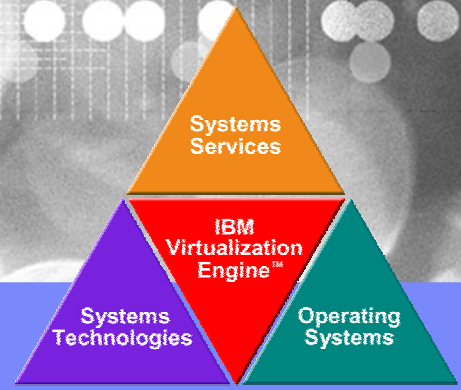
IBM EMEA ATS PSSC

IBM pSeries Power5 Advanced Power Virtualization

Jean-Armand Broyelle

IT Specialist,
EMEA system p5 benchmark center

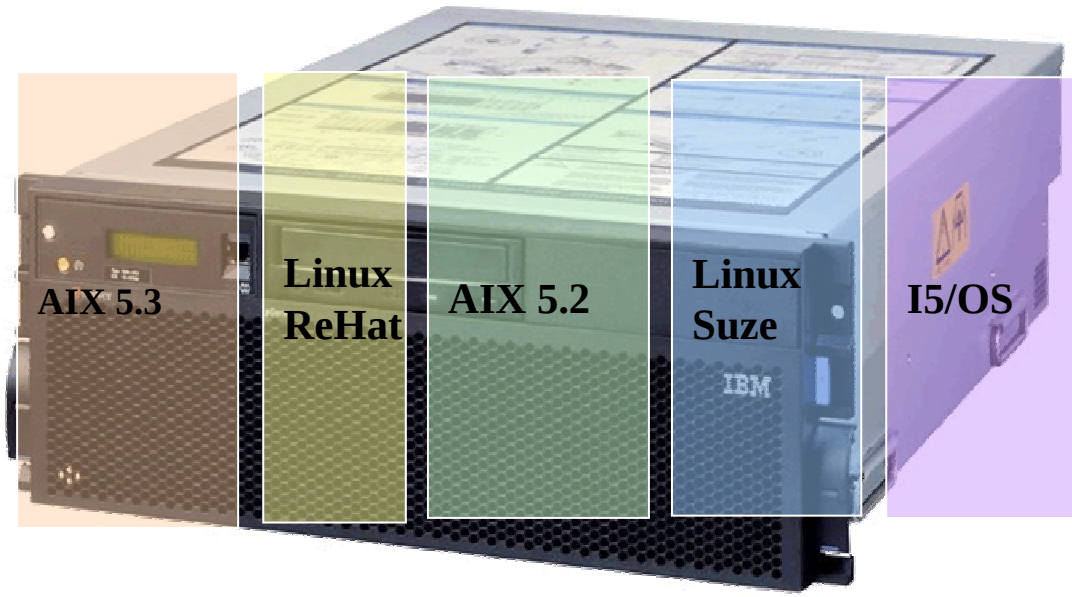


Agenda

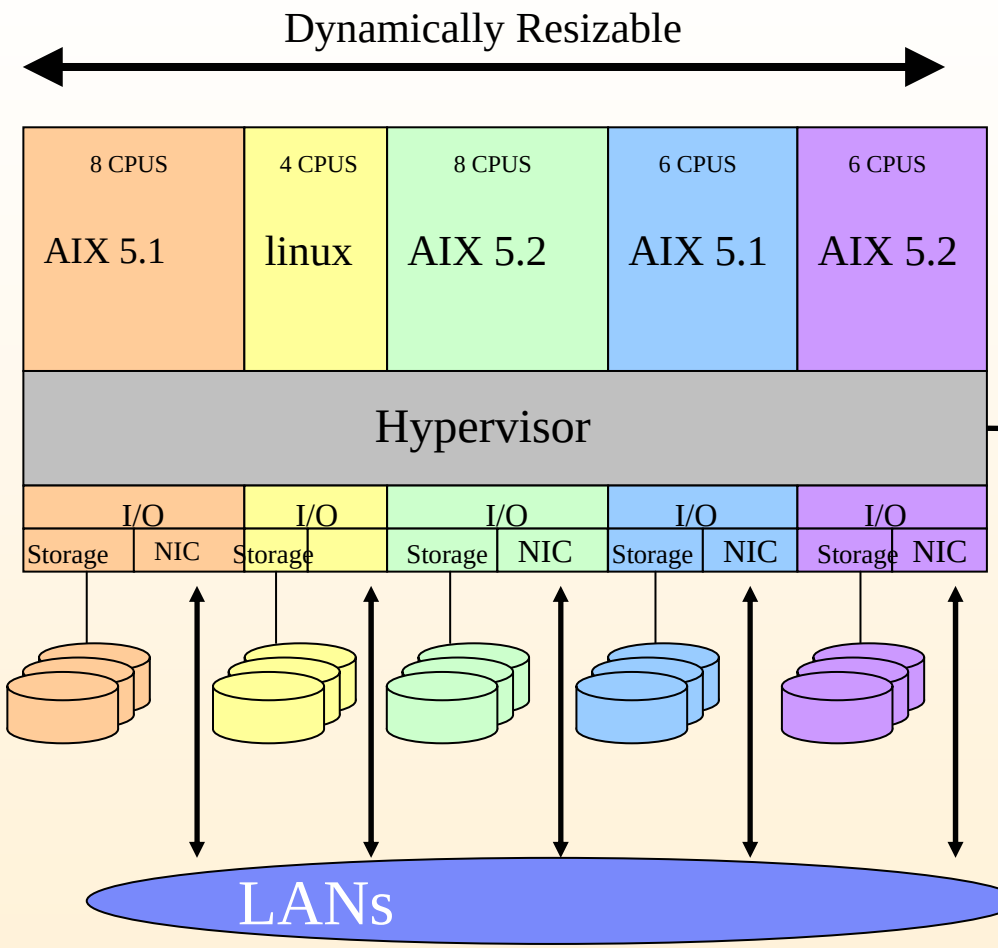
- POWER4 virtualisation reminder
- Adanced Power Virtualisation option
  - Shared processor LPAR (micropartition)
  - Virtual I/O
    - Virtual Ethernet Adapter
    - Virtual SCSI
  - Partition Load Manager
- Customer real life experiences
- Roadmap

Logical Partition : definition

Method of taking a single server and carving it up into multiple logical partitions each isolated from one another and each able to run a different OS



POWER4 Logical Partition : reminder



Hardware Management Console (**HMC**)

LPAR makes it possible to run **multiple, independent operating system images** of AIX 5L, Linux on a **single** pSeries server.

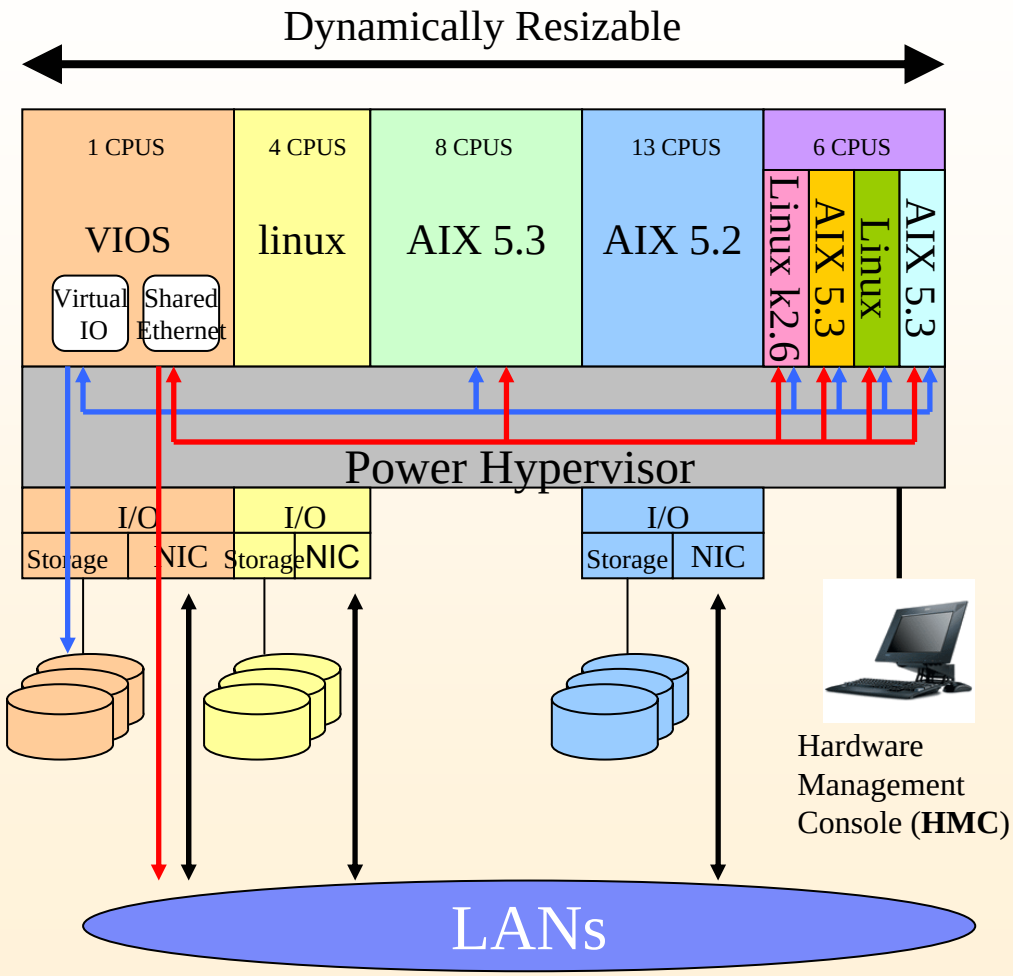
Dynamic Logical Partitions (Pwr4/AIX5.2)

- AIX5.2 Partition do not have to be rebooted to move ressources
- Hypervisor provided abstraction layer in hardware
- No ressources shared between partitions : ressources are **dedicated**

Resource granularity

- 1CPU per LPAR
- Single PCI I/O adapter
- Up to 32 Partitions with AIX 5.2 & Power 4

@server p5: advanced POWER virtualization option



Increase Physical resource utilization thru virtualization of processors, memory, network and disk ressources

Increased number of LPARS

- AIX support for 64 “dedicated” LPAR
- Virtual Ethernet-LPARs can communicate without having to use a physical I/O adapter
- APV :

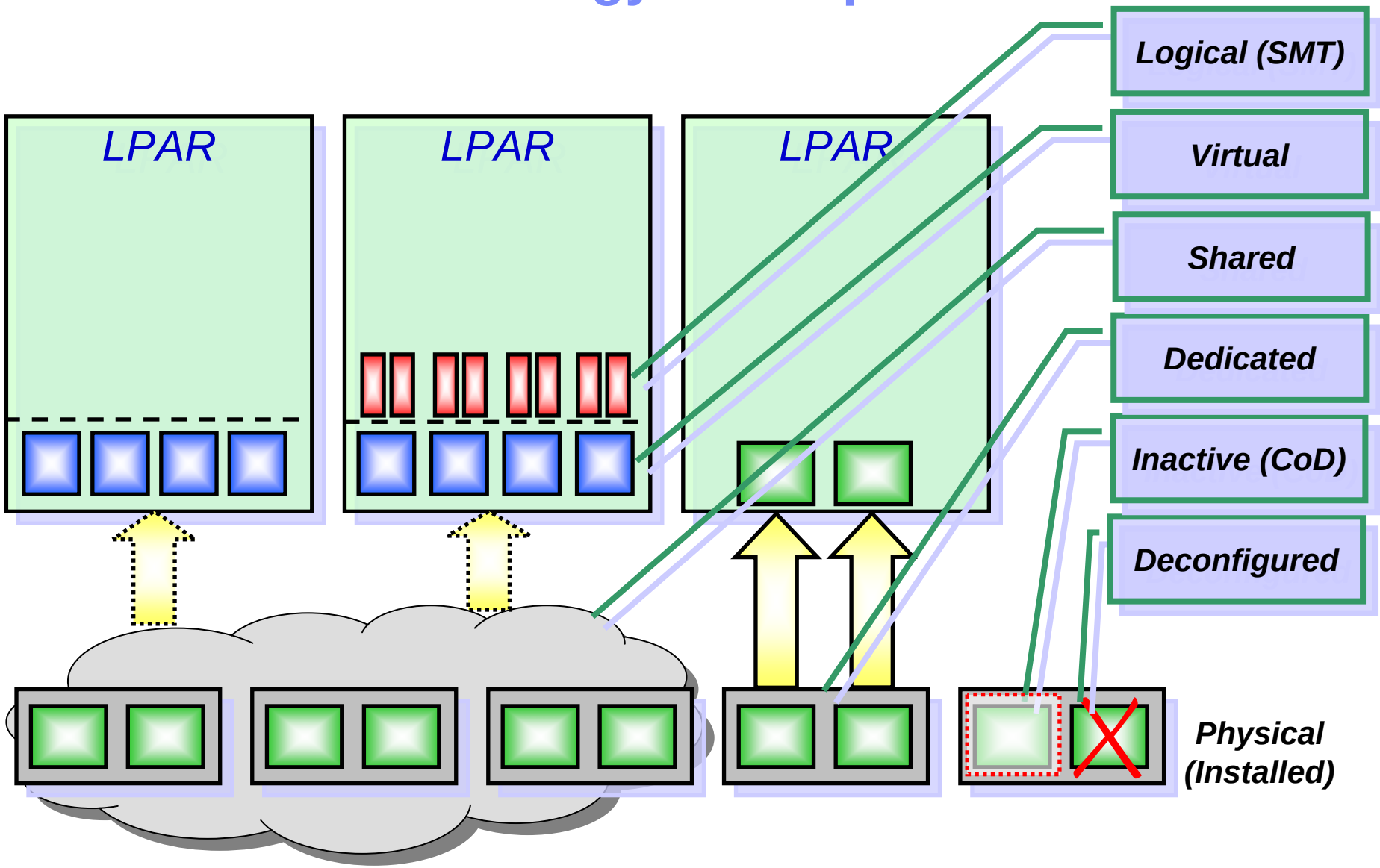
Micro-partition

- A single processor might be shared by up to 10 partitions
- Support for up to 254 partitions
- Requires Power5 and new Power Hypervisor

Virtual inter-partition Ethernet : Ethernet Sharing – LPARs can share external network connection

Virtual I/O disk : Client partitions can use logical disks hosted by another partition

Processor Terminology Concepts



Micro-Partitioning: definition

Min, Max and Desired processing units of capacity (Capacity Entitlement) :

- Processing capacity can be configured in fractions of 1/100 of a processor.
- The minimum amount or processing capacity which has to be assigned to a partition is 1/10 of a processor.

Min, Max and Desired number of Virtual processors : the whole number of concurrent operations that the operating system can use.

Capped and uncapped mode

- capped mode: The processor unit never exceeds the assigned processing capacity.
- uncapped mode: The processing capacity may be exceeded when the shared processing pool has spare processing power. When a partition is running an uncapped mode you have to specify the uncapped weight of that partition.

Memory and IO slots (physical and virtual): same as DLPAR

Micro-partition reminder through an example

- Hypothesis: 3 processors in the shared processor pool

  Capacity of the processor pool = **3.00**

- A micro-partition **Capacity Entitlement** is a guaranteed part of the shared processor pool capacity

  so **CE** is within 0.10 and 3.00

  The **sum of all CE** of the active micro-partitions is less than 3.00 (pool capacity)

- A micro-partition uses processor capacity of the shared processor pool through **virtual processors**

  Usual way to handle “processor execution concept” in all OS

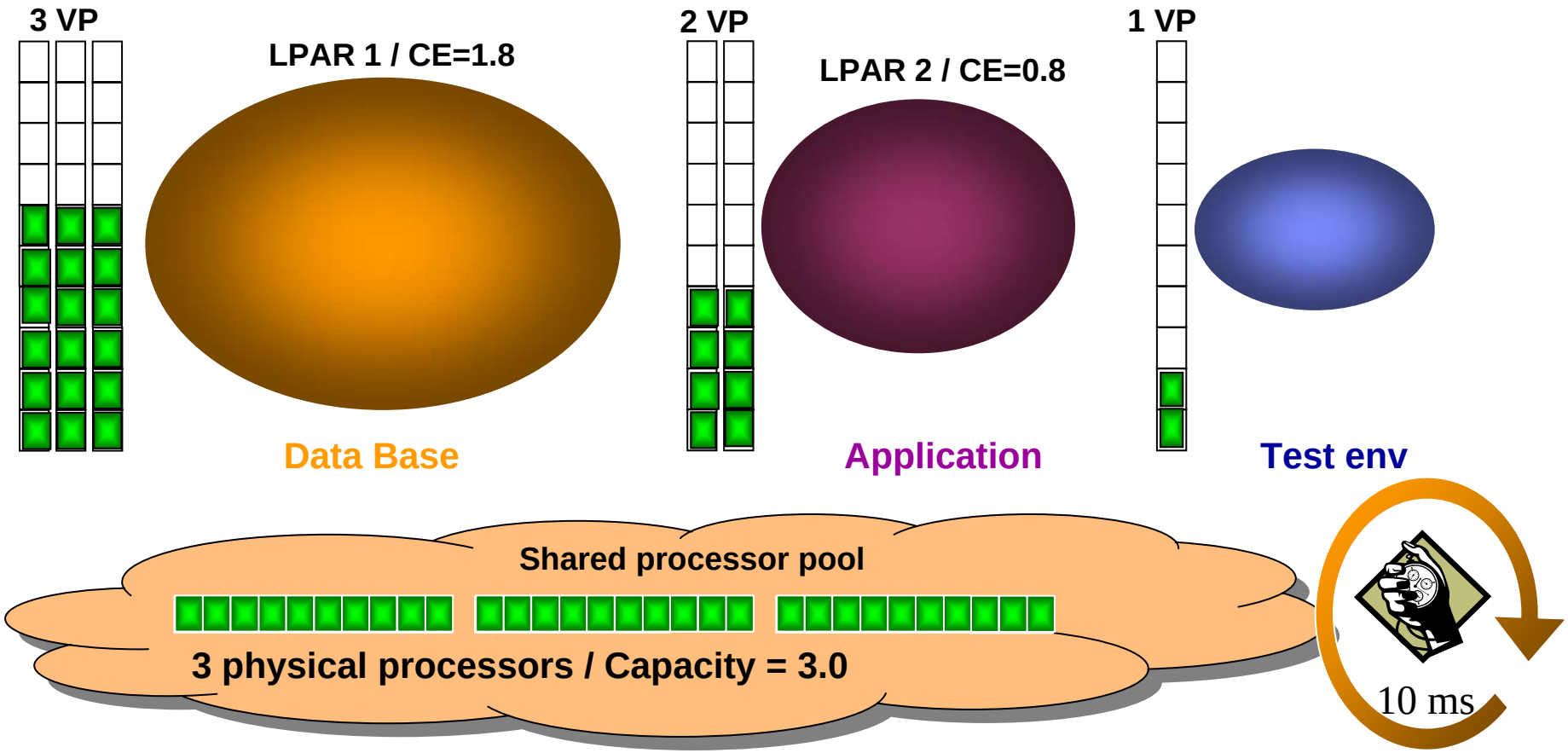
  Virtual processors are Physical CPU time slices

Micro-partitioning : CE example

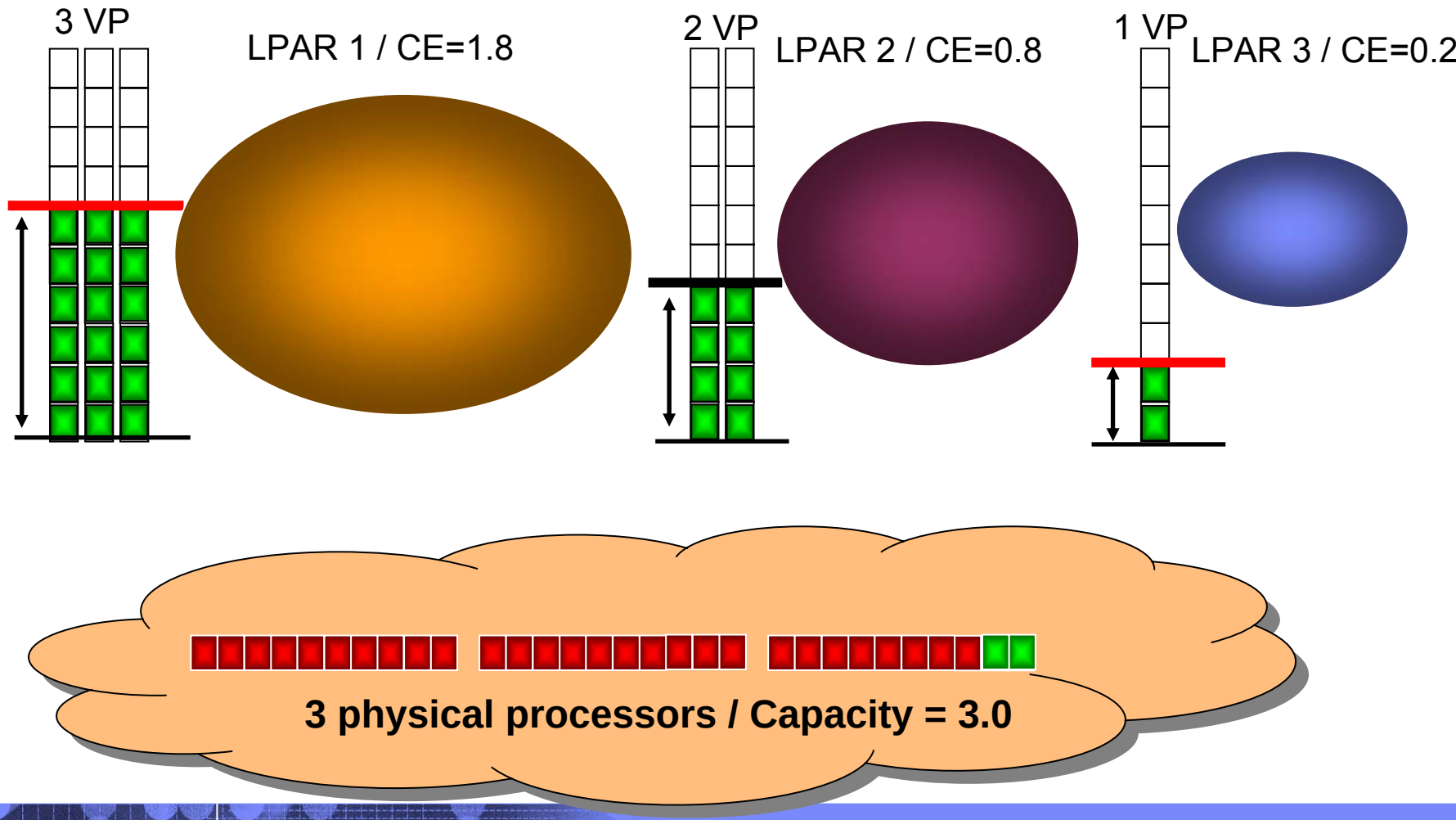
3 processors in the shared processor pool - pool **Capacity** = 3.00 (3x10x0,1)

Partition 1 : **Data Base**	CE=1.80, Virtual Proc = 3 (0,60 per processor)
Partition 2 : **Application**	CE=0.80, Virtual Proc = 2 (0,40 per processor)
Partition 3 : **Test environment**	CE=0.20, Virtual Proc = 1 (0,20 per processor)

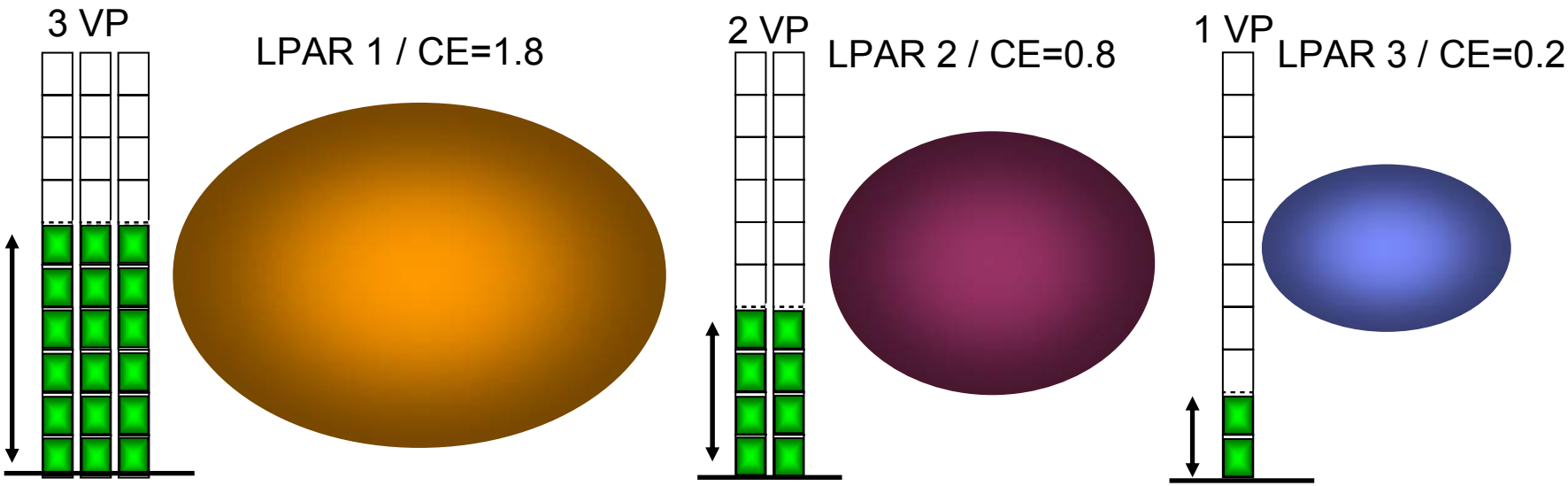
Total CE= 2.80, Total Virtual Proc = 6 (0.20 remaining in the pool)



Micro-partitions: Capped mode

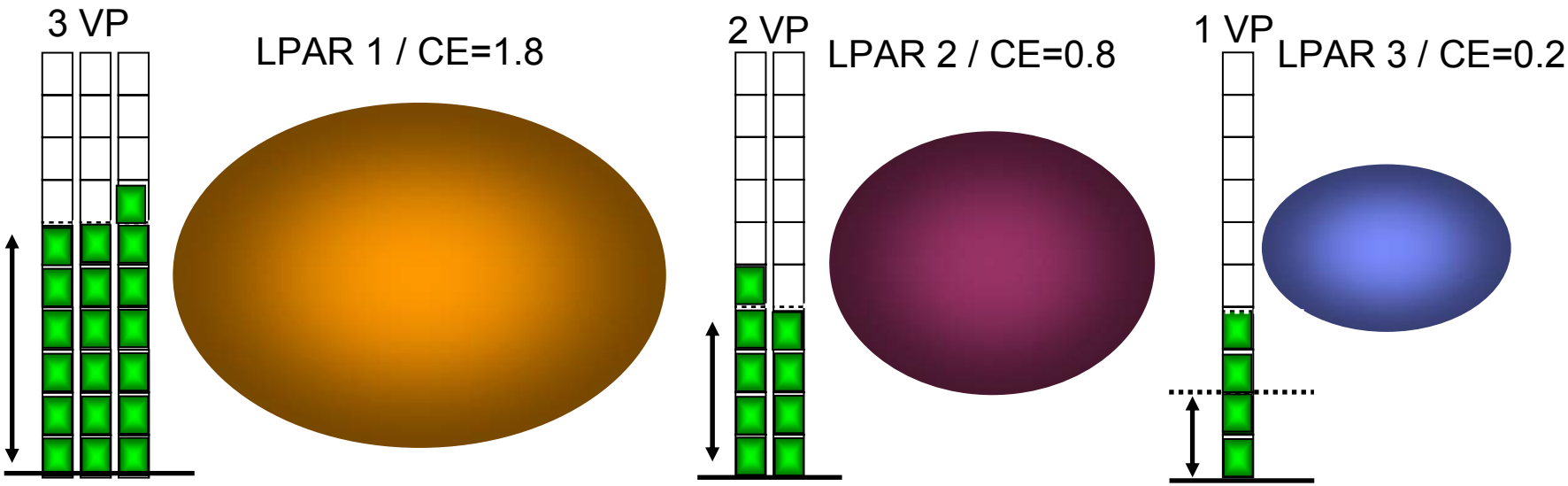


Micro-partitions: Uncapped

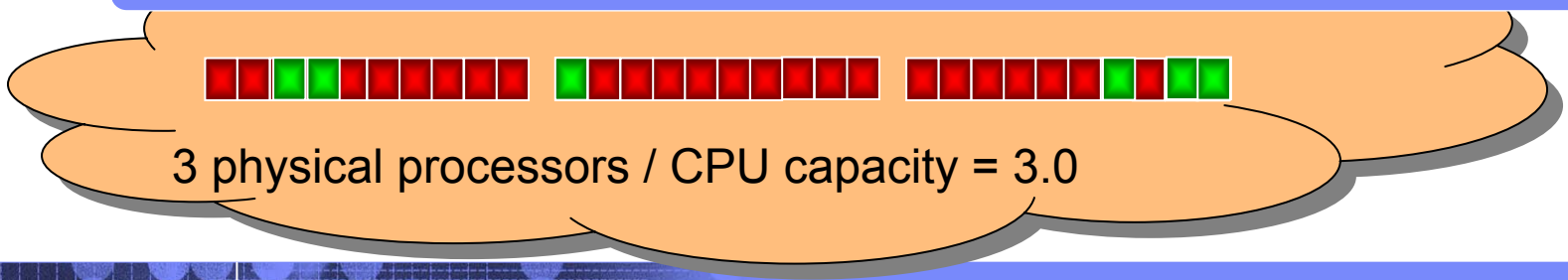


Micro-partitions: OS optimizations

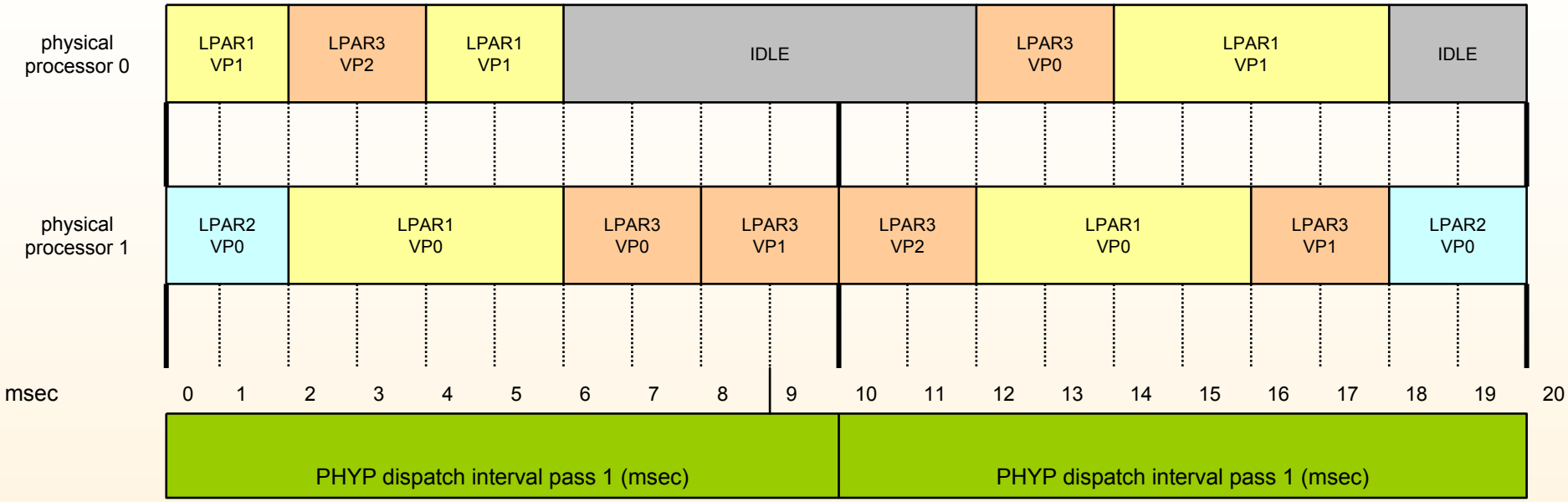
AIX5.3 and linux kernel 2.6 cede their idle CPU to the shared processor pool



SPLPAR allows a safe, dynamic and automatic adjustment of processor power according to the immediate workload need



Micropartitions : dispatching



	CE	VP	mode
LPAR1	0.8	2	capped
LPAR2	0.2	1	capped
LPAR3	0.6	3	capped

Micropartitions Increase Productivity

Fewer processors, & adapters, less memory

Without Partitioning

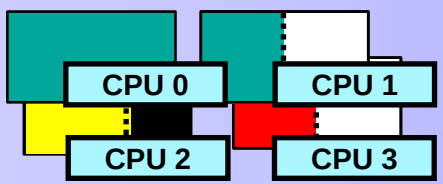
One OS image per server

Dedicated, Underutilized Resources

With Partitioning

SMP Partitioning on POWER4

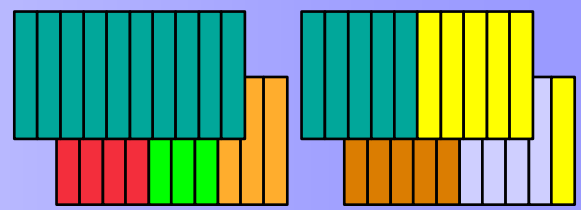
One partition per CPU



1 to 4 partitions in a 4-way SMP

SMP Partitioning on POWER5

Multiple partitions dynamically and responsively dispatched on each CPU



Up to 40 partitions in a 4-way SMP

System resource allocation can be fine tuned to adapt to rapidly changing business priorities.

Benchmark feed back: PeopleSoft

Configuration :

3* p5 595 : 64@1.9GHz 256GB RAM

SunFire 25K

DS 8300

Software stack

AIX 5.3 ML3, no VIO, GPFS

People Soft 8.46

Oracle RAC 10gR2

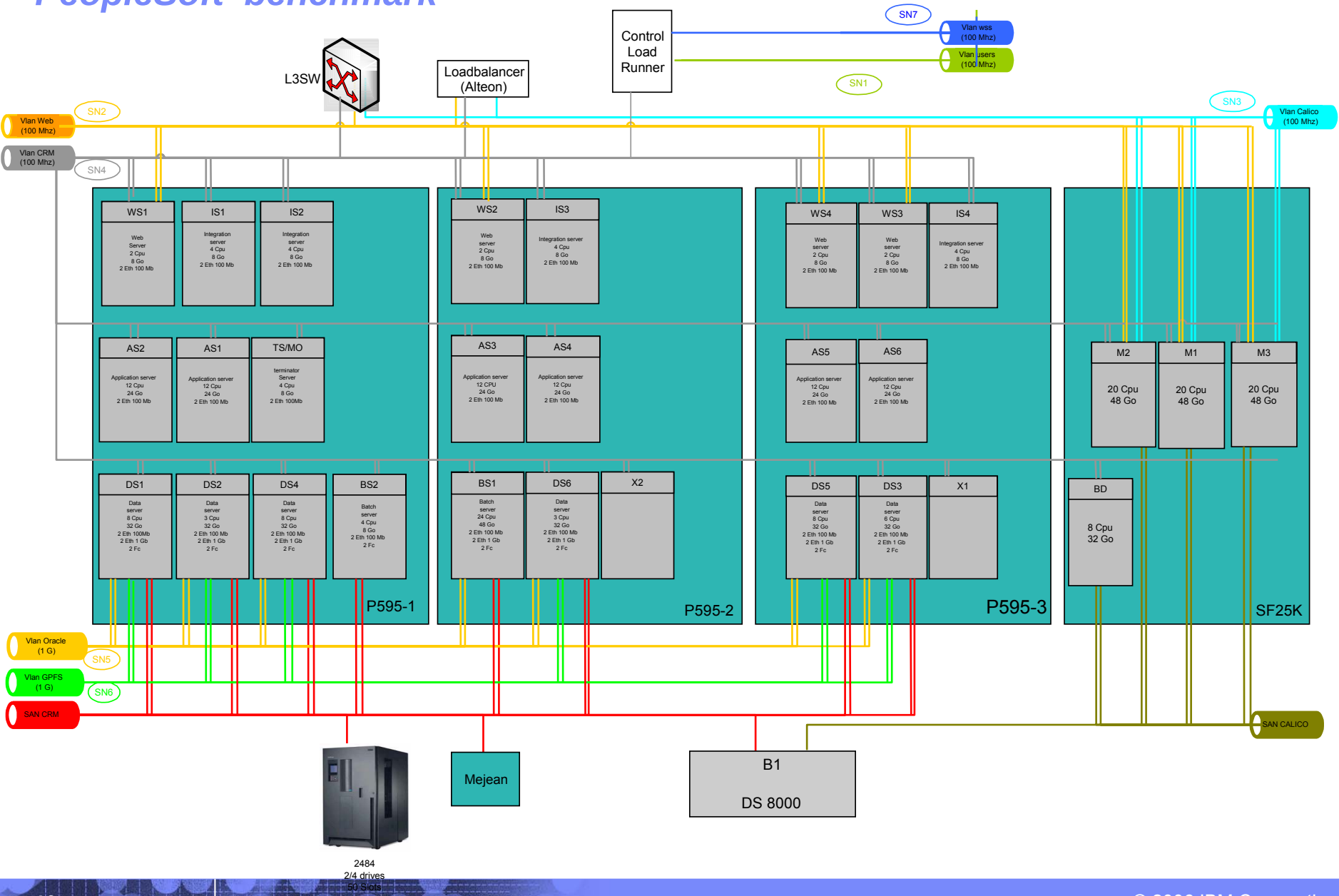
MQ 5.3 .05

WAS 5.1.1

Tuxedo

Benchmark target: OLTP and Batch Stress Test

PeopleSoft benchmark

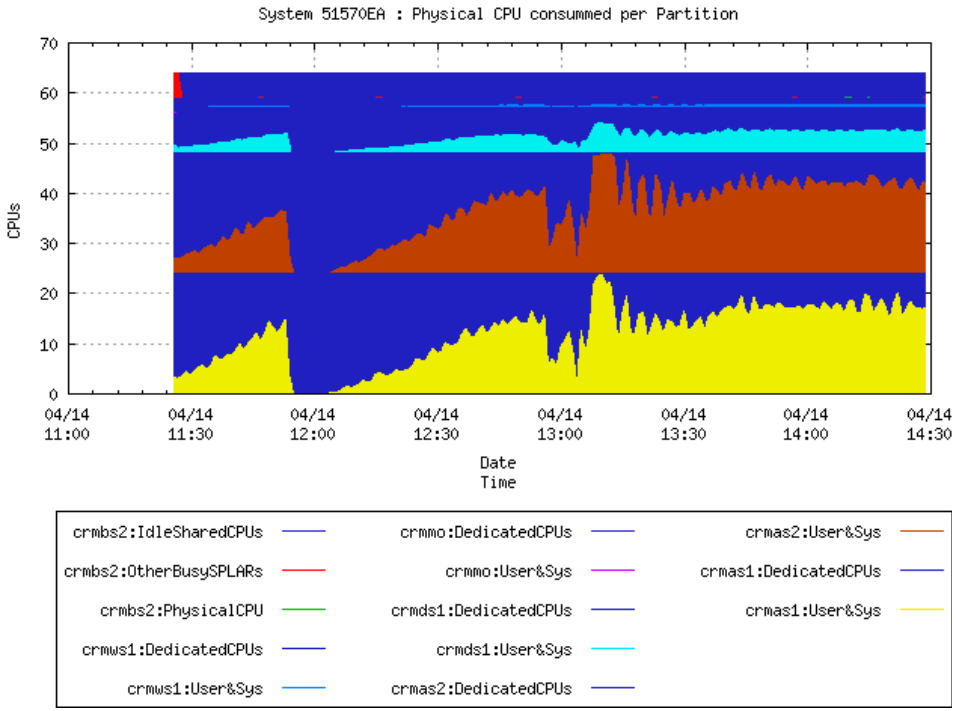


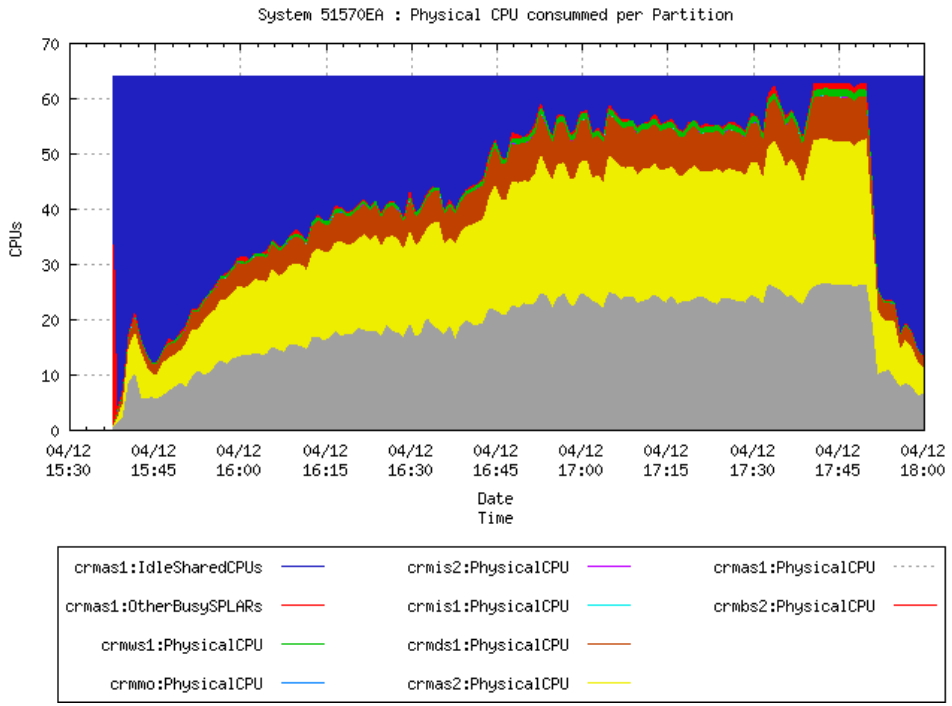
PeopleSoft benchmark:results

Dedicated/Shared configurations comparison: 2,400 running virtual users

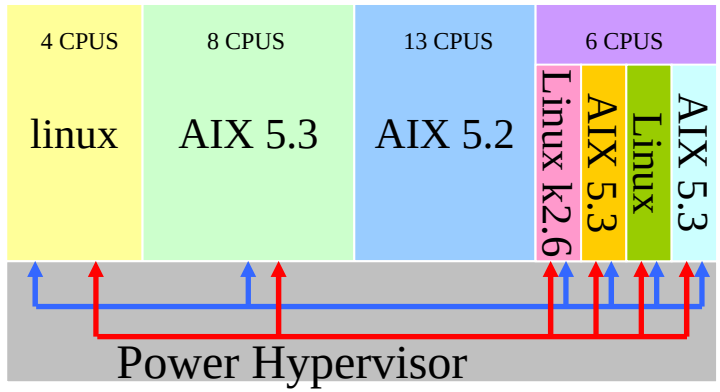
Dedicated mode :3 * 64 CPUs Dedicated
sustain Throughput :1,110,000 bytes/s
Action_Transaction: 59.07 sec

Shared mode : 3*64 shared processors,
sustain Throughput :1,112,000 bytes/s
Action_Transaction: 53.256 sec





VLAN Switch



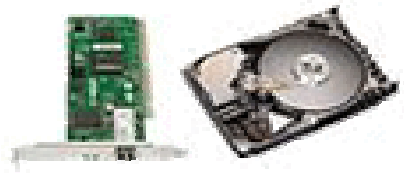
IEEE VLAN style implementation
- Consistent with IEEE 802.1Q
- Upto 4094 VLANs starting with VID 2
- PVID tag for untagged packets

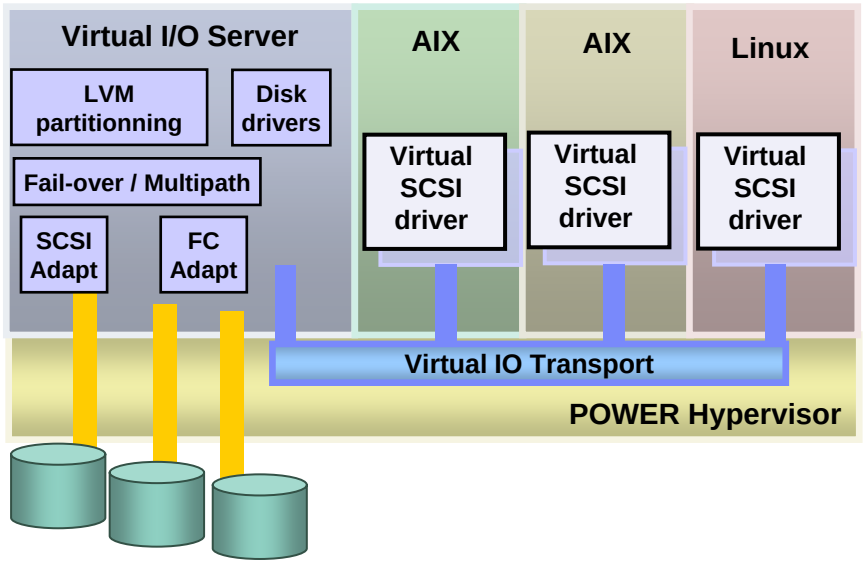
Switch configuration through HMC
- Multiple ports per LPAR
- Multiple (upto 18) VIDs per port
- HMC generates MAC address
- Locally administered ethernet address
- Option to override prefix

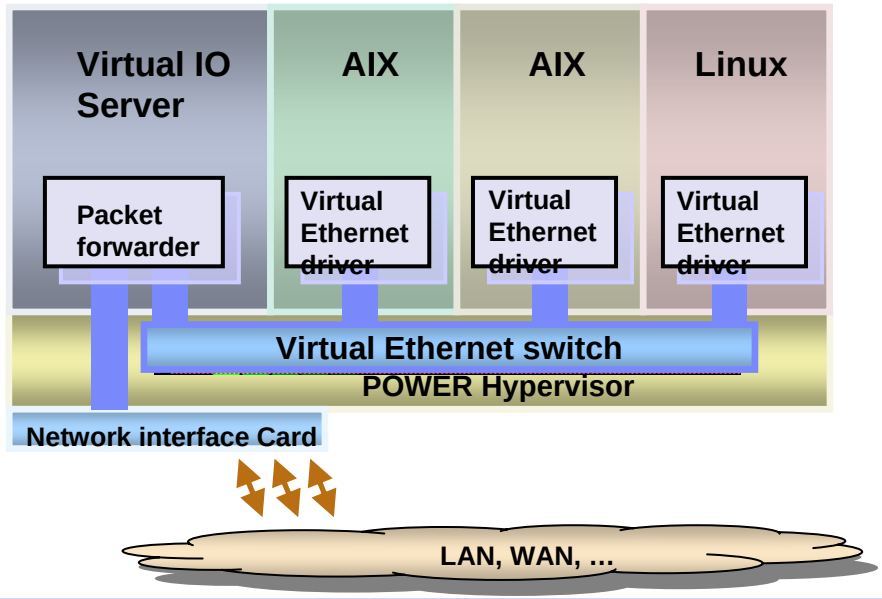
Memory based Interpartition LAN
- Packets copied between LPARs
- Network adapters are not needed for Interpartition communication

Virtual I/O Server

- Special type of LPAR used to share I/O devices on a p5 Server
- Created like other LPARs but loaded with the VIO Server code
- Physical I/O devices are attached to VIO Server and then shared among other partitions as virtual devices







VIOS : virtual disks

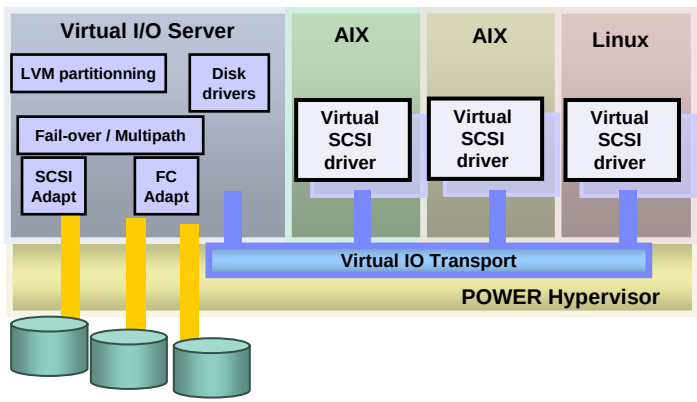
- Virtual SCSI bus: Link between VIOS and a client partition for virtualized disks

  A Virtual disk is attached to a virtual SCSI bus

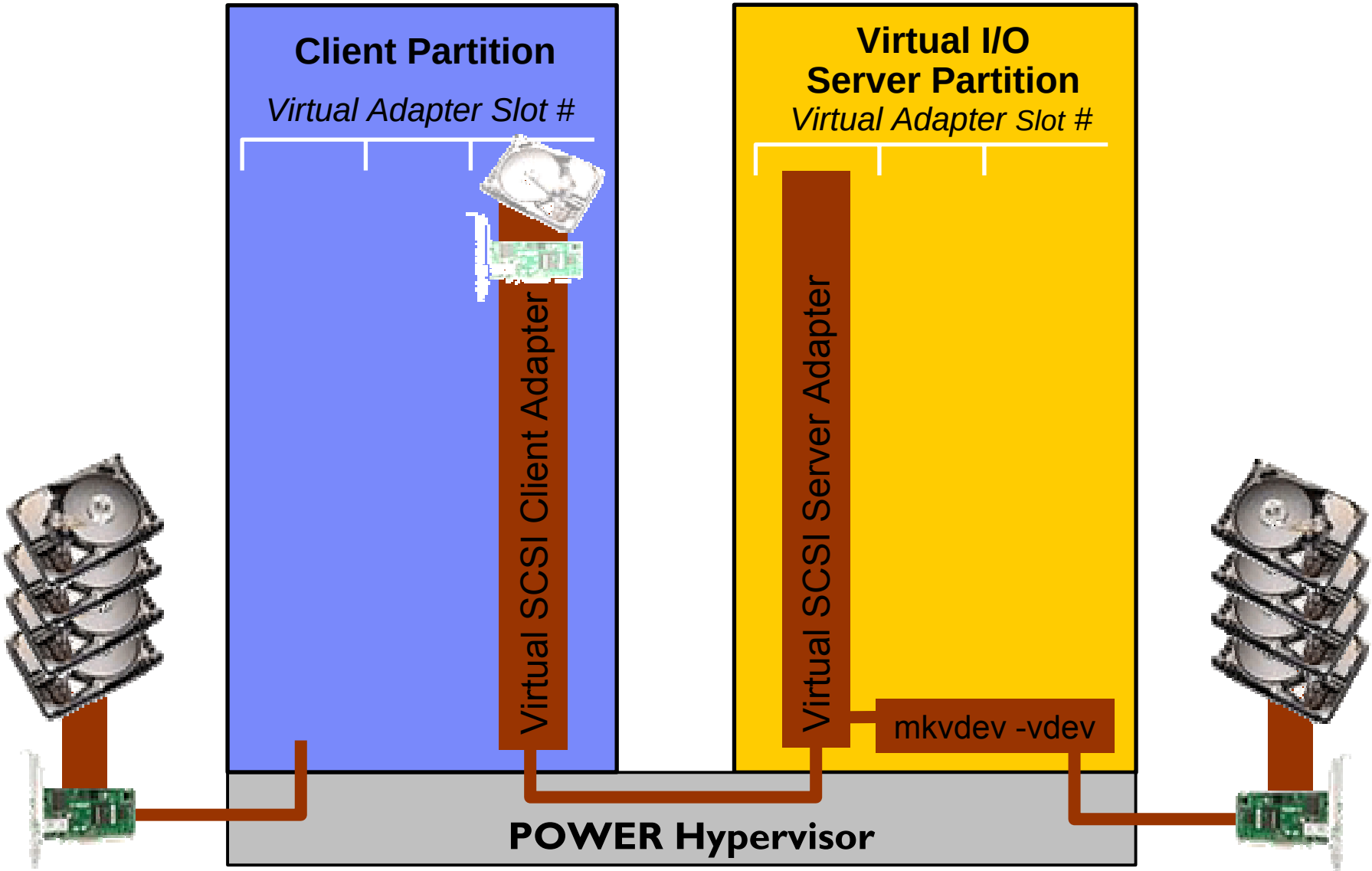
- 2 ways to create Virtual disks :

  One physical drive can be split into multiple virtual disks with LVM slicing

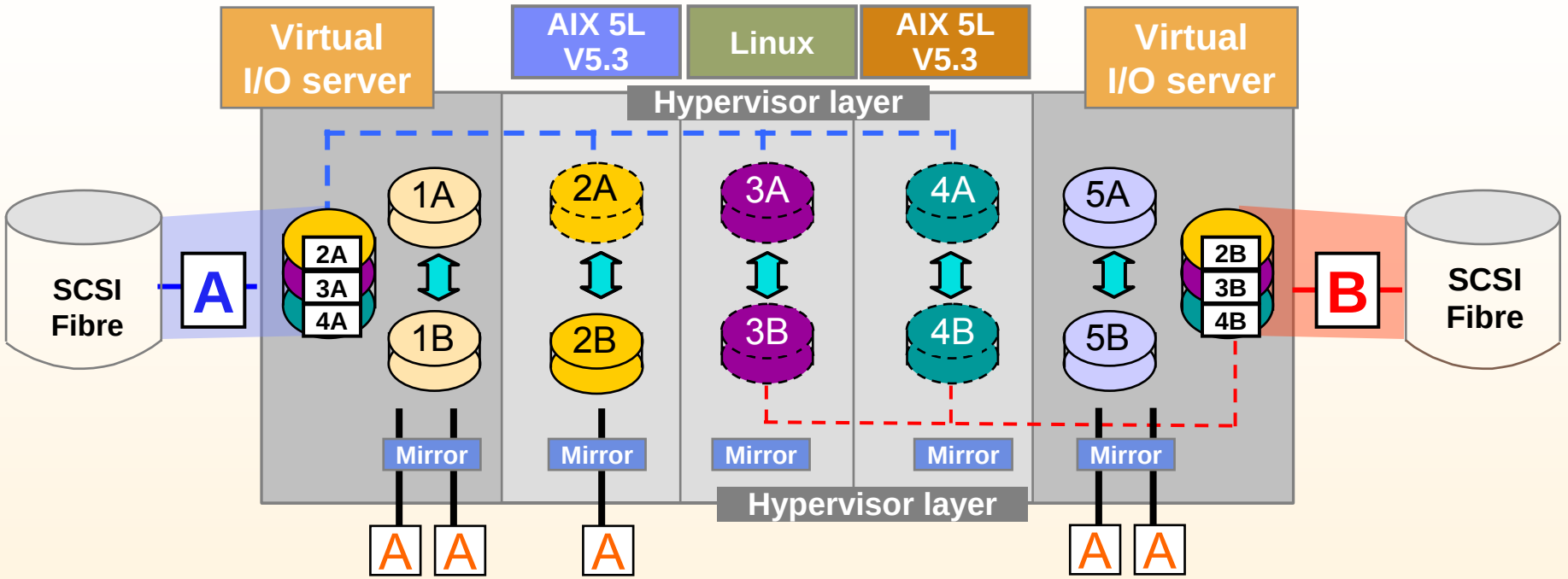
  LUNs can also be mapped “as is” to virtual disks



Virtual SCSI

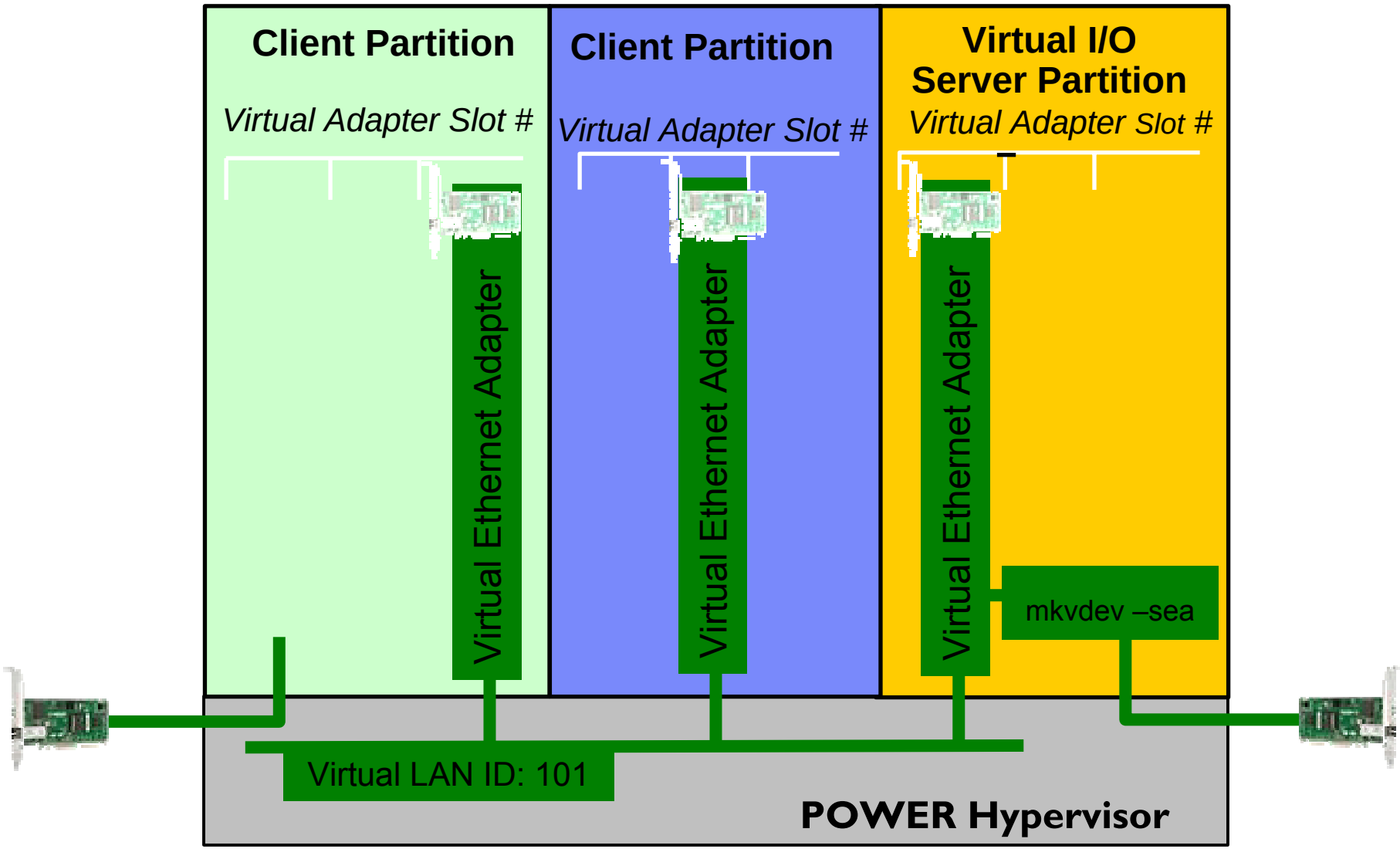


Virtual I/O server disk sharing

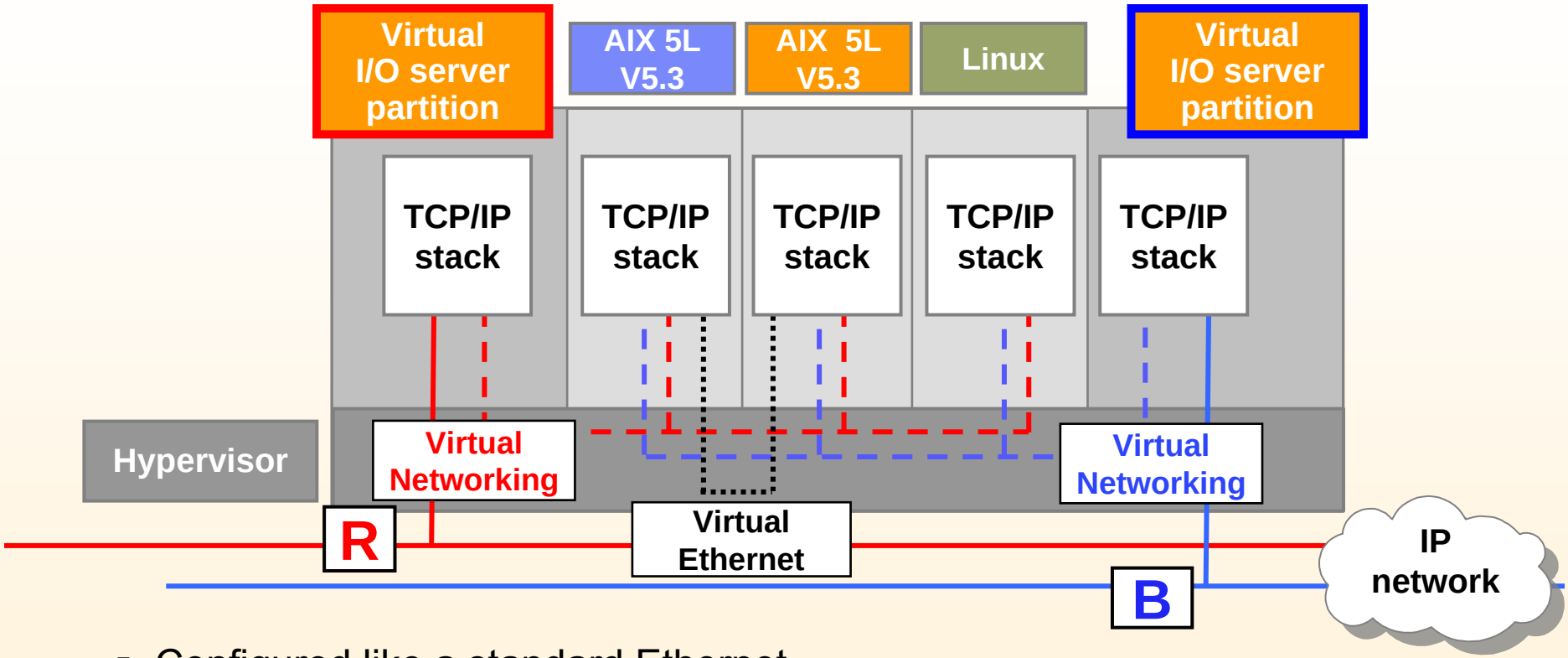


- One physical drive can appear to be multiple logical drives
  - LUNs appear as individual logical drives
- Minimizes the number of adapters
- Can have mixed configuration (virtual and real adapters)
- SCSI and Fibre supported
- Supports AIX 5L V5.3 and Linux partitions

Virtual LAN / Shared Ethernet Adapter



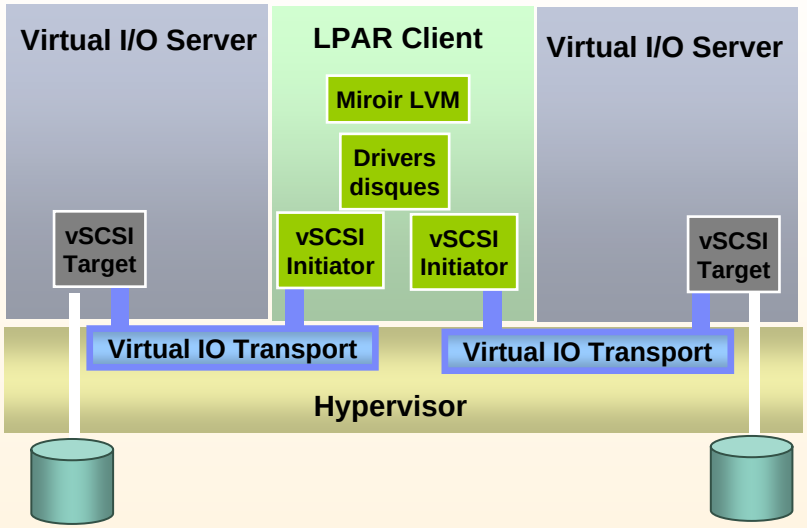
Virtual I/O server Ethernet sharing

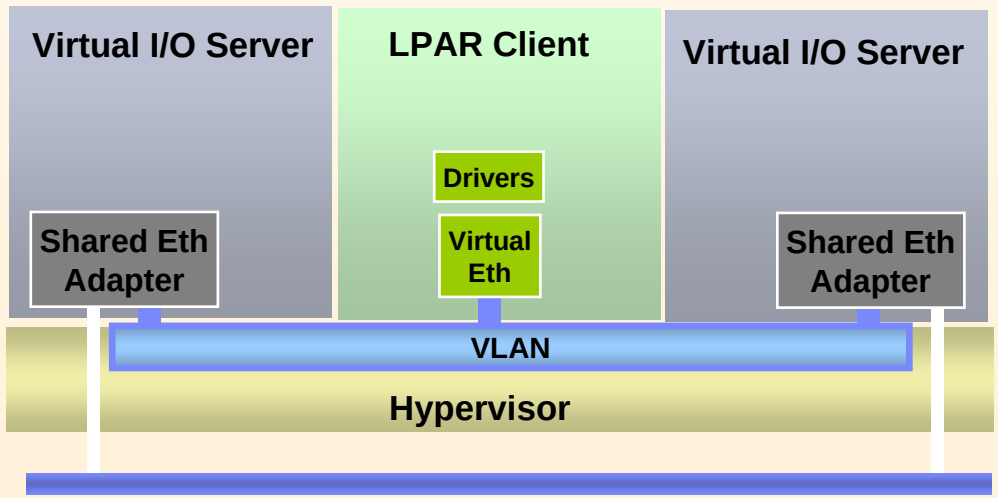


- Configured like a standard Ethernet
- Ethernet bridging provided by I/O server partition
- Can have multiple connections per partition
- Virtual “MAC” addressing
- Each adapter can support 16 virtual Ethernet LANs

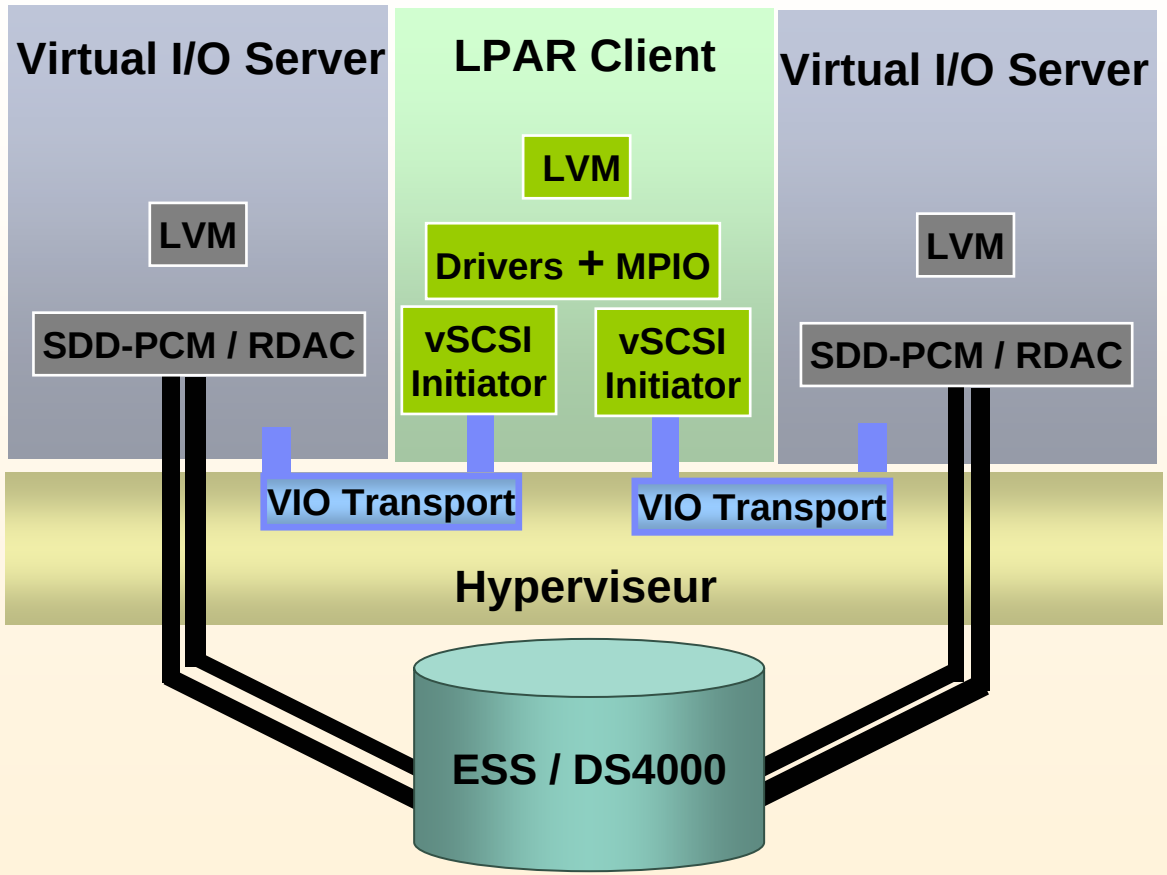
VIOS : High Available configuration

- Protection against a VIOS stop
- Security against adaptors or disks crashes



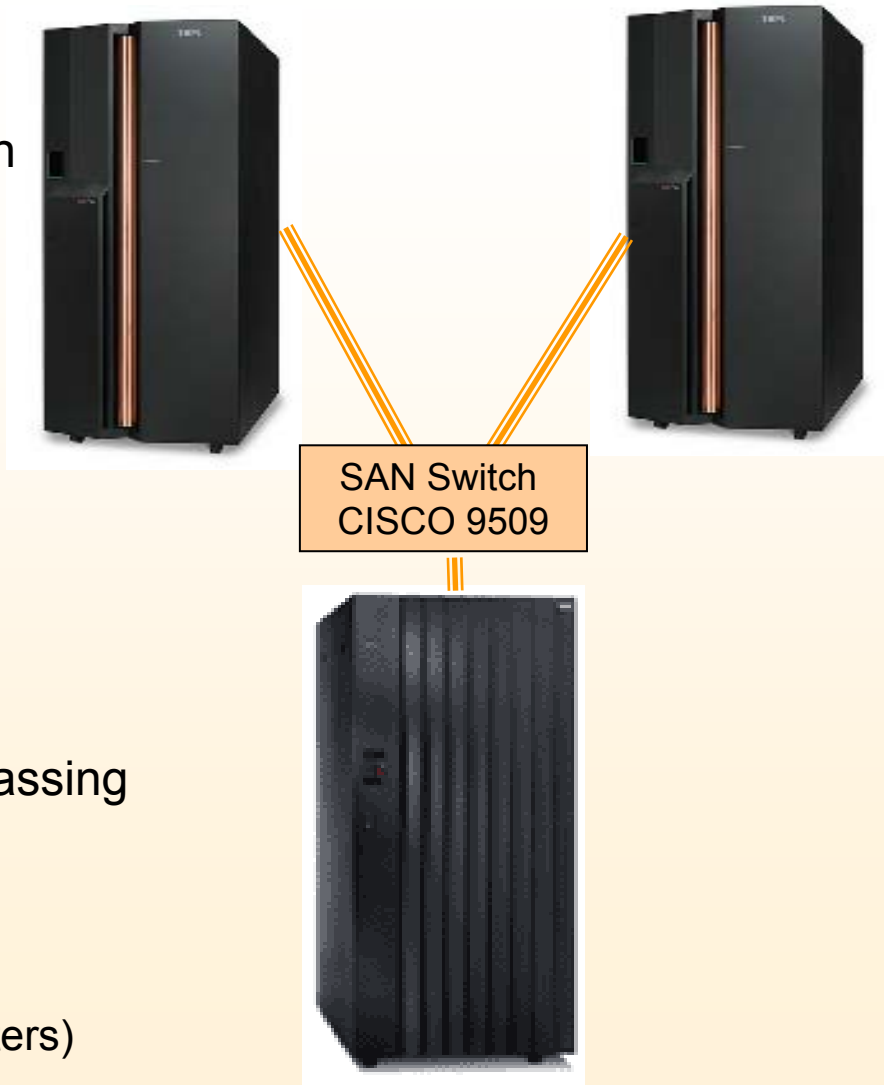


Virtual I/O server : client MPIO

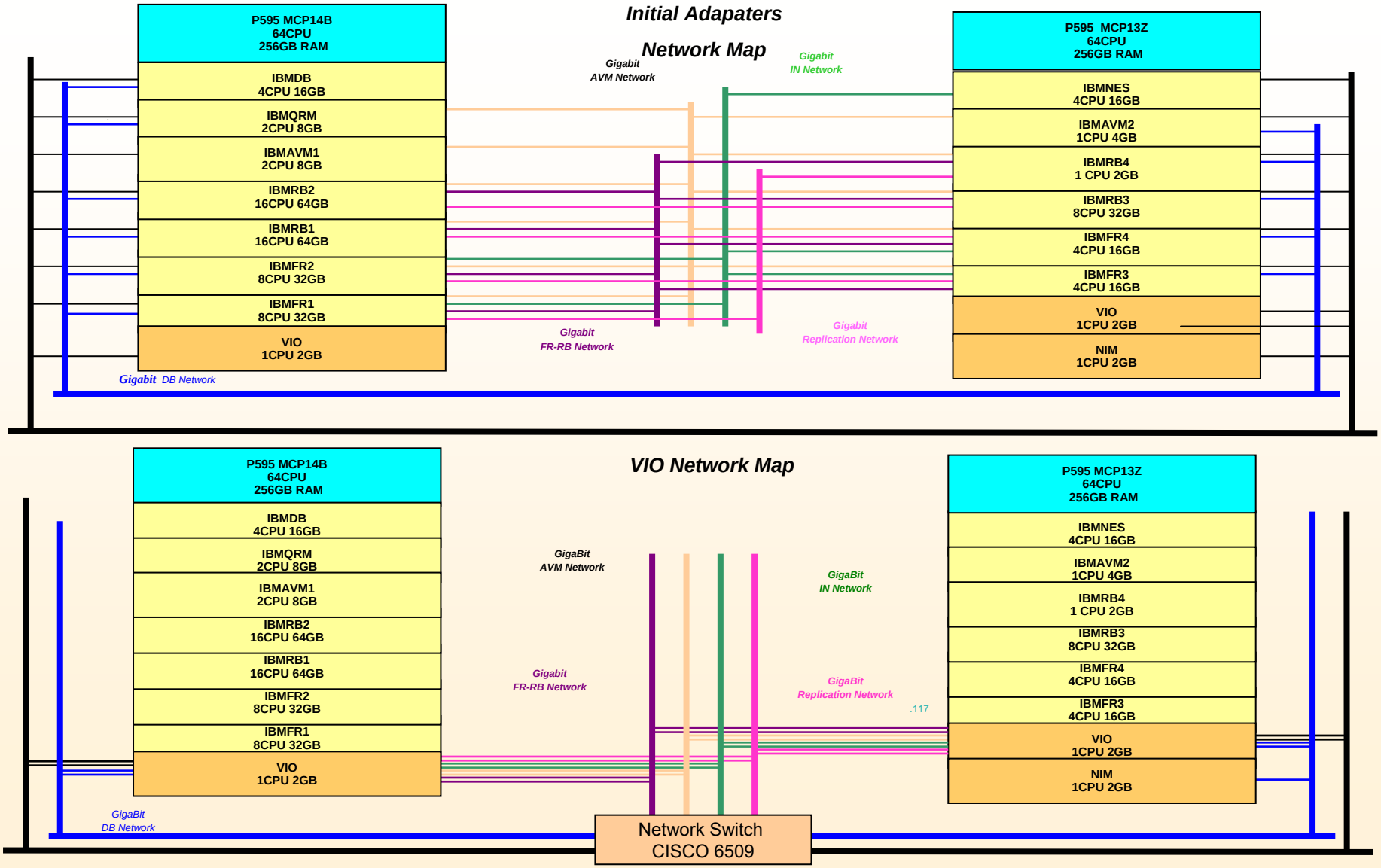


Benchmark feed back : AMDOCS benchmark

- Telco software : postpaid application
- configuration :
  - 2 * P5 595
    - 64 Power5@1.9GHz
    - 256 GB RAM
    - 5 LPARs per system
  - 1 DS8300
  - AIX 5.3 ML3, VIOS 1.2
  - AMDOCS/Oracle 10g/TimesTen
- telco billing and rating: Messages passing application
- VIOS to share all network adapters:
  - EtherChannel (2 physical adapters)
  - 1SEA per VLAN

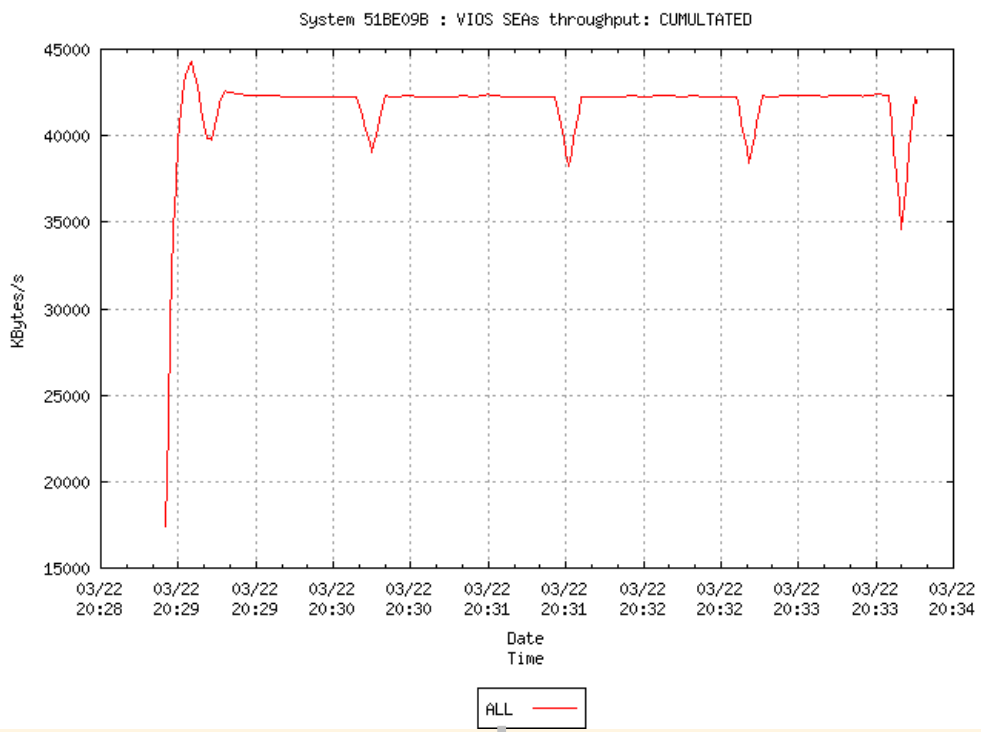


AMDOCS benchmark : network configuration



AMDOCS benchmark: results and analysis

–Results :
- – VIOS configuration :
  - – 1 Cpu Dedicated
- – VIOS bandwidth →
- – VIOS CPU consumption :
  - – ~70% CPU



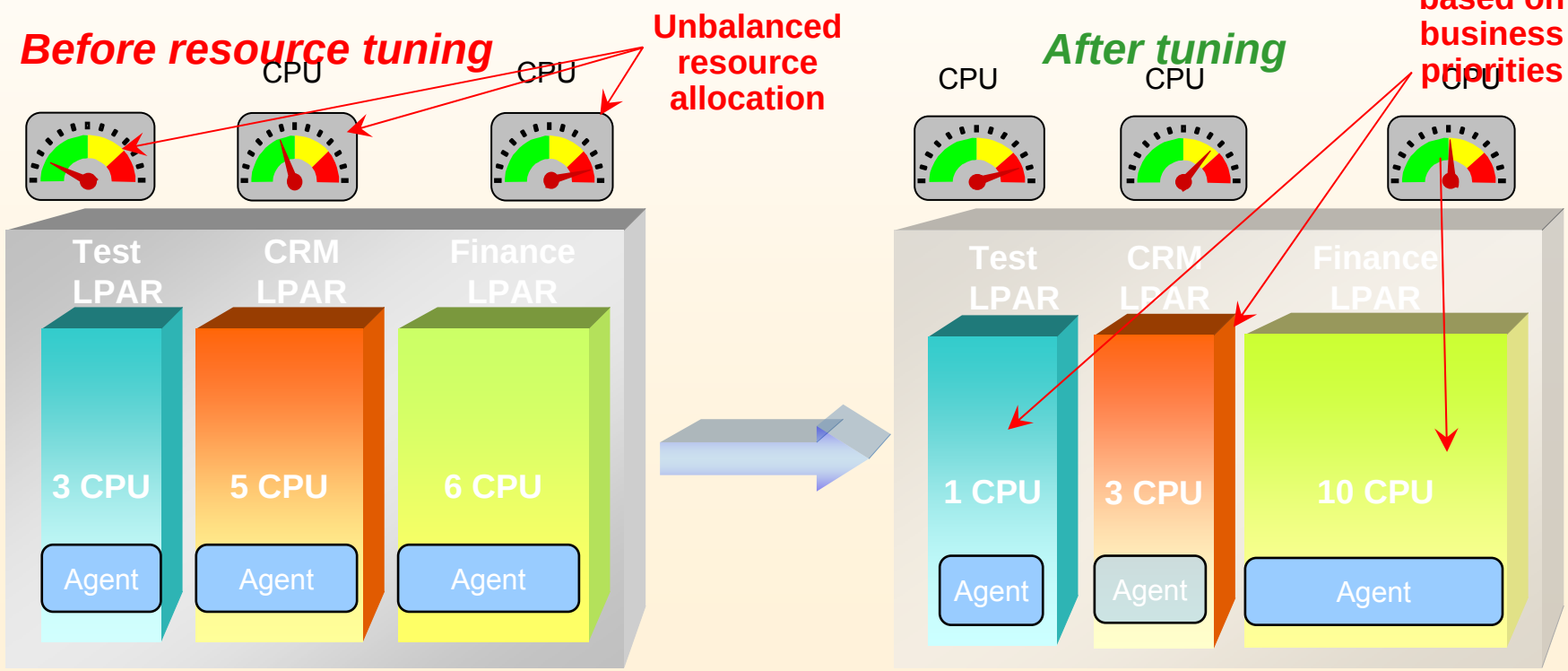
– Analysis :
- – SEA latency and throughput OK regarding AMDOCS expectations
- – Final tests reached the DS8k disks performance limits.

www.amdocs.com :

"IBM Demonstrates Breakthrough Performance for Amdocs Real-time Convergent Charging "

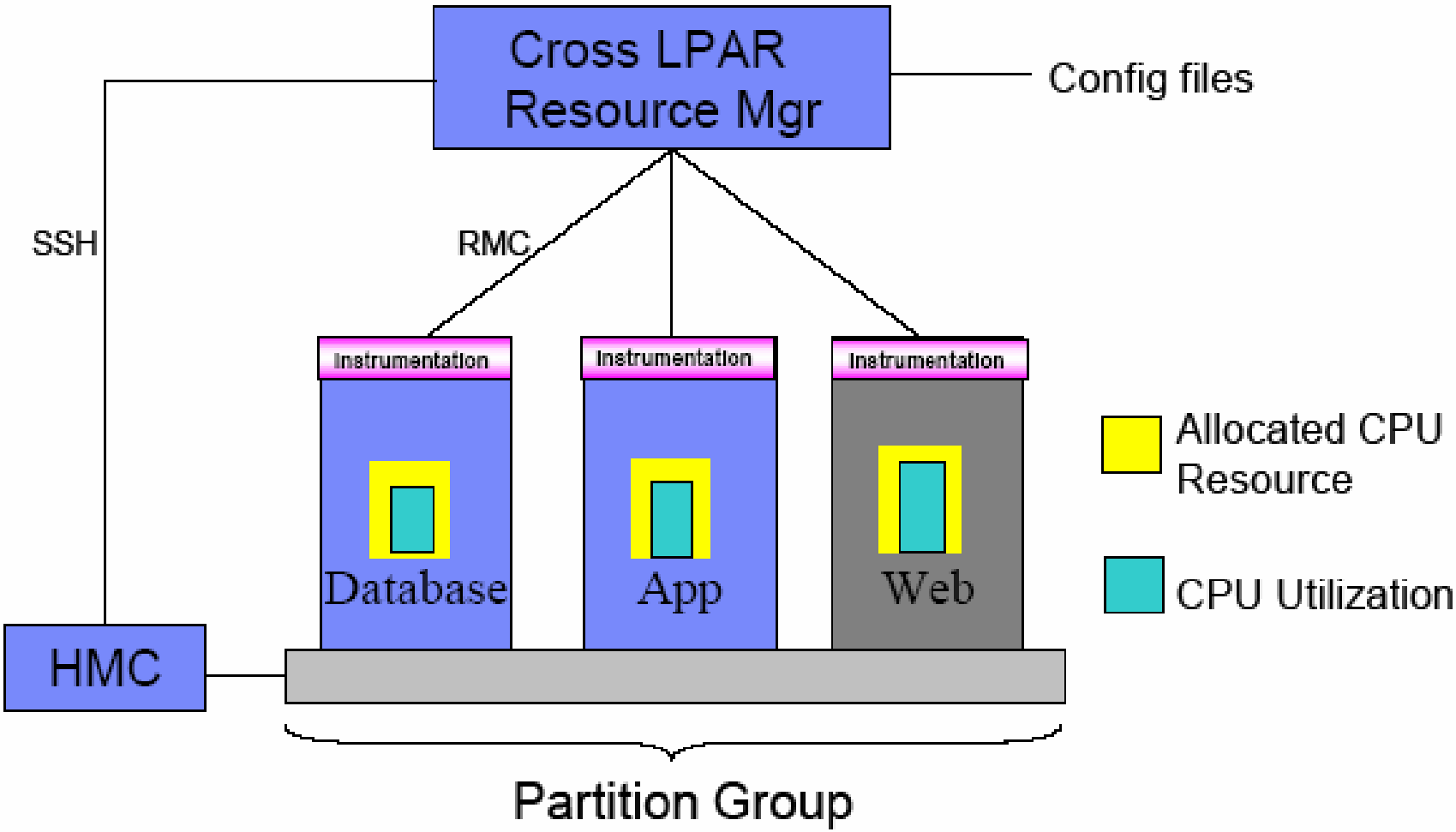
Partition Load Manager

- Policy-based, automatic partition resource tuning
- Can dynamically adjust CPU and memory allocations
- Supports AIX 5L V5.3/V5.2 partitions
- p5-520, p5-550, and p5-570 systems



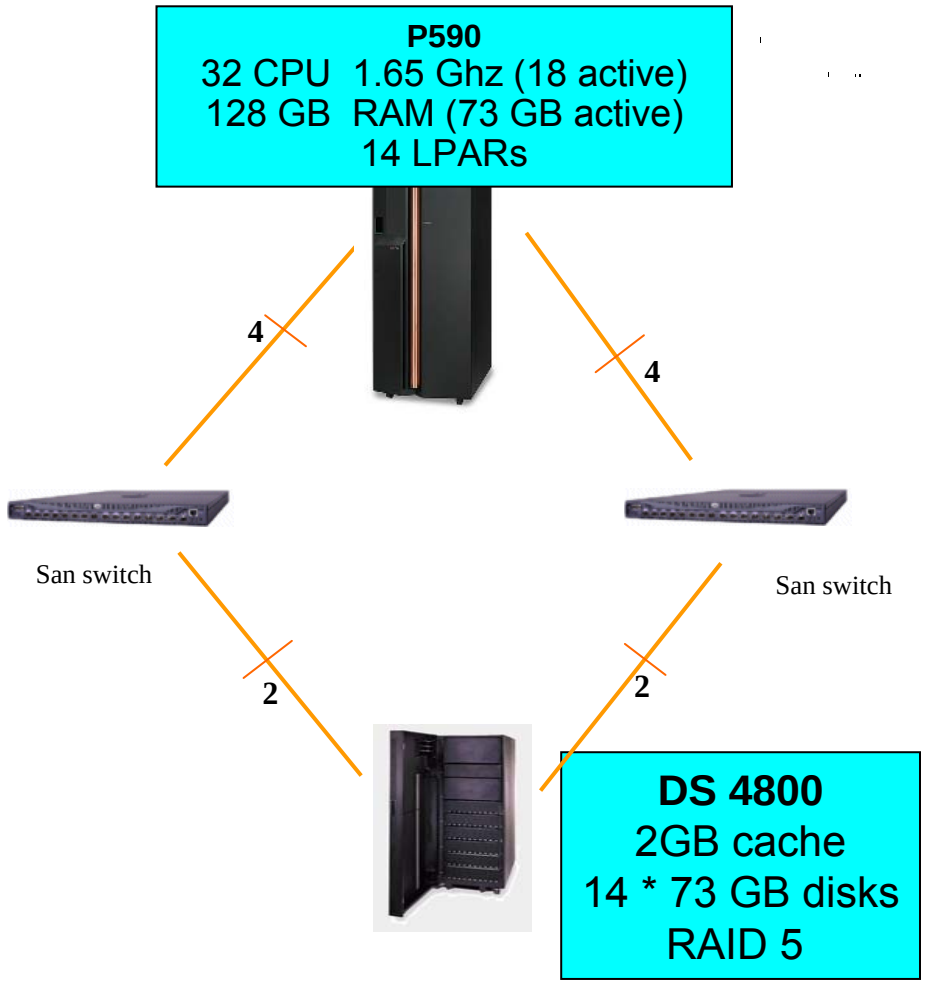
* Note: optional feature

Cross LPAR Resource Manager (PLM)

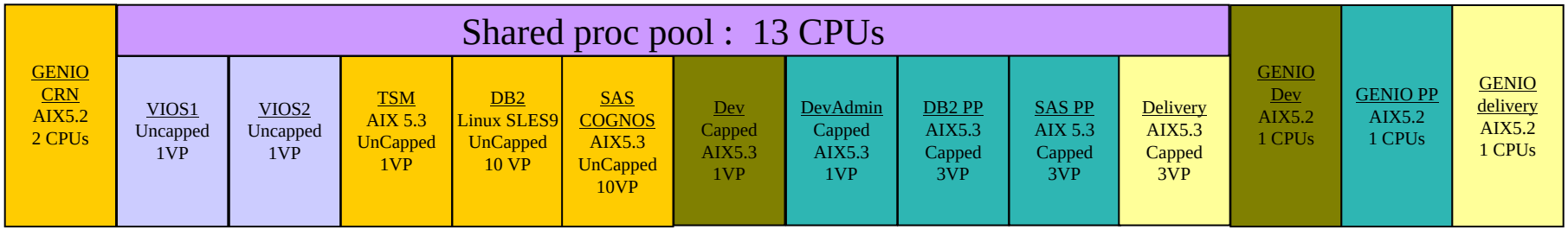


Bank : Business intelligence environment

- One Application → four environments
  - Production
  - Delivery
  - Pre-production
  - Development
- Requirements :
  - Price/performance
  - Flexibility and scalability

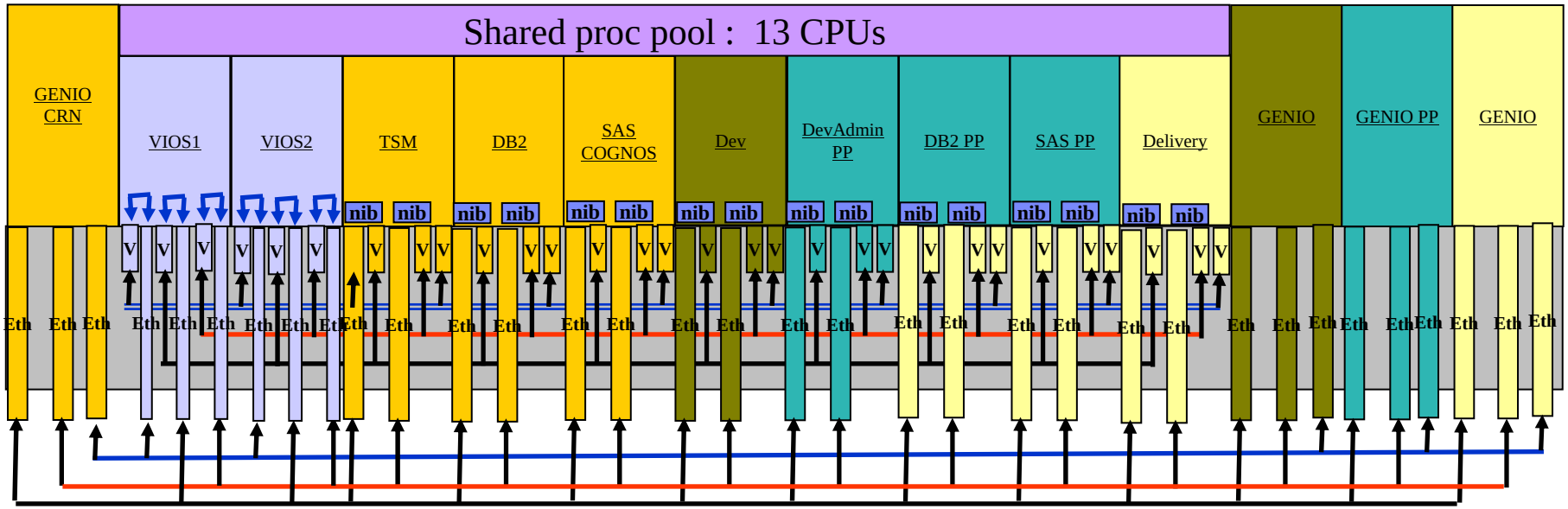


Bank : LPARs configuration



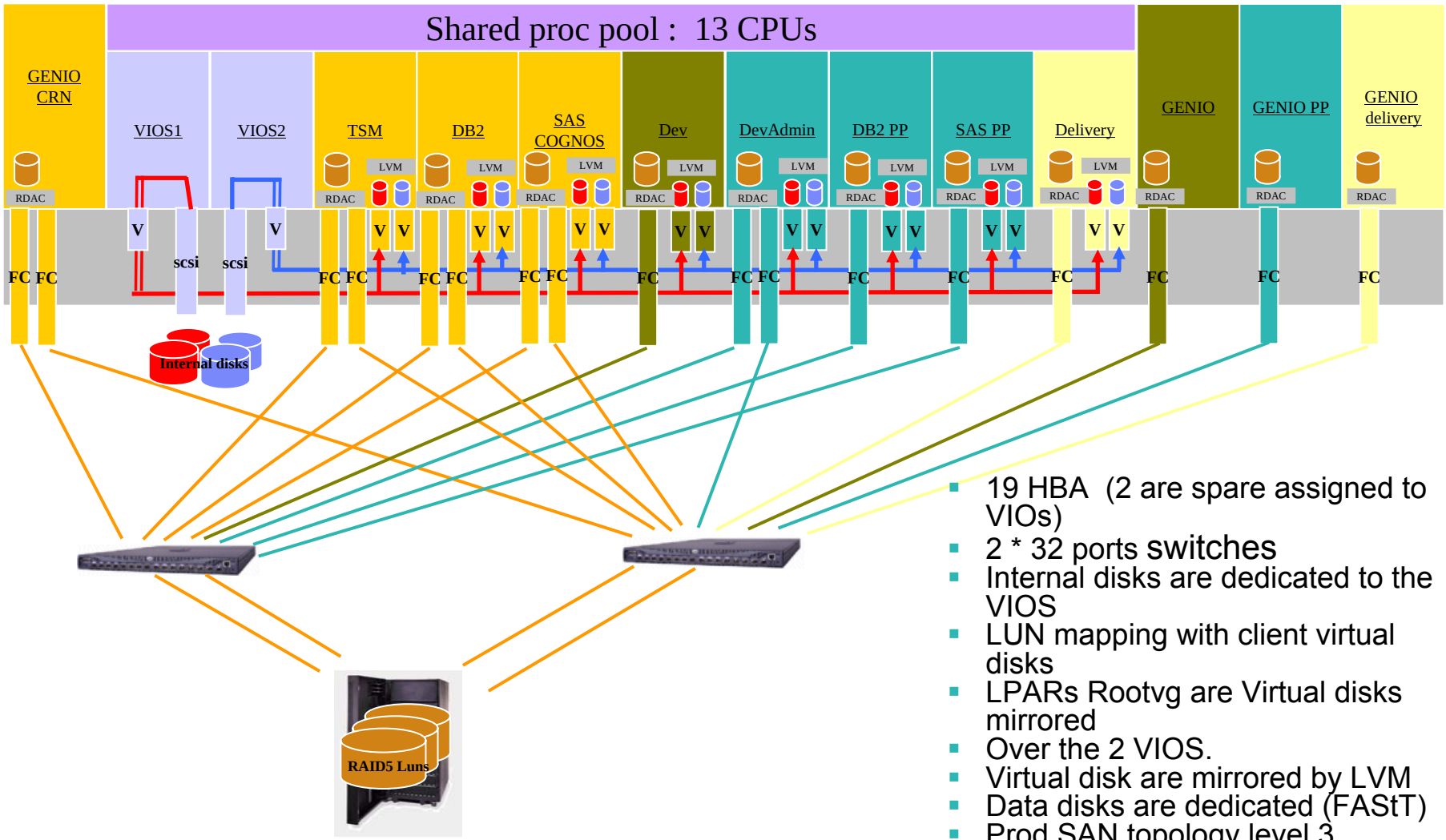
- 4 environments :
  - Production → Uncapped, enough VPs to consume Pool
  - Pre-Production → Capped
  - Delivery → Capped
  - Dev → Capped
- SPLPARs :
  - Dedicated LPAR required for Genio software (not released on AIX 5.3)
  - CE and VP are subject to changes, if the workload is required
  - Version : AIX 5.2 ML2 , VIOS 1.1

Bank : networks configuration



- 3 VLANs :
  - Admin (100Mb) All partitions, 2 Virtual VLAN (one per VIO) + client NIB,
  - Data(GigE)
  - Users(GigE)
- Physical Adapters : 17 dual Ethernet, 6 mono (VIOS)
- Use of AIX Network interface backup (nib), with physical adapter as primary, and virtual as secondary (same VLAN thanks to a VIOS SEA)
- No use of 802.1Q ethernet tagging.

Bank : disks configuration



- 19 HBA (2 are spare assigned to VIOs)
- 2 * 32 ports switches
- Internal disks are dedicated to the VIOS
- LUN mapping with client virtual disks
- LPARs Rootvg are Virtual disks mirrored
- Over the 2 VIOS.
- Virtual disk are mirrored by LVM
- Data disks are dedicated (FAStT)
- Prod SAN topology level 3

I/O Virtualization roadmap

I/O Hosting Partition V1 (GA 9/04)
- AIX based
- virtual disk(LV and phys volume backed)
- SEA
- command line interface

I/O Server V1.2 (GA 10/05)
- virtual optical(DVD, CDROM)
- HMC-Lite + I/O Server
- HA SEA
- Performance management

I/O Server V3
- Nport ID Virtualization(NPIV)
- virtual tape

Alpha partition
- HV and blades environment(HMC-less)

Futures
- Performance management, QOS
- LPAR migration support
- on-demand storage provisioning

Additional Resources

- **Redbook: Introduction to Advanced POWER Virtualization on IBM p5 Servers - SG24-7940-00**
- **Redbook: Advanced POWER Virtualization on IBM eServer p5 Servers Architecture and Performance Considerations - SG24-5768-00**

WWW.redbooks.ibm.com

Any question ?