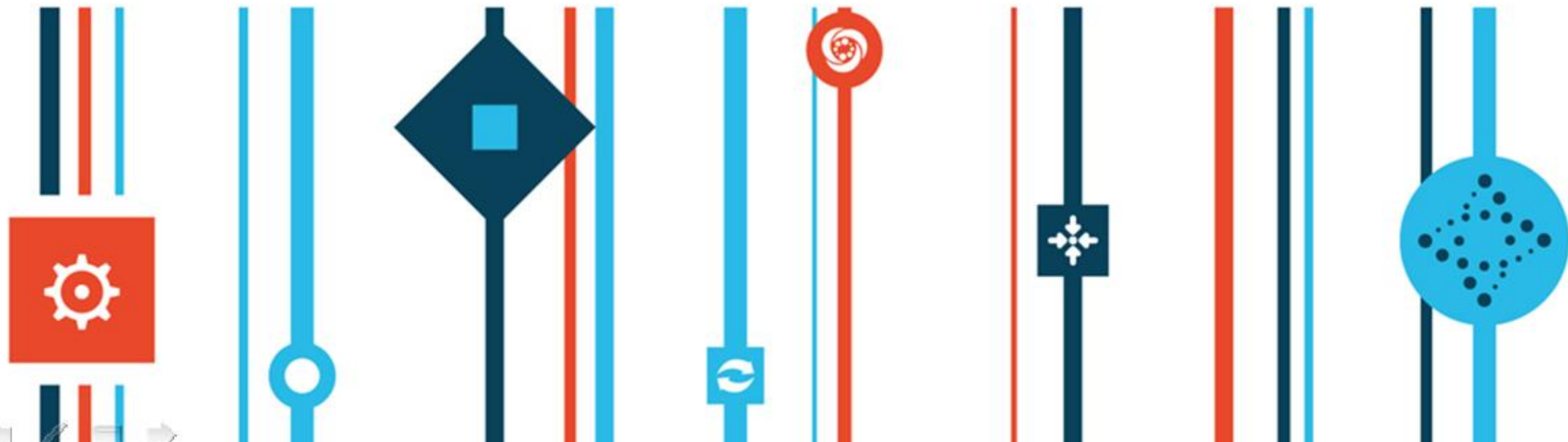


# POWER智慧系统的完美虚拟化技术 —— 云计算坚实可靠的源动力

李一峰

IBM系统与科技部

liyifeng@cn.ibm.com

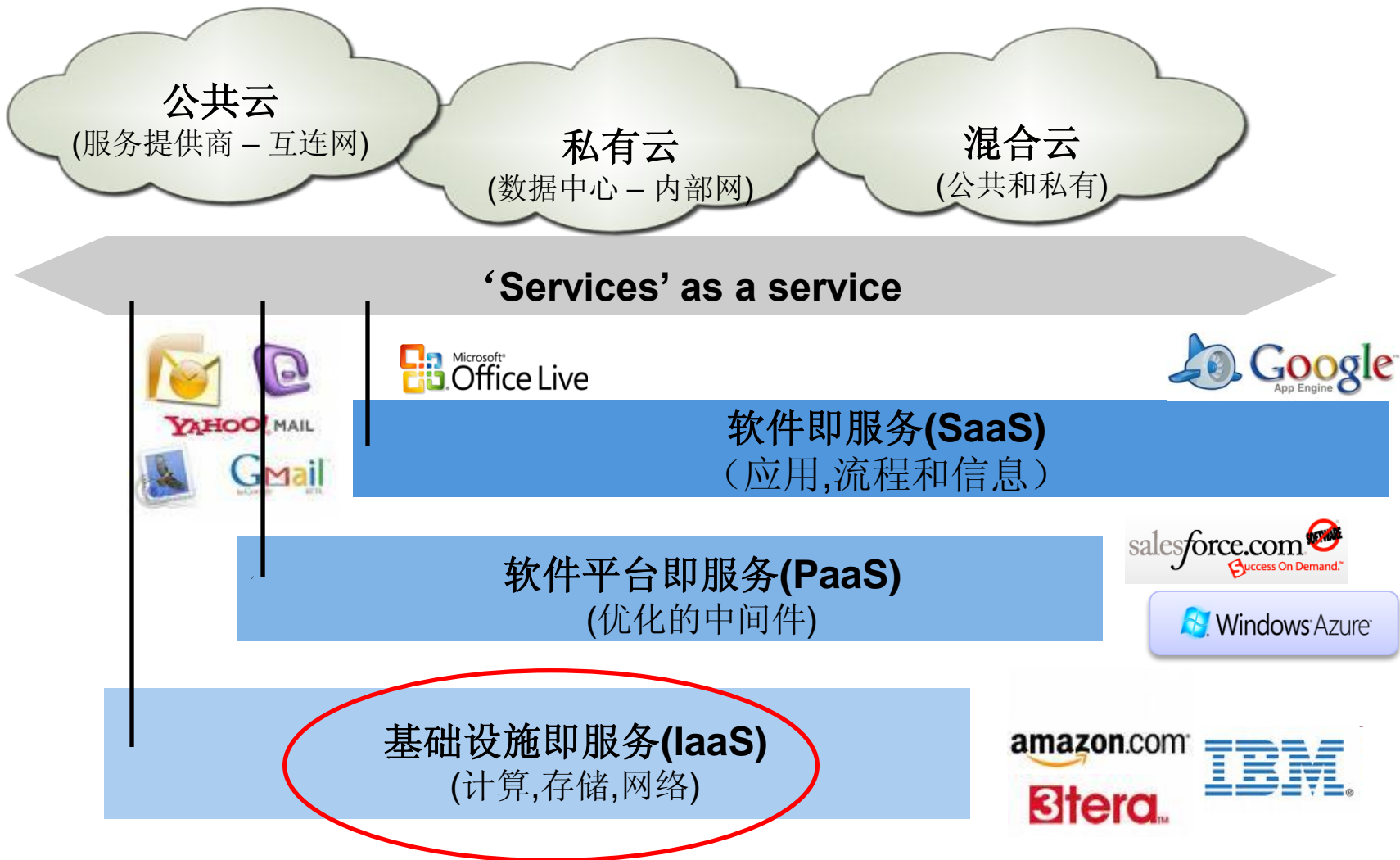


# 议题

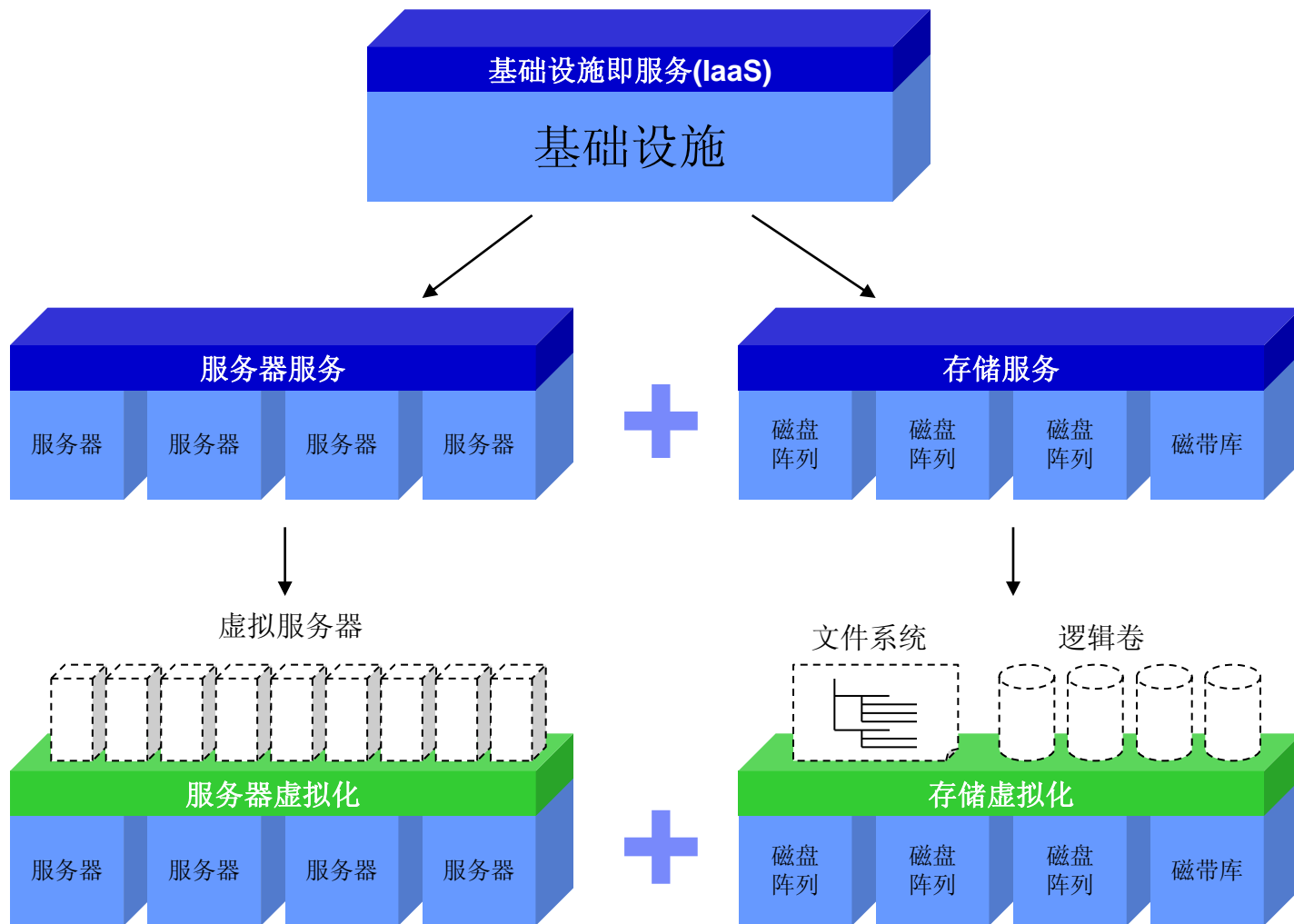
- 服务器虚拟化技术概览
- IBM PowerVM虚拟化技术介绍
  - CPU虚拟化
  - I/O虚拟化
  - 内存虚拟化
  - 分区在线迁移
  - 负载分区
- 实施服务器虚拟化的考虑
- 服务器虚拟化的案例
- 虚拟化的管理



# 云计算



# 虚拟化是云计算的技术基础



# 为什么在大型服务器上实施虚拟化



## 更高的服务器利用率

- CPU & memory
- I/O ports & adapters
- Large or small workloads



## 更有效的数据中心

- Power and generator
- Cooling
- Floor space



## 提高生产力

- Order, track, retire
- Install, cable, fix
- Patch fewer hypervisors

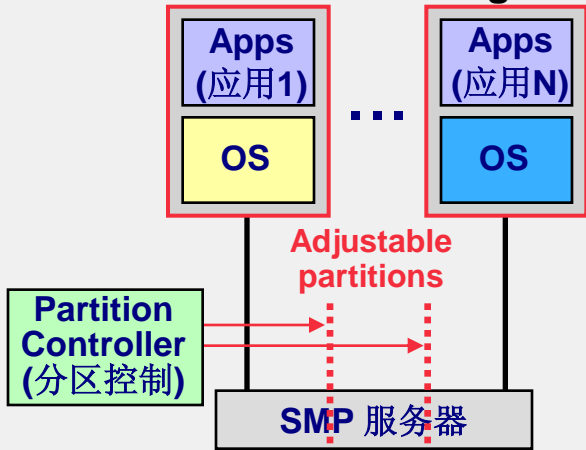


## 强大的RAS

- More RAS features
- Fewer parts
- Capacity on Demand

# 服务器虚拟化技术概览

## 硬件分区技术 Hardware Partitioning



Server is subdivided into fractions each of which can run an OS

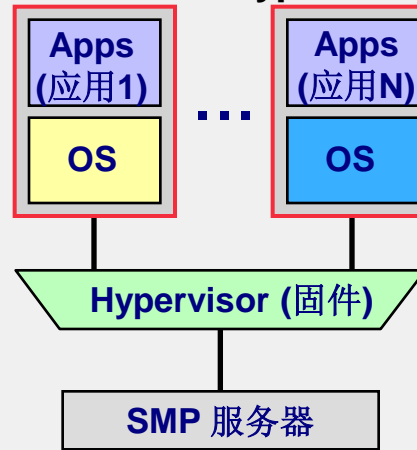
### Board-level partitioning

- S/370 SI->PP & PP->SI
- Sun Dynamic Domains
- HP nPartitions (nPAR)

### Core/thread-level partitioning

- HP vPartitions
- Sun Logical Domains
- Original POWER4 LPAR

## 固件分区技术 Bare Metal Hypervisor

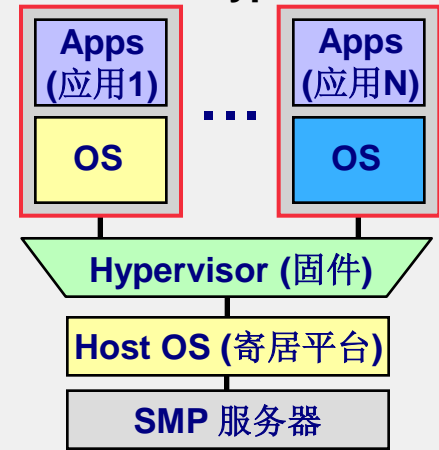


Hypervisor provides fine-grained timesharing of all resources

### Hypervisor software/firmware runs directly on server

- System z PR/SM and z/VM
- POWER Hypervisor
- VMware ESX Server
- Xen Hypervisor
- Kernel VM
- Microsoft Hyper-V

## 软件分区技术 Hosted Hypervisor



Hypervisor uses OS services to do timesharing of all resources

### Hypervisor software runs on a host operating system

- AIX Workload Partition (WPAR)
- Solaris Container
- HP Integrity VM
- VMware GSX
- Microsoft Virtual Server

- 硬件分区受到限制很多,发展有限;因此采用硬件分区技术将越来越少。
- 固件分区技术成为主流虚拟化实现的方式,能提供丰富功能,灵活性,可用性。
- 软件分区技术将主要被重视操作系统层整合的客户采用。

# IBM — 虚拟化技术的领导者

*A 40-year track record in virtualization innovation continues with PowerVM™*

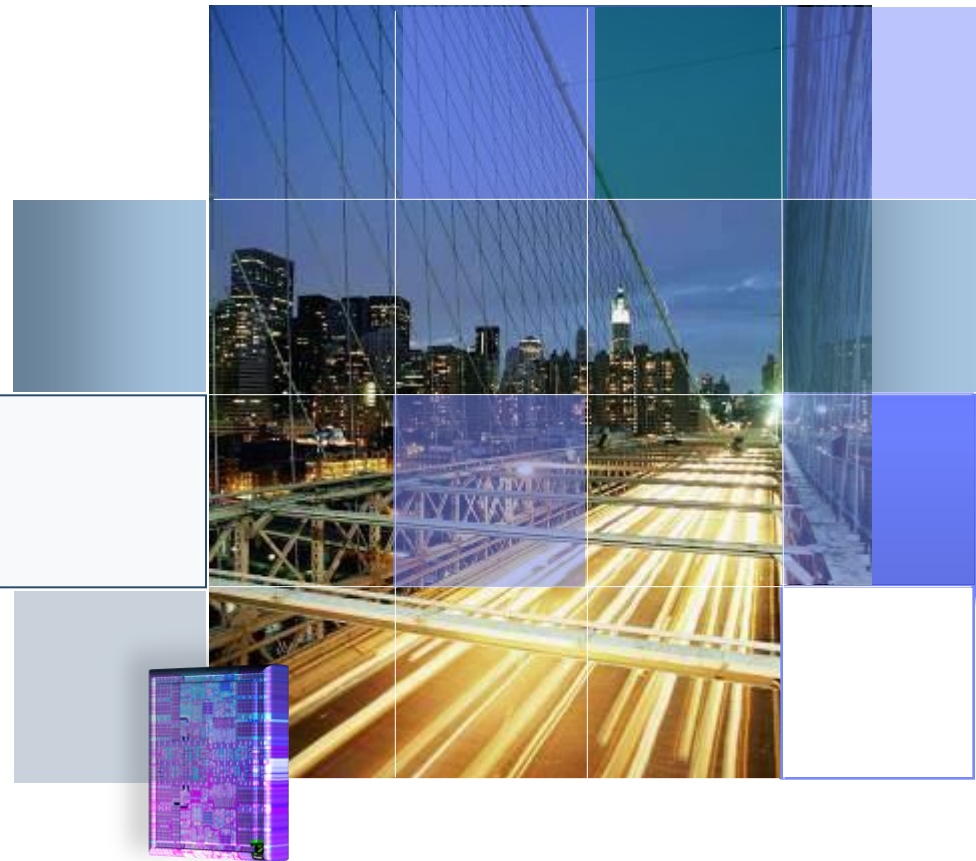


<p>1967</p> <p>IBM develops <b>hypervisor</b> that would become VM on the mainframe</p>	<p>1973</p> <p>IBM announces first machines to do <b>physical partitioning</b></p>	<p>1987</p> <p>IBM announces <b>LPAR on the mainframe</b></p>	<p>1999</p> <p>IBM announces <b>LPAR on POWER</b></p>	<p>2004</p> <p>IBM intro's <b>POWER Hypervisor™</b> for System p and System i™</p>	<p>2007</p> <p>IBM announces <b>POWER6™</b>, the first <b>UNIX servers with Live Partition Mobility</b></p>	<p>2008</p> <p>IBM announces <b>PowerVM</b></p>
---	--	---	---	--	---	---

采用IBM虚拟化技术实现服务器集中的TCO分析		
费用	预计节省	分析
维护成本	69% - 76%	服务器数量的减少降低了维护所需费用
软件	65% - 69%	更少的服务器数量以及更少的CPU数量，显著降低软件的许可费用、升级和服务费用
人工	31% - 45%	更少的服务器数量和简化的IT架构，自动化的管理工具
运营成本	52% - 61%	节省机房空间、空调、用电等。



## PowerVM虚拟化技术介绍





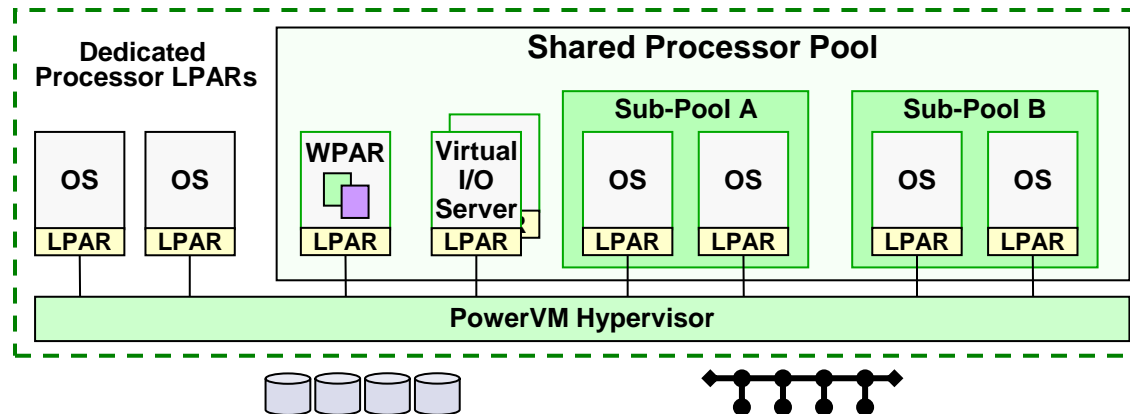
# PowerVM 主要内容

## Processor

- Shared or dedicated LPARs
- Capped or uncapped LPARs
- Multiple shared processor pools
- Dynamic LPAR operations (add/remove)
- Shared dedicated LPARs

## I/O

- Shared and/or dedicated I/O
- Virtual Ethernet, virtual SCSI
- Dynamic LPAR operations (add/remove)
- Integrated Virtual Ethernet
- Virtual FC (N\_Port ID Virtualization)
- Virtual Tape Support



## Memory

- Dedicated memory
- Active Memory Sharing
- Dynamic LPAR operations (add/remove)
- Active Memory Expansion (POWER7/AIX 6.1)

## Other

- Integrated Virtualization Manager
- Live LPAR mobility
- Workload partitions (AIX 6.1)
- Workload partition mobility (AIX 6.1)
- Lx86 for Linux applications (Linux)
- IBM i Virtualization (IBM i)

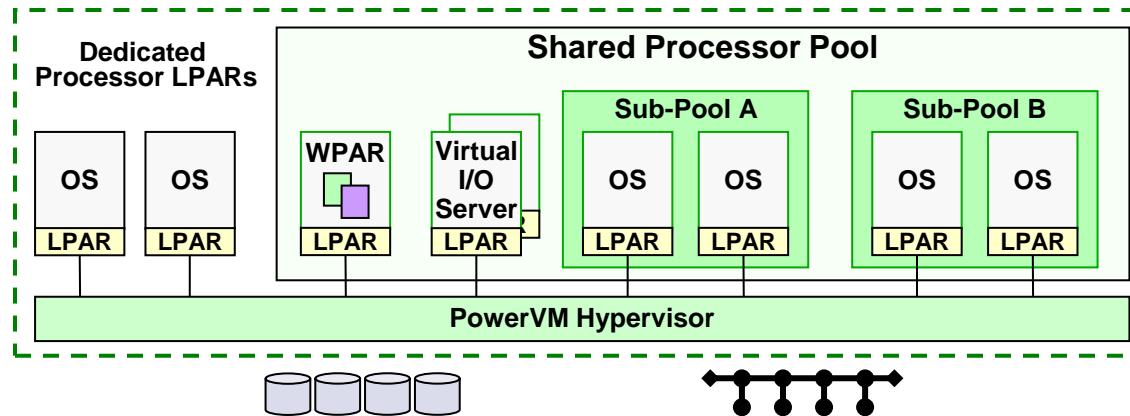
# PowerVM 主要内容

## Processor

- 微分区或专有分区
- 处理能力分区内受限或不受限
- 多个共享处理器池
- 动态分区 (增加或移走CPU)
- 共享专有分区的处理能力

## I/O

- 虚拟I/O和真实I/O可以同时存在
- 虚拟以太网和虚拟SCSI设备
- 动态分区 (增加或移走I/O设备)
- 集成的虚拟以太网卡 (IVM)
- 虚拟FC卡 (NPIV)
- 虚拟磁带



## Memory

- 专有内存
- 共享内存 (AMS)
- 动态分区 (增加或移走内存)
- 动态内存扩展 (AME)

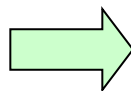
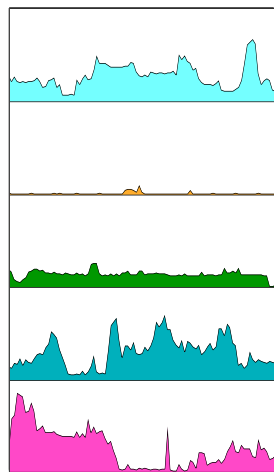
## Other

- 集成的虚拟化管理 (IVM)
- 分区在线迁移 (LPM)
- 负载分区 (WPAR)
- 负载分区在线迁移
- Lx86 for Linux applications (Linux)
- IBM i Virtualization (IBM i)

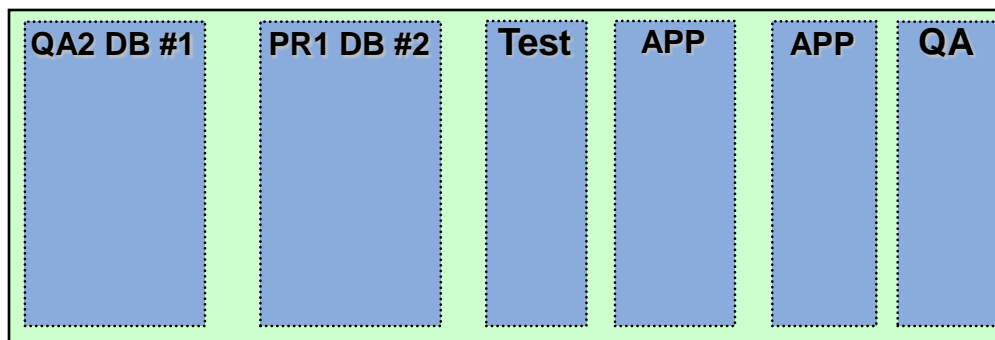
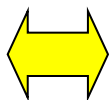
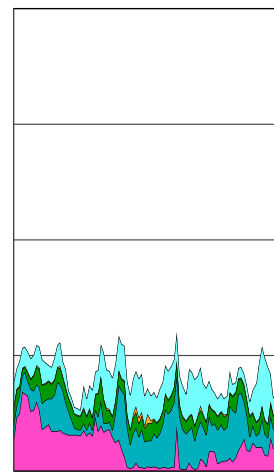
# CPU虚拟化 — 专有分区和微分区

- 专有分区(Dedicated-processor partition): 颗粒度是1个CPU的分区
- 微分区(Shared-processor partition): 颗粒度是0.1个CPU的分区
- 动态分区(DLPAR): 调整分区内的资源（例如：CPU或内存等）时，不需重启分区上的操作系统。

不使用分区或  
仅使用DLPAR  
技术



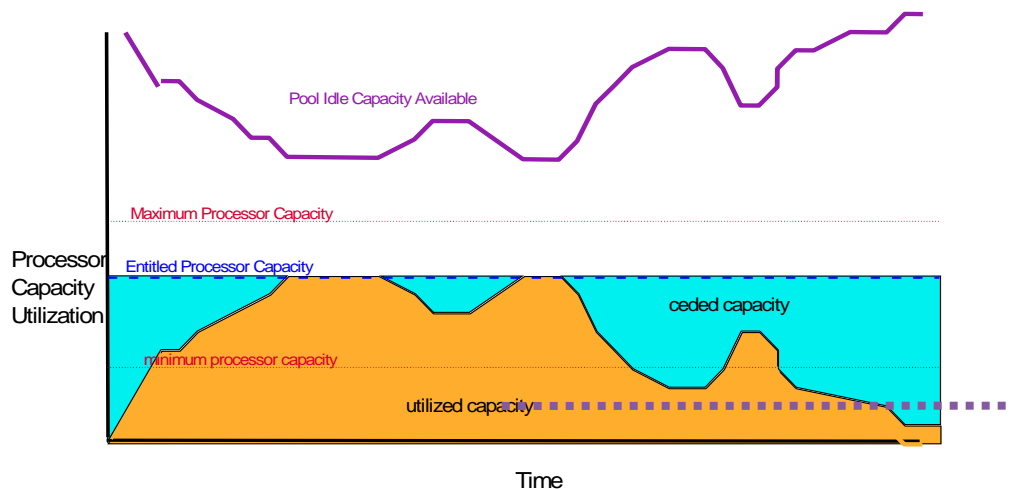
使用微分区和  
共享处理器池  
技术



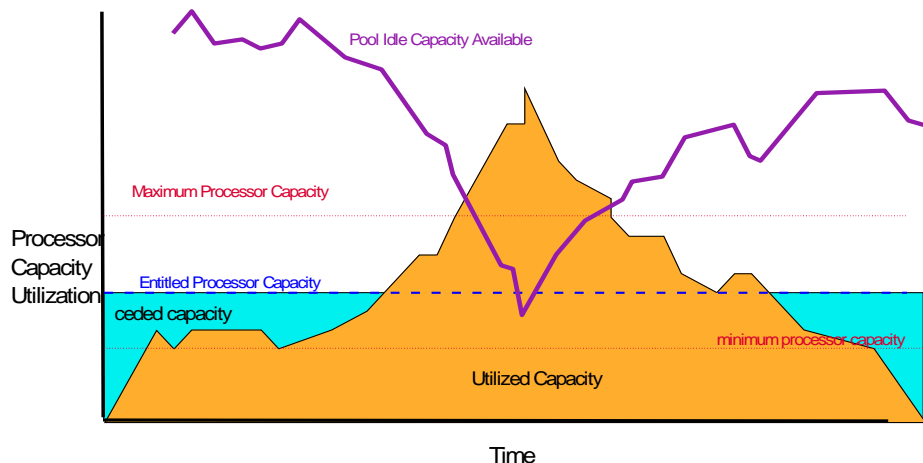
# CPU虚拟化 — 受限(Capped)和不受限(Uncapped)分区

- **受限分区**
  - 不允许超过其分配的CPU容量
- **不受限分区**
  - 允许超过其分配的CPU容量
- **容量加权**
  - 用于为不受限分区分配优先级
  - 范围值：0-255
  - 数值0称为“软限制”

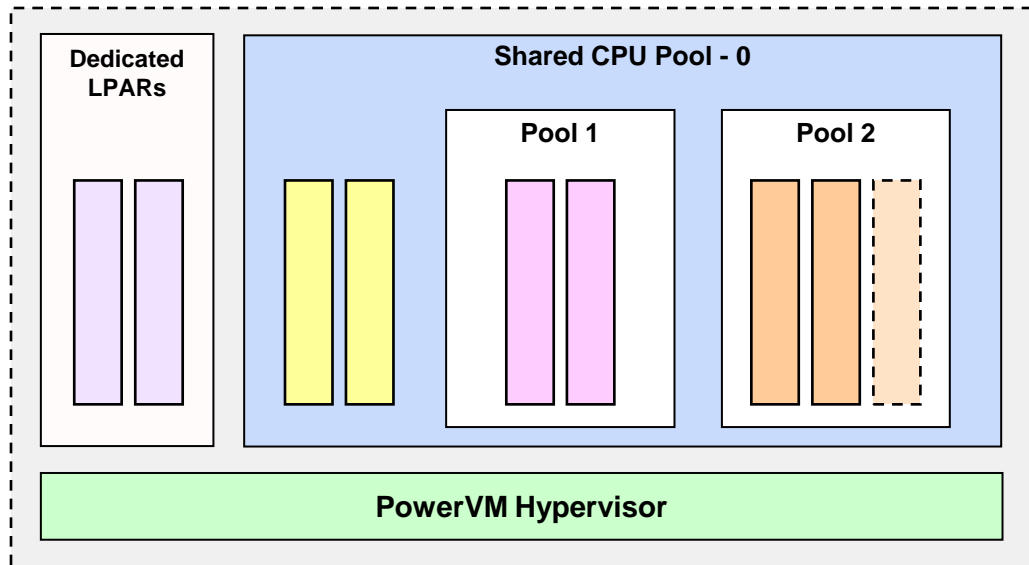
## Capped Shared Processor LPAR



## Uncapped Shared Processor LPAR



# CPU虚拟化 — 处理能力共享



## ➤微分区技术

- CPU划分的颗粒度是0.1颗CPU

## ➤虚拟CPU的个数

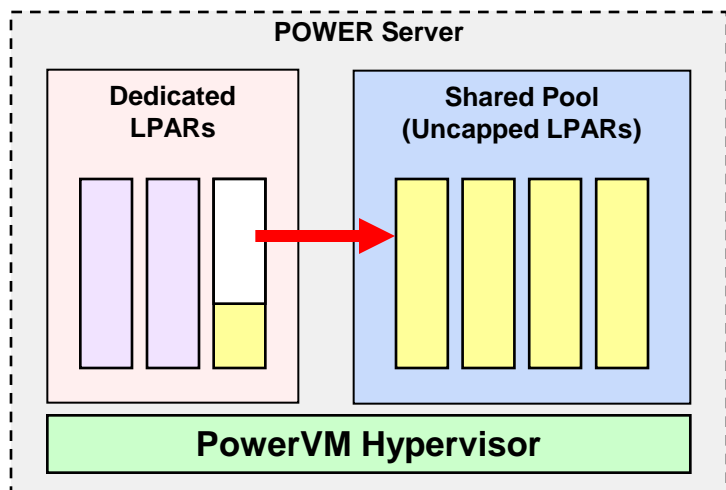
- 最小/最大

## ➤受限和非受限分区

## ➤处理能力自动共享

## ➤多个CPU池

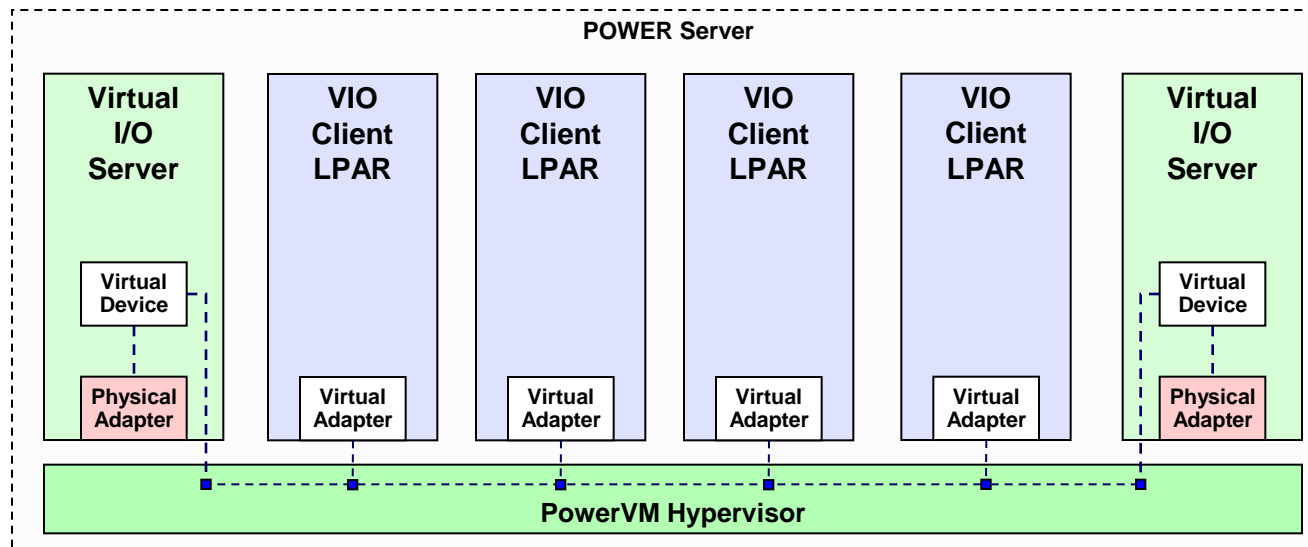
## ➤动态分区



## ➤可共享专有分区的CPU处理能力

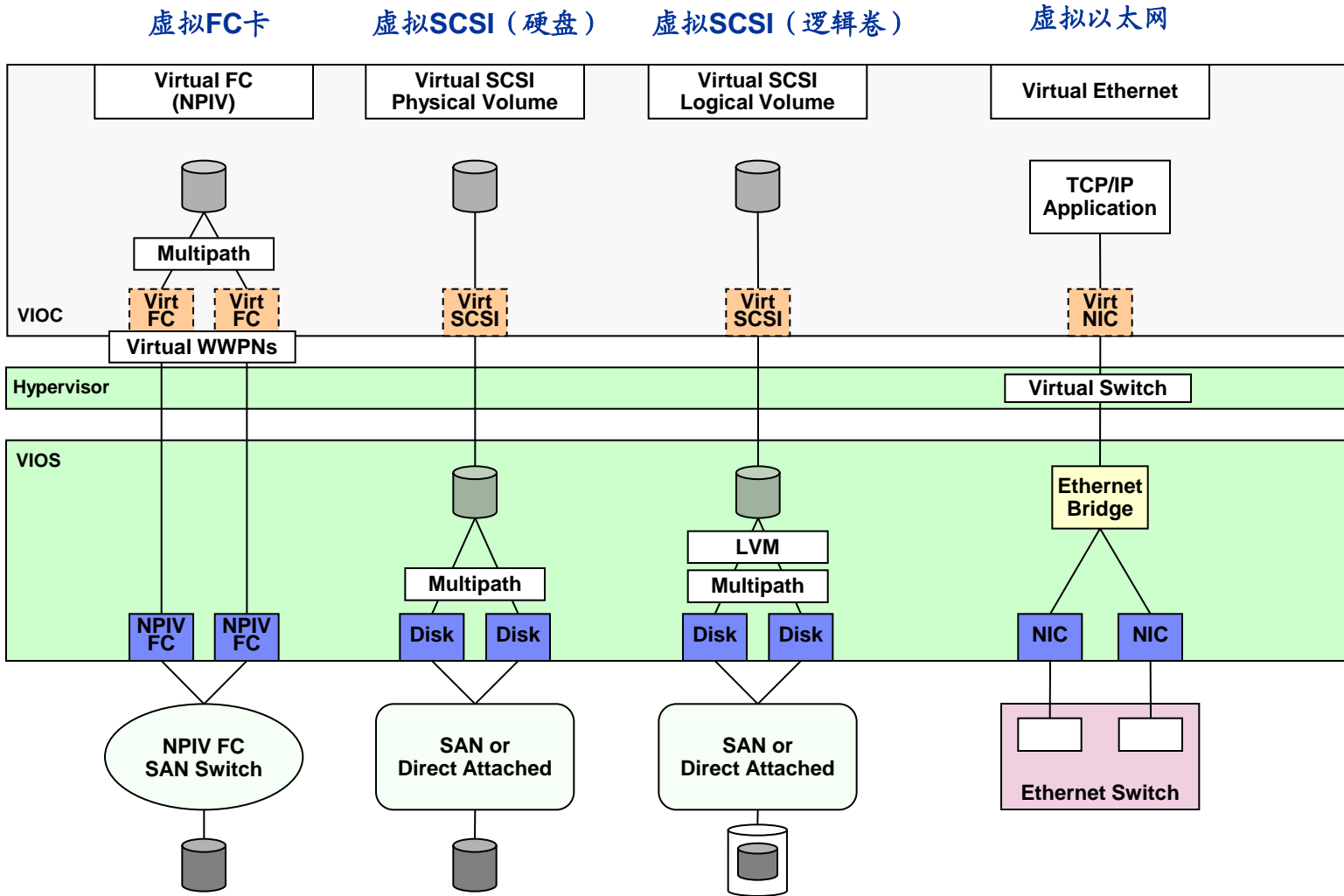
# I/O虚拟化 — Virtual I/O Server

- No I/O devices in Hypervisor
- Multiple VIOS LPAR support
  - High availability VIOS options
  - VIOS for specific purposes
  - VIOS sized to match workloads
- Performance
  - Virtual FC/SCSI - logical redirected DMA avoids data copies
  - PowerVM MTU sizes up to 65280 bytes
  - TCP/IP checksum bypass option
- Benefits
  - Better availability and security
  - Better performance



- VIO技术解放物理I/O对虚拟服务器的束缚
- 每个LPAR中，真实I/O和虚拟I/O可以并存。

# Virtual I/O 概览

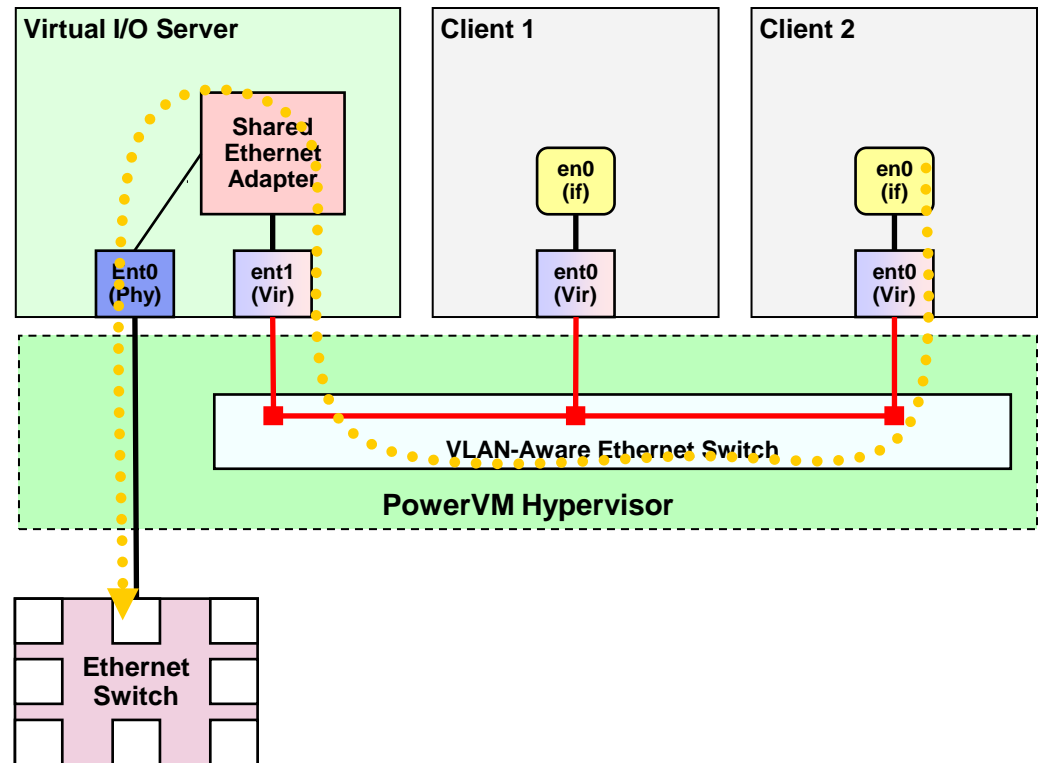


Note: Dual VIO server options are commonly used - these options are not shown above.

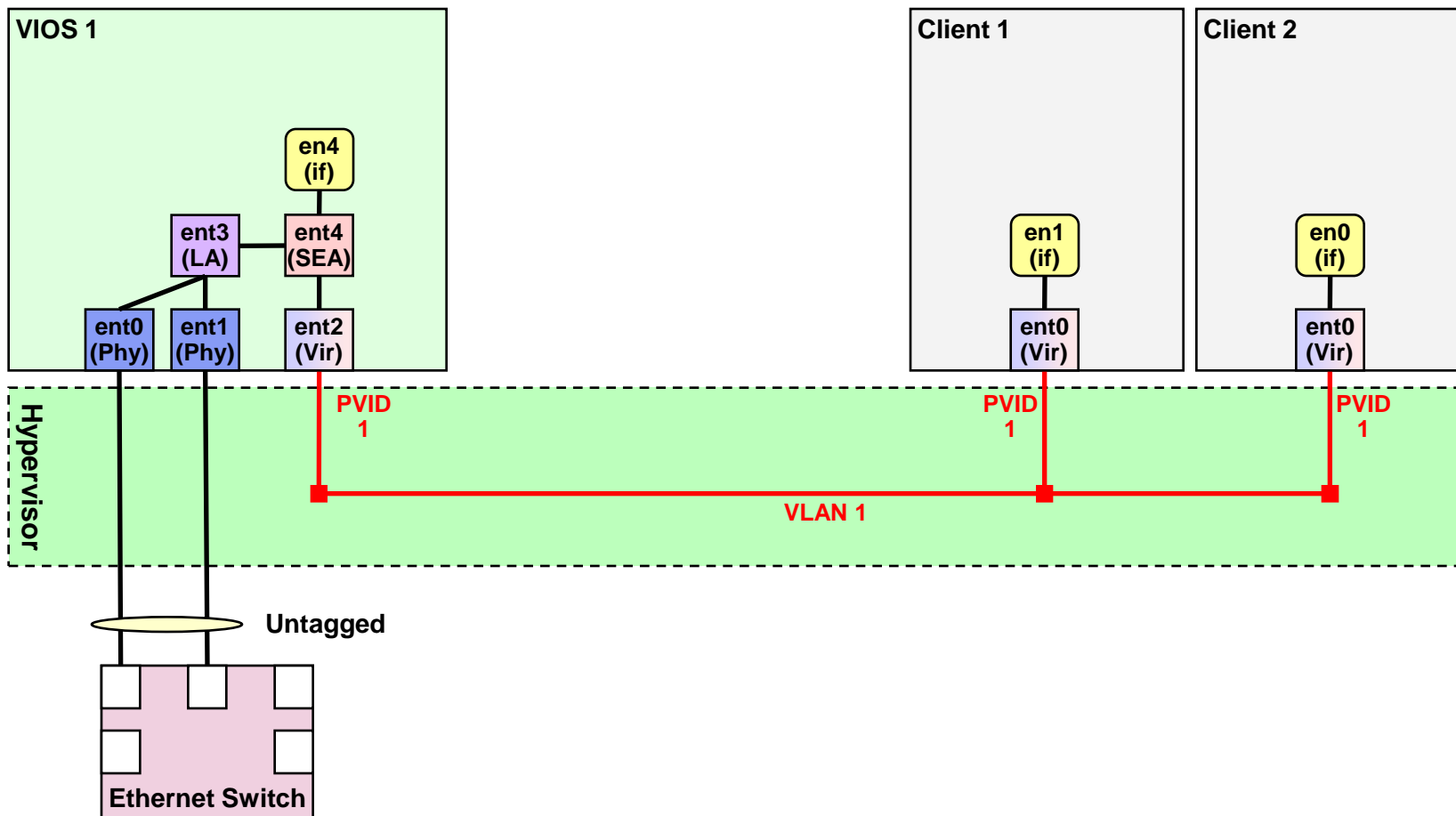


# I/O虚拟化 — 虚拟以太网

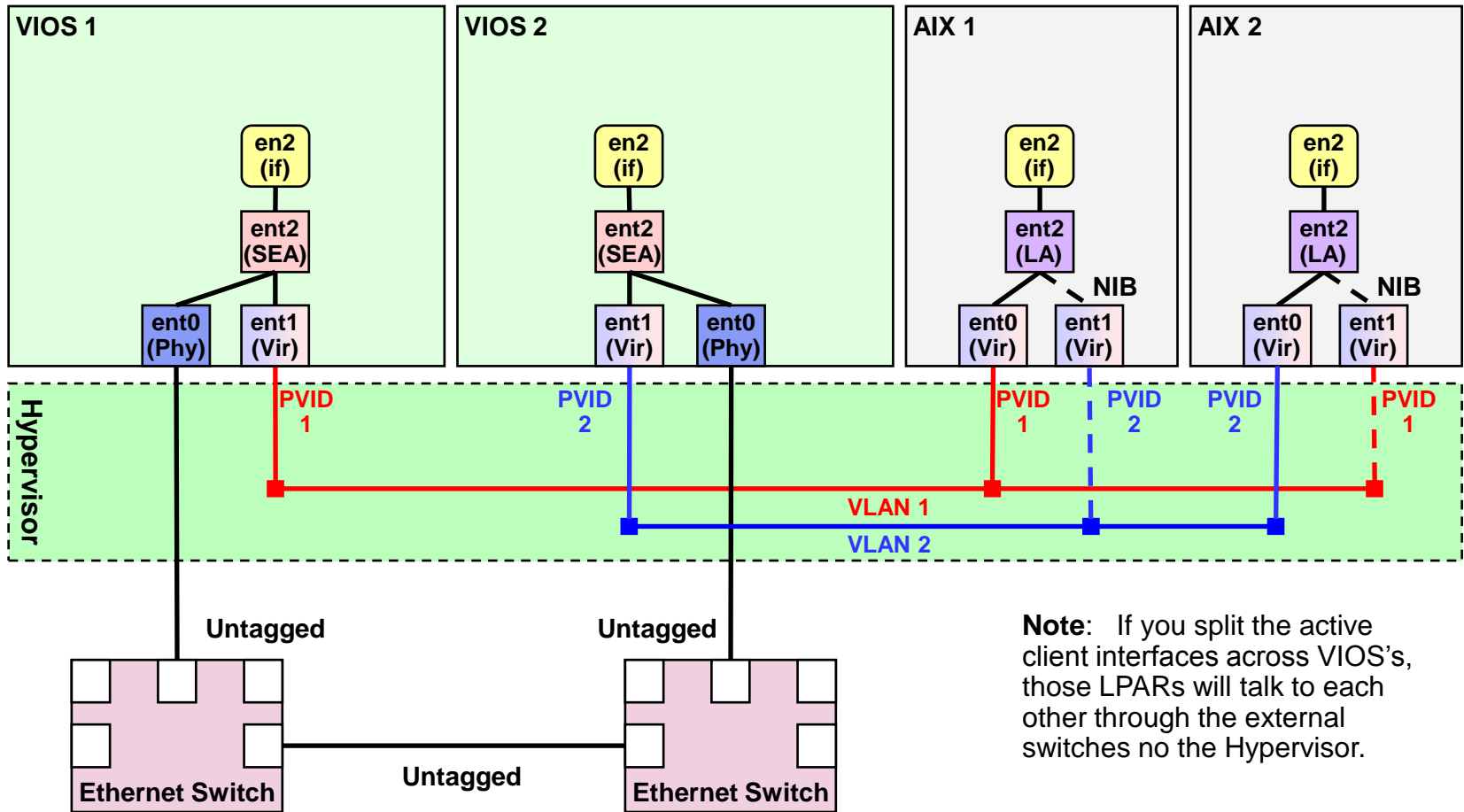
- PowerVM Ethernet switch
  - Part of PowerVM Hypervisor
  - Moves data between LPARs
  
- Shared Ethernet Adapter
  - Part of the VIO server
  - Logical device
  - Bridges traffic to and from external networks
  
- Additional capabilities
  - VLAN aware
  - Link aggregation for external networks



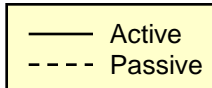
# 单VIO服务器，VIOS内双网卡备份



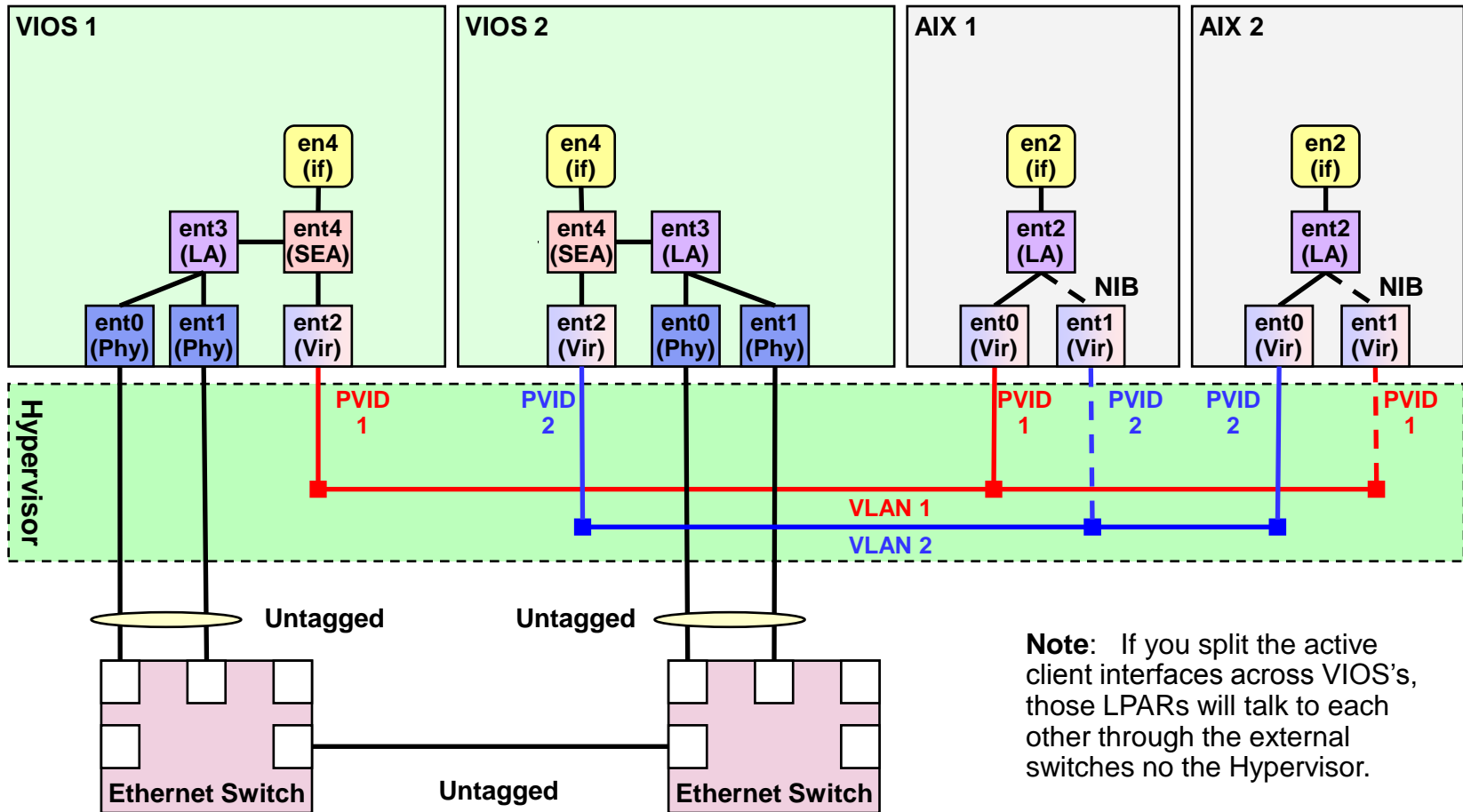
# 双VIO服务器，分区内双网卡备份



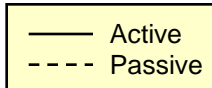
**Note:** If you split the active client interfaces across VIOS's, those LPARs will talk to each other through the external switches no the Hypervisor.



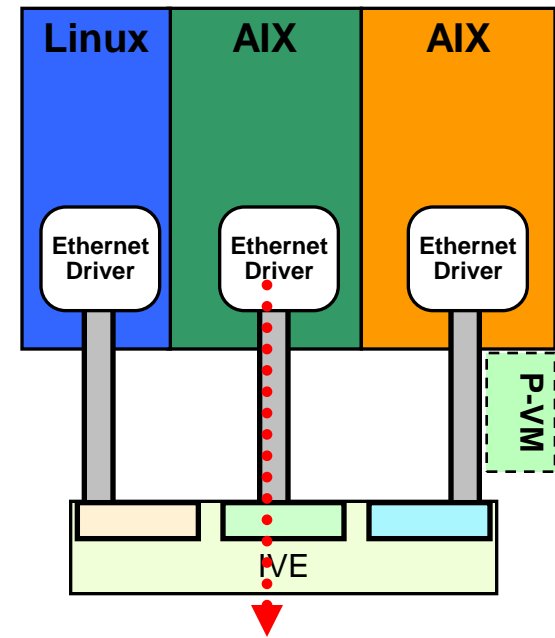
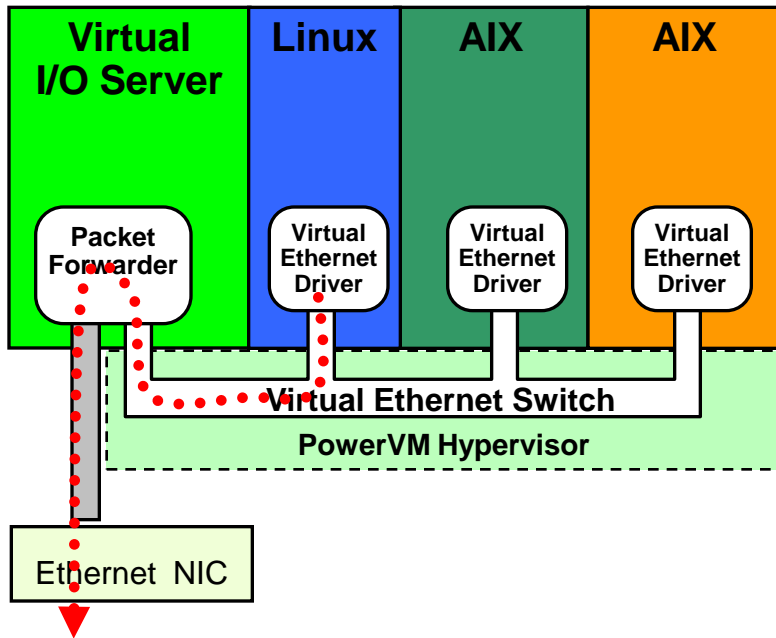
# 双VIO服务器， VIOS内双网卡备份。分区内双网卡备份



**Note:** If you split the active client interfaces across VIOS's, those LPARs will talk to each other through the external switches no the Hypervisor.

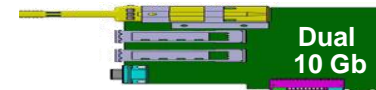
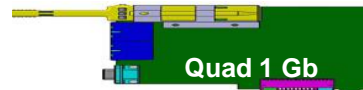
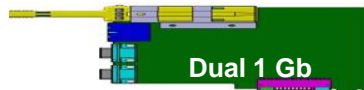


# 集成的虚拟网卡(IVE — Integrated Virtual Ethernet)

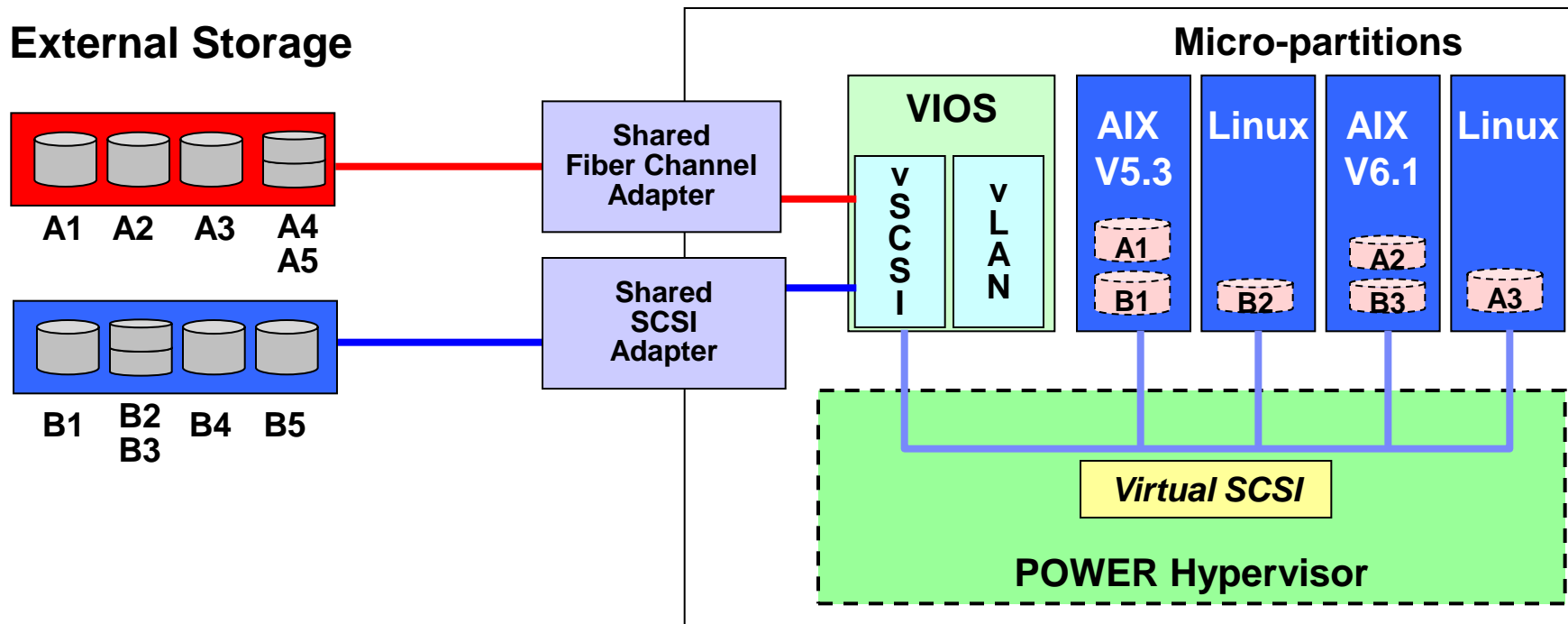


- Virtual I/O Server
  - Allows sharing of PCI Ethernet adapters
  - Some small amount of CPU time spent by packet forwarder

- Integrated Virtual Ethernet
  - Removes SW packet forwarding function of VIO server
  - Provides equivalent performance to a dedicated Ethernet adapter



# I/O虚拟化 — 虚拟SCSI



VIOS 拥有物理的磁盘资源

- 在VIO Server上的存储划分LVM
- 存储设备可以通过光纤或者SCSI连接

分区把这些磁盘看成虚拟SCSI设备 (Virtual SCSI)

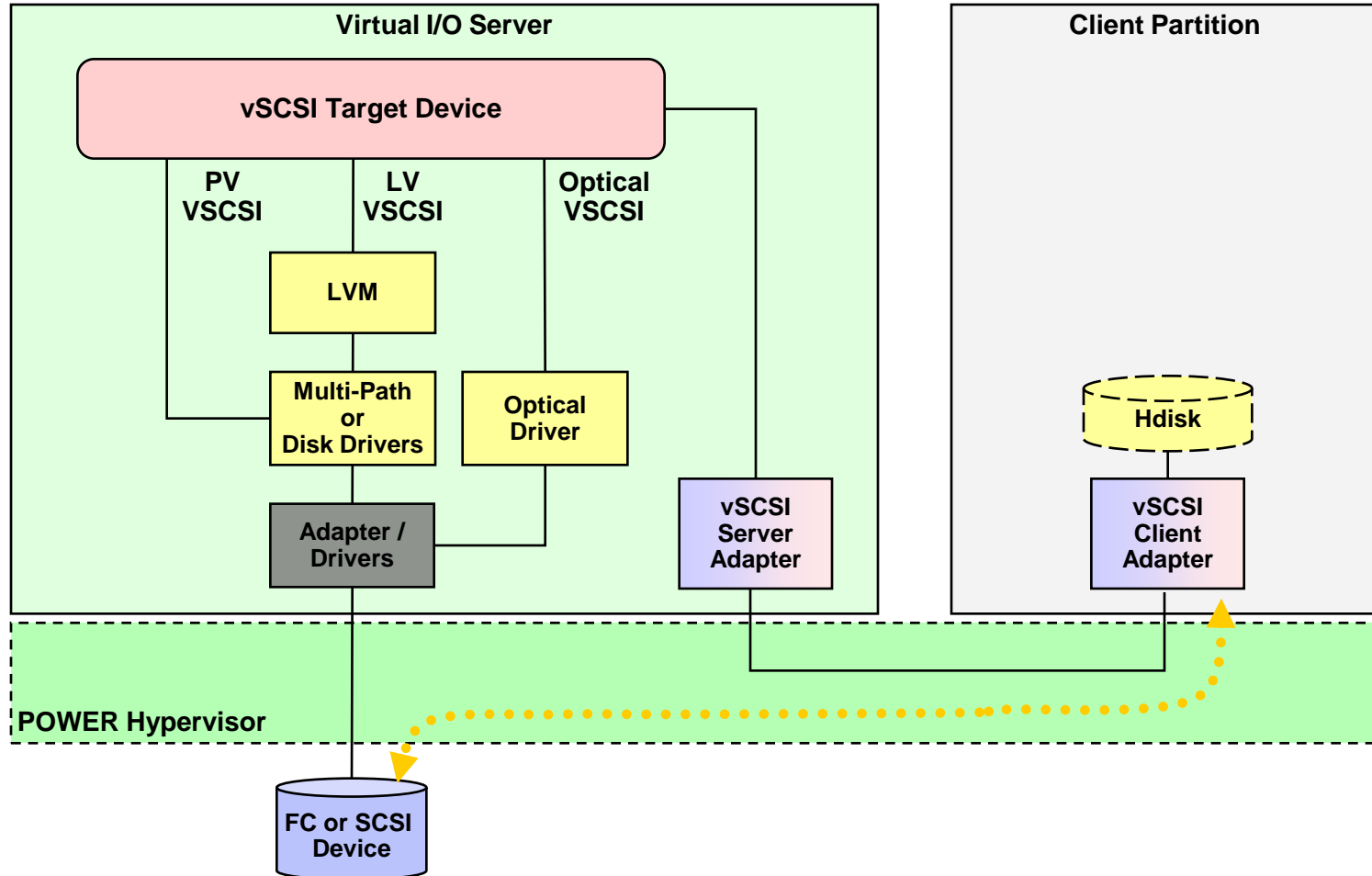
- 通过HMC把虚拟设备配置在分区上
- 分区上配置这些硬盘是虚拟硬盘
- VIOS需要一定的配置

多个分区可以使用相同和不同的物理硬盘

- 在VIOS配置成逻辑卷
- 在分区上显示为一个hdisk
- 可以配置一整块物理硬盘给一个分区使用

Virtual I/O 通过共享磁盘减少硬件投资

# 虚拟 SCSI 基本架构



- No I/O devices in Hypervisor
- Virtual FC/SCSI - logical redirected DMA avoids data copies



## 虚拟I/O服务器(VIOS)多路径支持

- **IBM XIV Storage System**
- **2105 Enterprise Server (800, 750, Fxx)**
- **IBM TotalStorage SAN Volume Controller**
- **DS83000**
  - 2107 Model 921 IBM System Storage DS8100
  - 2107 Model 922 IBM System Storage DS8300
  - 2107 Model 9A2 IBM System Storage DS8300
  - 2107 Model 92E IBM System Storage DS8000 Expansion Unit
  - 2107 Model 9AE IBM System Storage DS8000 Expansion Unit
- **DS6000**
  - 1750 Model 511 IBM System Storage DS6800
  - 1750 Model EX1 IBM System Storage DS6000 Expansion Unit
- **NSeries**
  - IBM System Storage N3700, N5000 and N7000 (Fibre Channel and iSCSI attach)
- **DS5000**
  - IBM System Storage DS5100
  - IBM System Storage DS5300
- **DS4000**
  - IBM System Storage DS4800
  - IBM System Storage DS4700
  - IBM TotalStorage DS4500 (FAStT900)
  - IBM TotalStorage DS4400 (FAStT700)
  - IBM TotalStorage DS4300 (FAStT600)
  - IBM System Storage DS4200
  - IBM TotalStorage DS4100 (FAStT100)
- **FastT**
  - FAStT200 Storage Server
  - FAStT500 Storage Server
- **DS3000**
  - IBM System Storage DS3400
- **Under-the-cover IBM parallel SCSI devices**

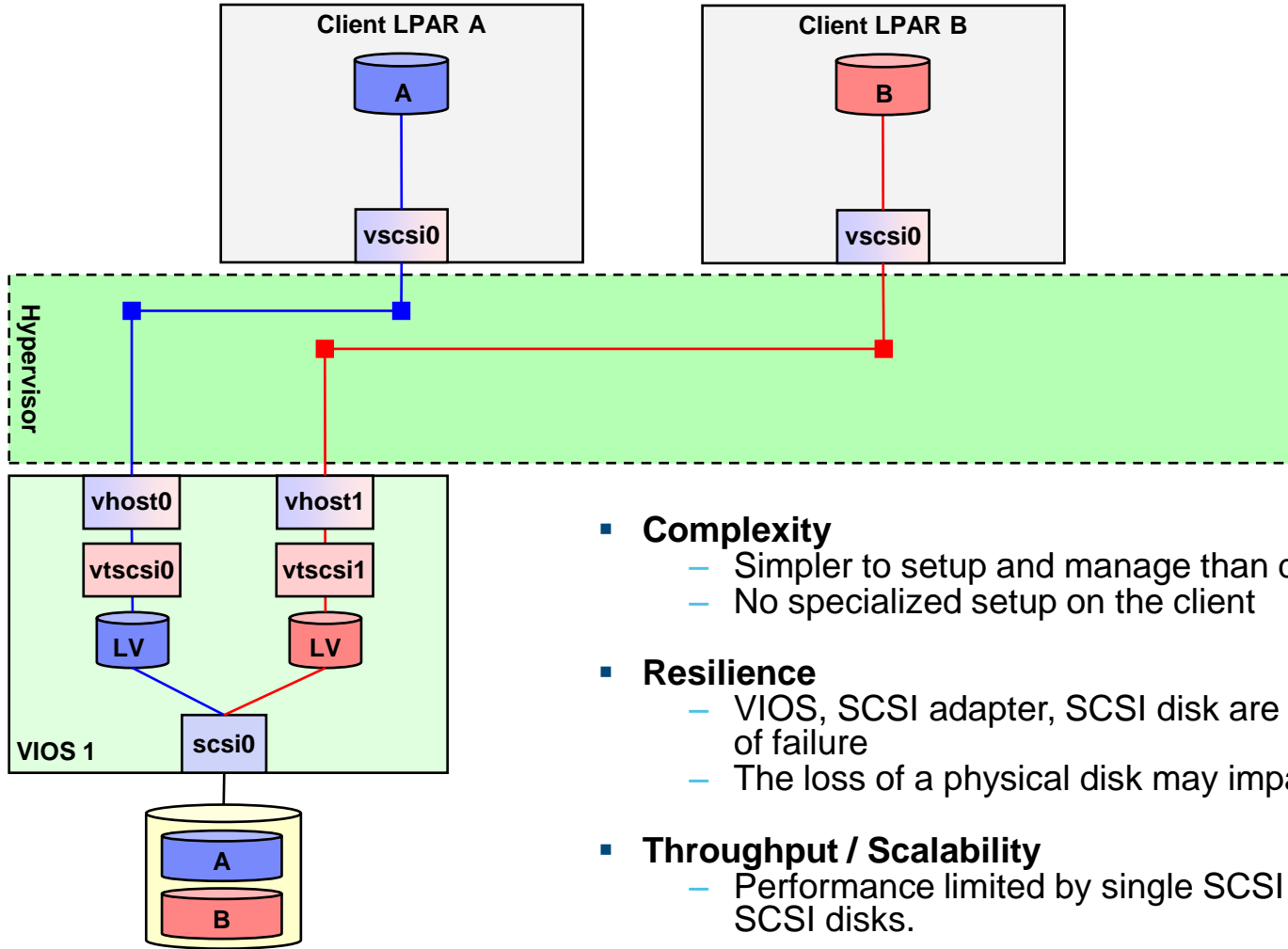
<http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html#multipath>

## 虚拟I/O服务器(VIOS)多路径支持

- The Virtual I/O Server has been tested on selected configurations, which include specific models from the following disk subsystems suppliers:
  - EMC
  - HDS
  - HP
  - NetApp
  
- Contact your storage vendor for model details.

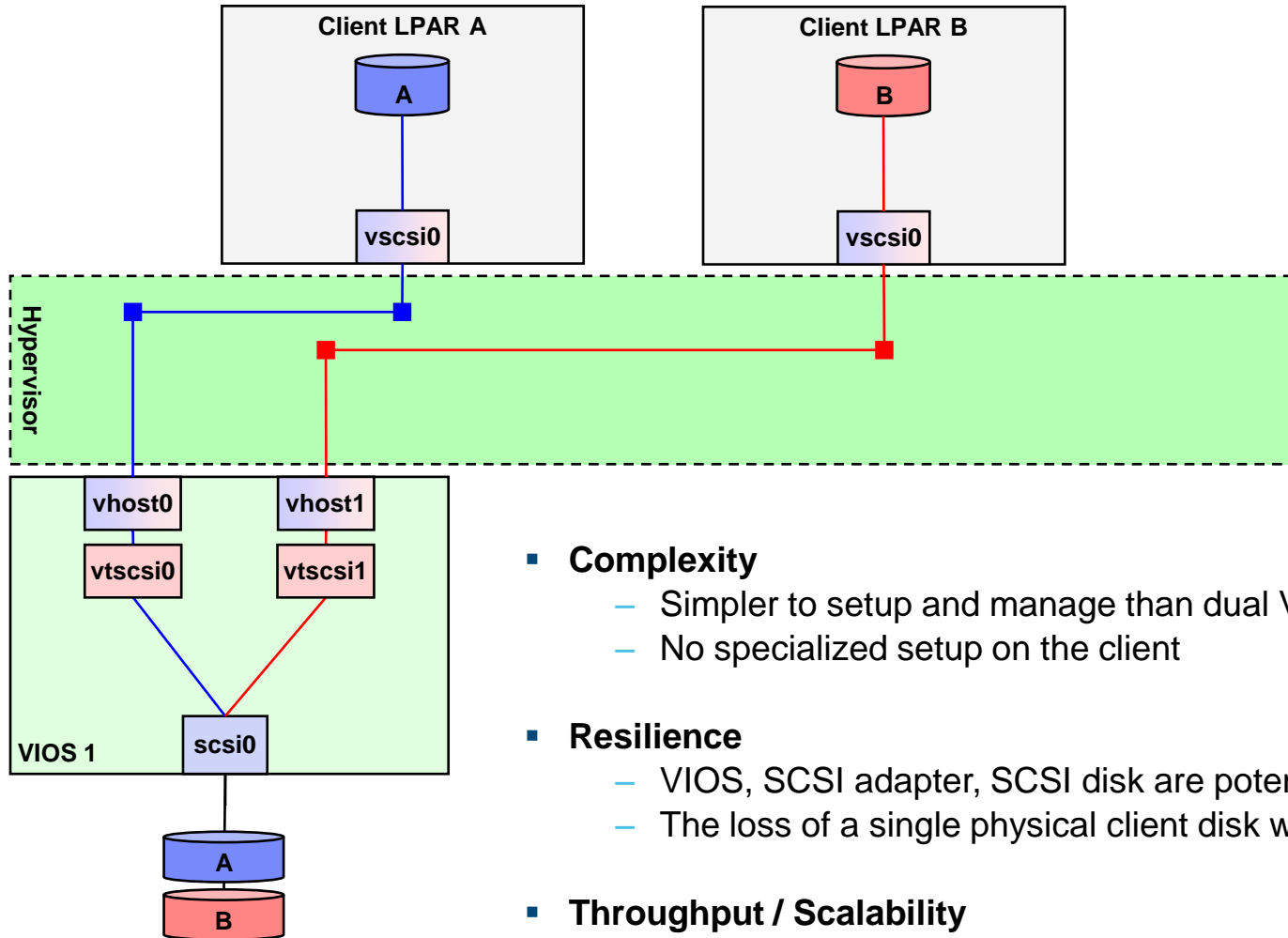
<http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html#multipath>

# 单VIO服务器，逻辑卷(LV)映射成虚拟SCSI盘



- **Complexity**
  - Simpler to setup and manage than dual VIOS
  - No specialized setup on the client
- **Resilience**
  - VIOS, SCSI adapter, SCSI disk are potential single points of failure
  - The loss of a physical disk may impact more than one client
- **Throughput / Scalability**
  - Performance limited by single SCSI adapter and internal SCSI disks.
- **Notes**
  - Low cost disk alternative

# 单VIO服务器, 硬盘(PV)映射成虚拟SCSI盘



- **Complexity**

- Simpler to setup and manage than dual VIOS
- No specialized setup on the client

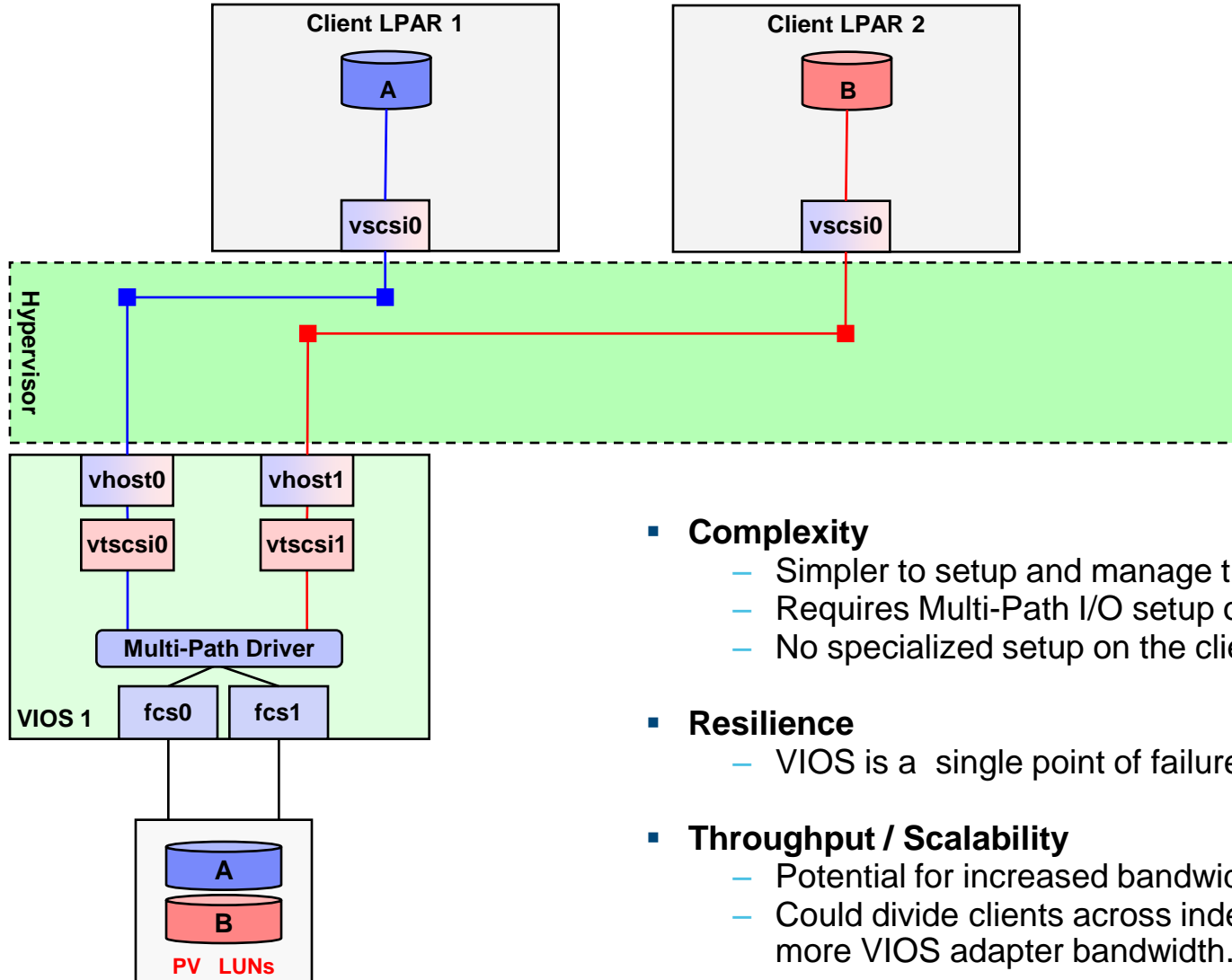
- **Resilience**

- VIOS, SCSI adapter, SCSI disk are potential single points of failure
- The loss of a single physical client disk will affect only that client

- **Throughput / Scalability**

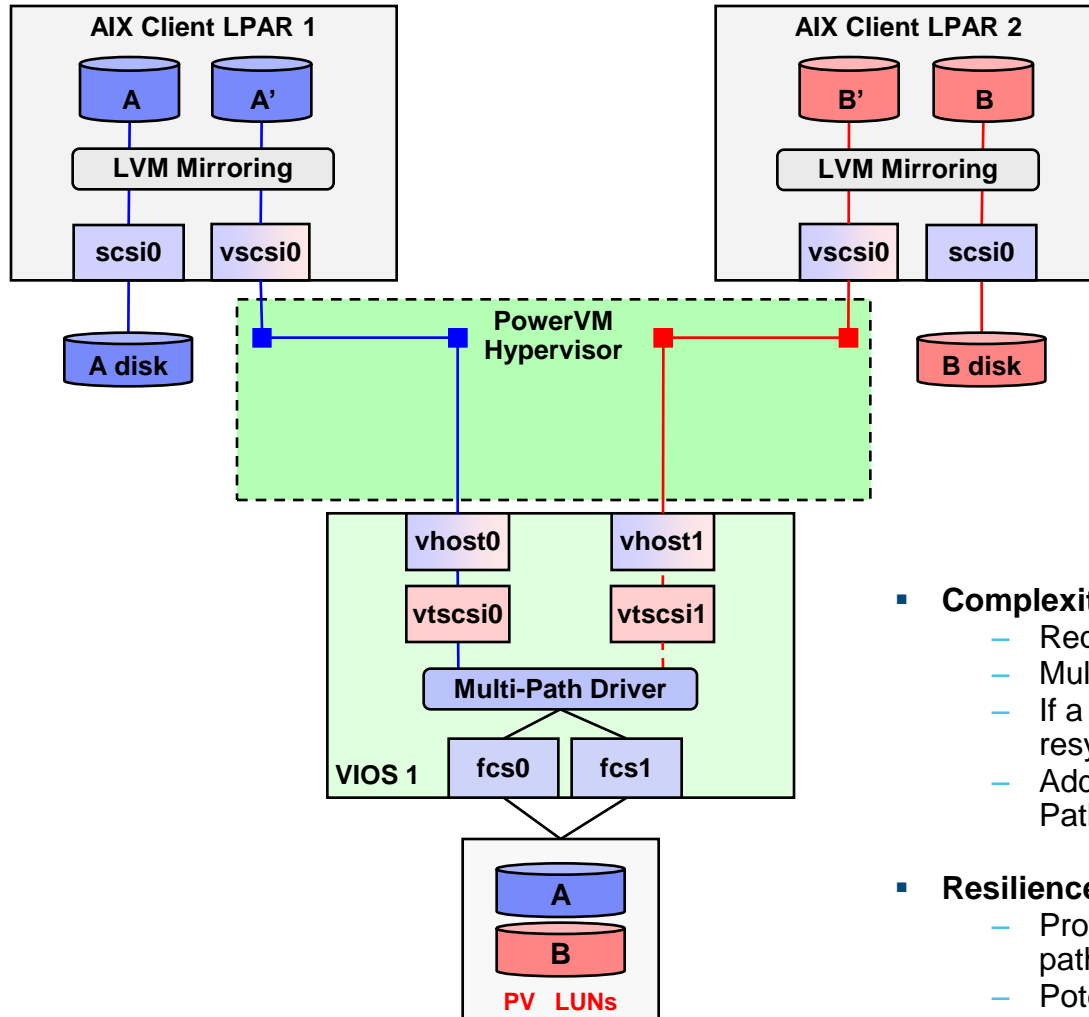
- Performance limited by single SCSI adapter and internal SCSI disks.

# 单VIO服务器, I/O多路径



- **Complexity**
  - Simpler to setup and manage than dual VIO servers
  - Requires Multi-Path I/O setup on the VIOS
  - No specialized setup on the client
- **Resilience**
  - VIOS is a single point of failure
- **Throughput / Scalability**
  - Potential for increased bandwidth due to multi-path I/O.
  - Could divide clients across independent VIOS allowing more VIOS adapter bandwidth.

# 分区内真实硬盘与虚拟硬盘的逻辑卷镜像(LVM Mirroring)



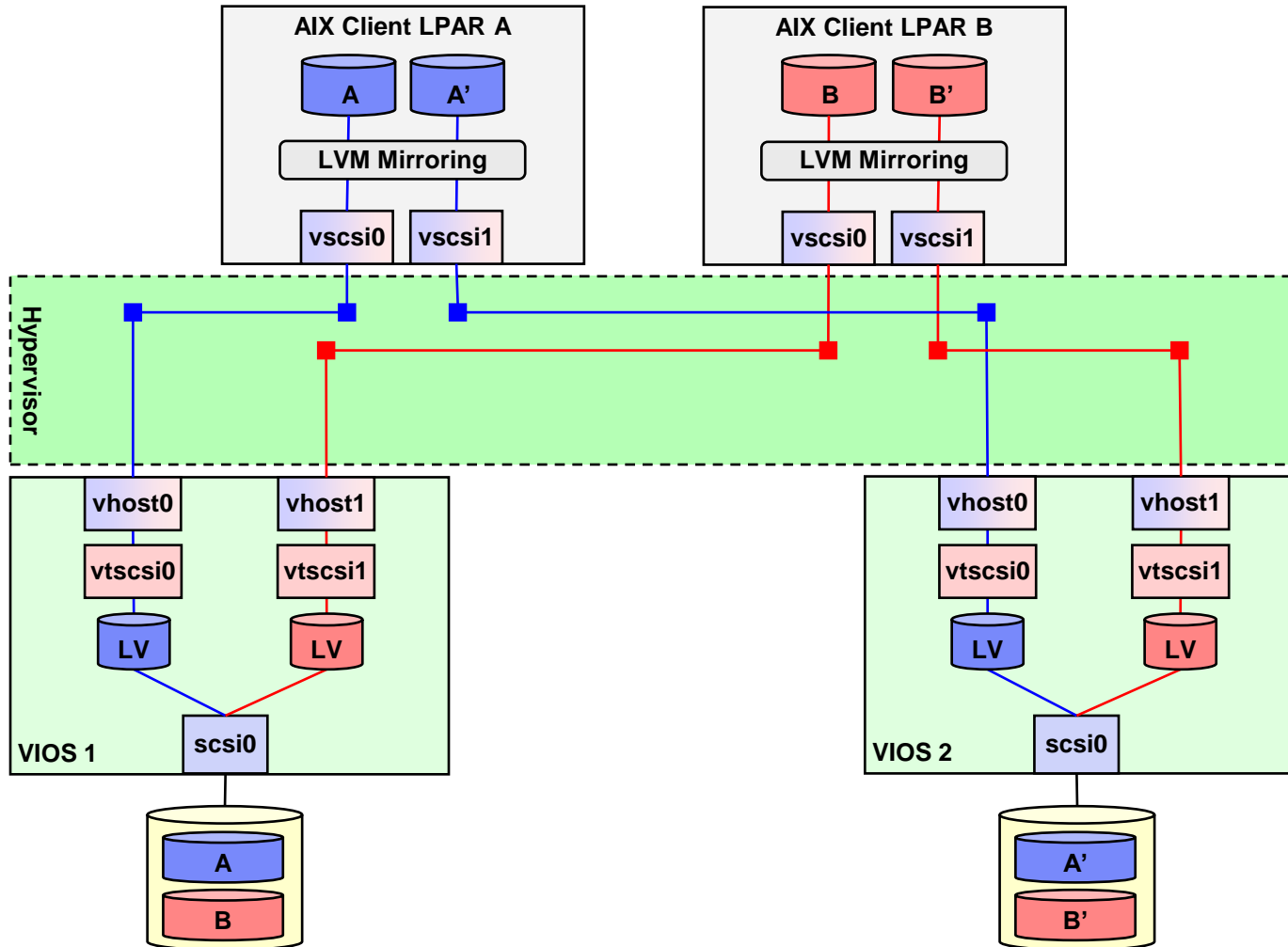
## Complexity

- Requires LVM mirroring to be setup on the VIOC
- Multi-Path I/O setup on the VIOS
- If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- Additional complexity due to multiple disk types, Multi-Path I/O setup, and client mirroring

## Resilience

- Protection against failure of single adapter failure (or path) or disk
- Potential protection against FC adapter failures within VIOS (if Multi-Path I/O is configured)

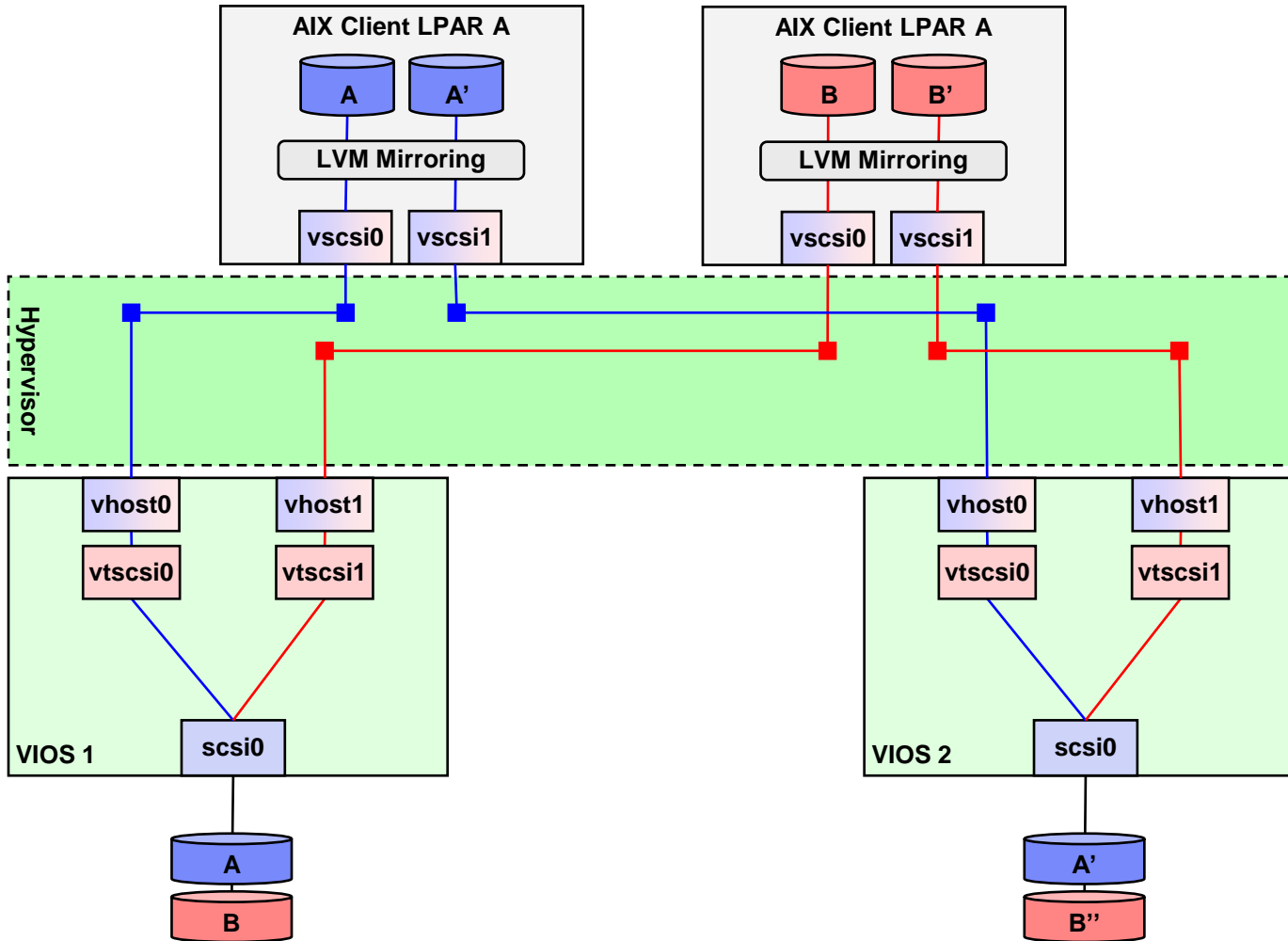
# 双VIO服务器, 逻辑卷(LV)映射成虚拟SCSI盘。分区内逻辑卷镜像



- **Complexity**
  - More complicated than single VIO server but does not require SAN ports or setup
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC.
- **Resilience**
  - Protection against failure of single VIOS / SCSI disk / SCSI controller.
- **Throughput / Scalability**
  - VIOS performance limited by single SCSI adapter and internal SCSI disks.

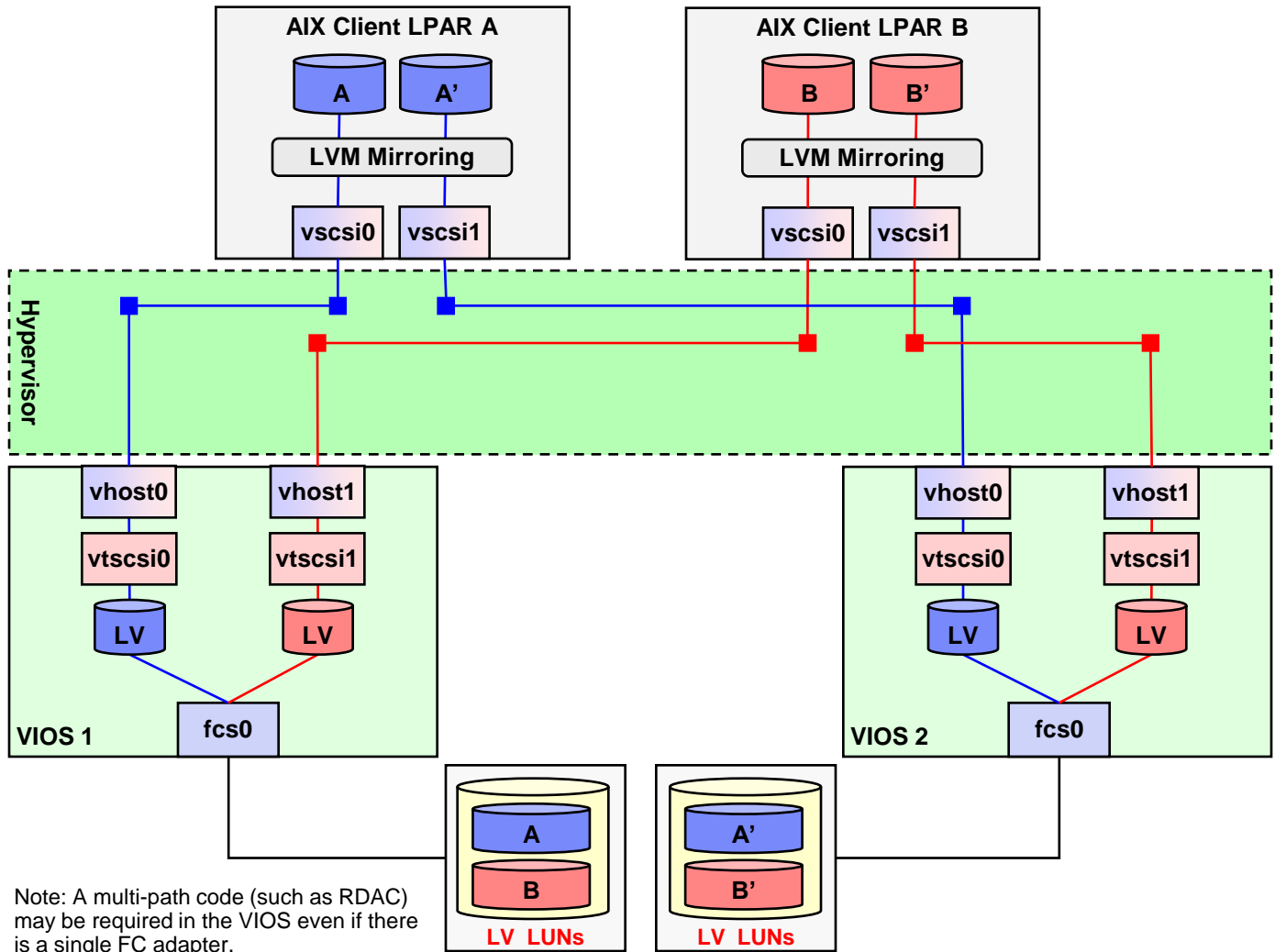


# 双VIO服务器, 硬盘(PV)映射成虚拟SCSI盘。分区内逻辑卷镜像



- **Complexity**
  - More complicated than single VIO server but does not require SAN ports or setup
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
  
- **Resilience**
  - Protection against single VIOS / SCSI disk / SCSI controller
  - The loss of a single physical disk would affect only one client
  
- **Throughput / Scalability**
  - VIOS performance limited by single SCSI adapter and internal SCSI disks.

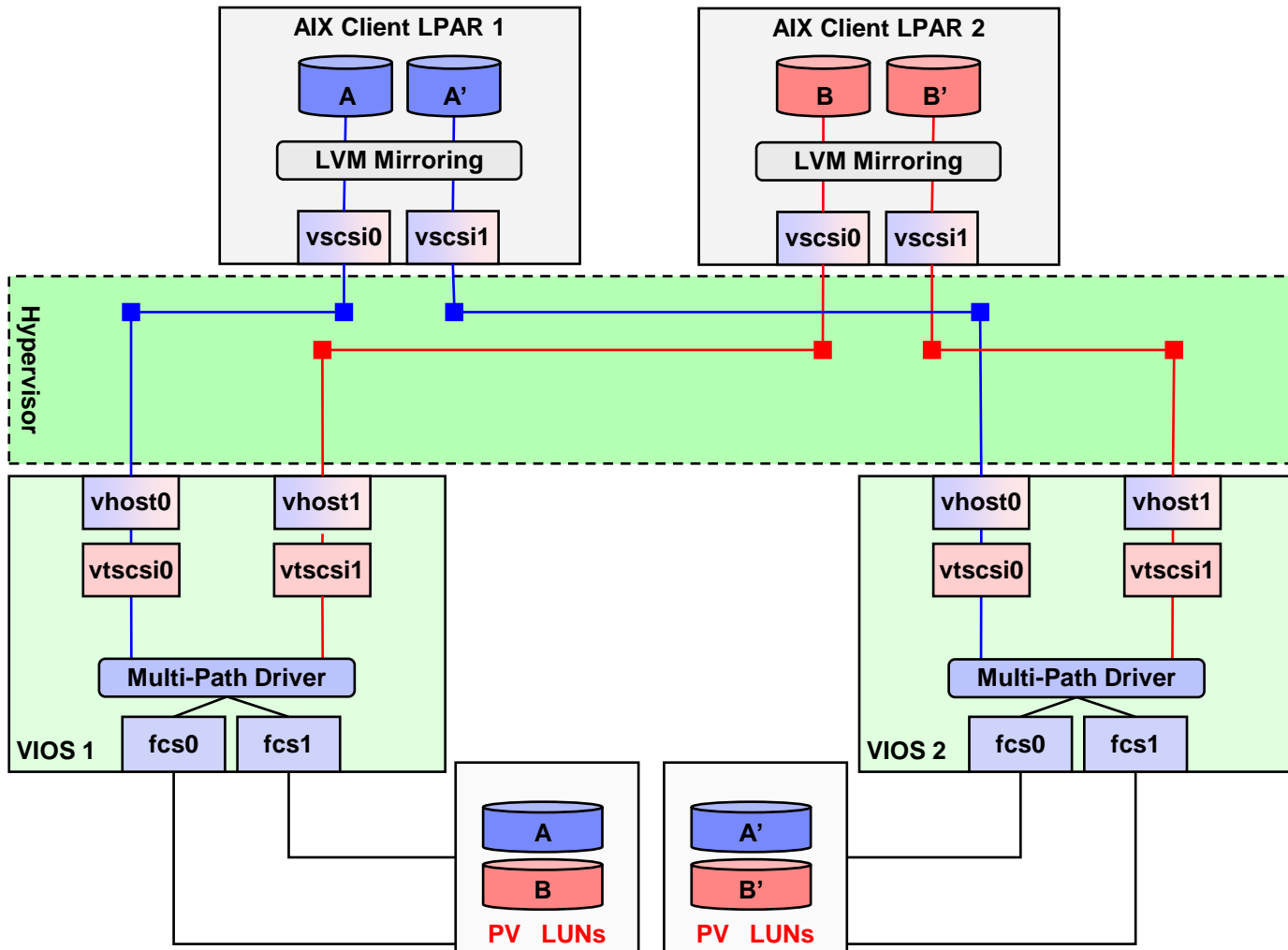
# 双VIO服务器, I/O单路径, LV或PV映射成虚拟SCSI盘。 分区内逻辑卷镜像



Note: A multi-path code (such as RDAC) may be required in the VIOS even if there is a single FC adapter.

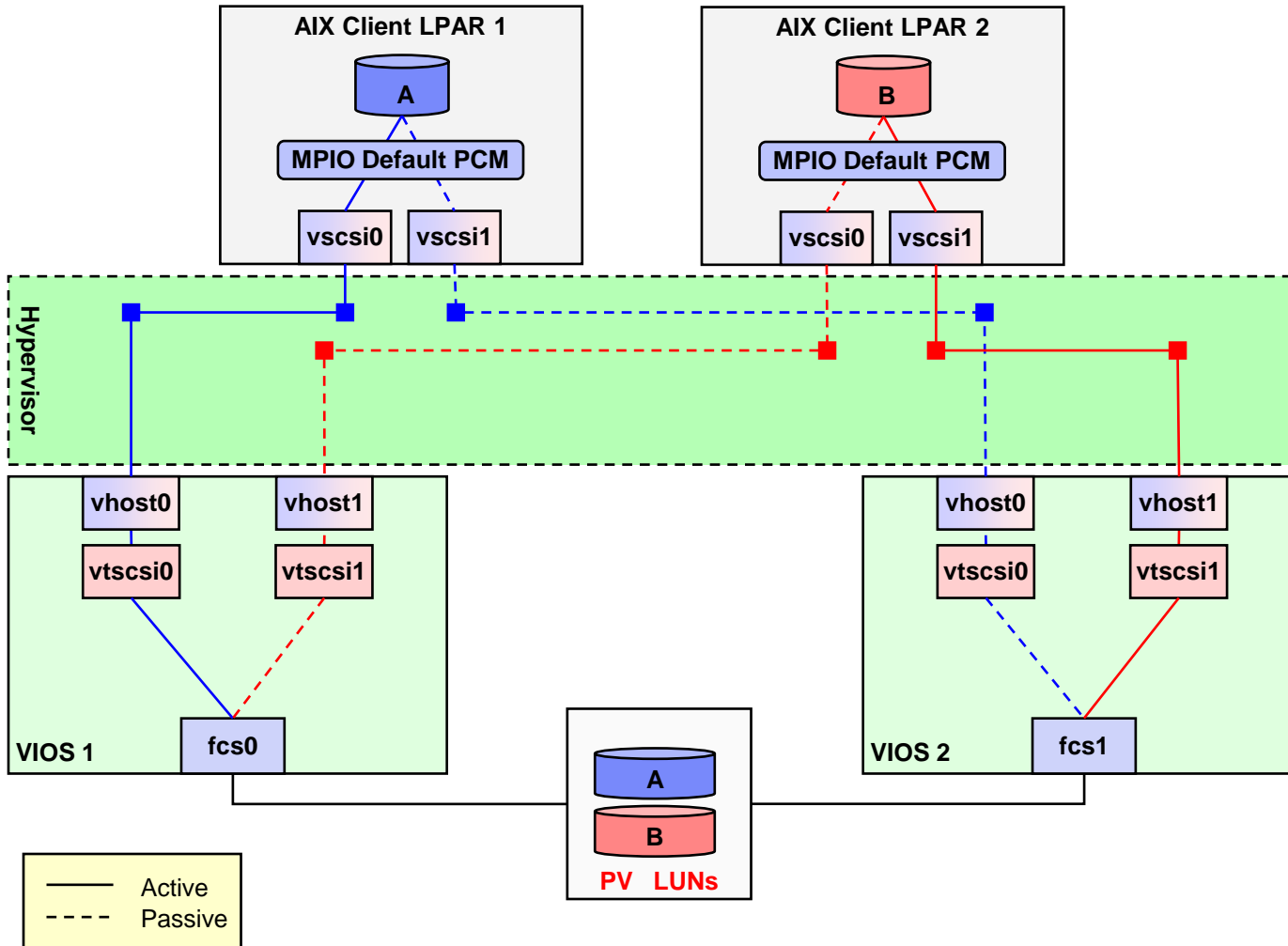
- **Complexity**
  - Requires SAN configuration
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- **Resilience**
  - Protection against failure of single VIOS / FC adapter (or path)
  - No protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Performance limited by a single FC adapter.
- **Notes**
  - Disks must only be seen by one VIO server.
  - LV VSCSI LUNs could also be PV VSCSI LUNs.

# 双VIO服务器, I/O多路径, LV或PV映射成虚拟SCSI盘。 分区内逻辑卷镜像



- **Complexity**
  - Requires LVM mirroring to be setup on the VIOC
  - Requires Multi-Path I/O setup on the VIOS
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- **Resilience**
  - Protection against failure of single VIOS / FC adapter failure (or path)
  - Protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to multi-path I/O
- **Notes**
  - LUNs used for this purpose can only be assigned to a single VIOS
  - LV VSCSI LUNs could also be PV VSCSI LUNs.

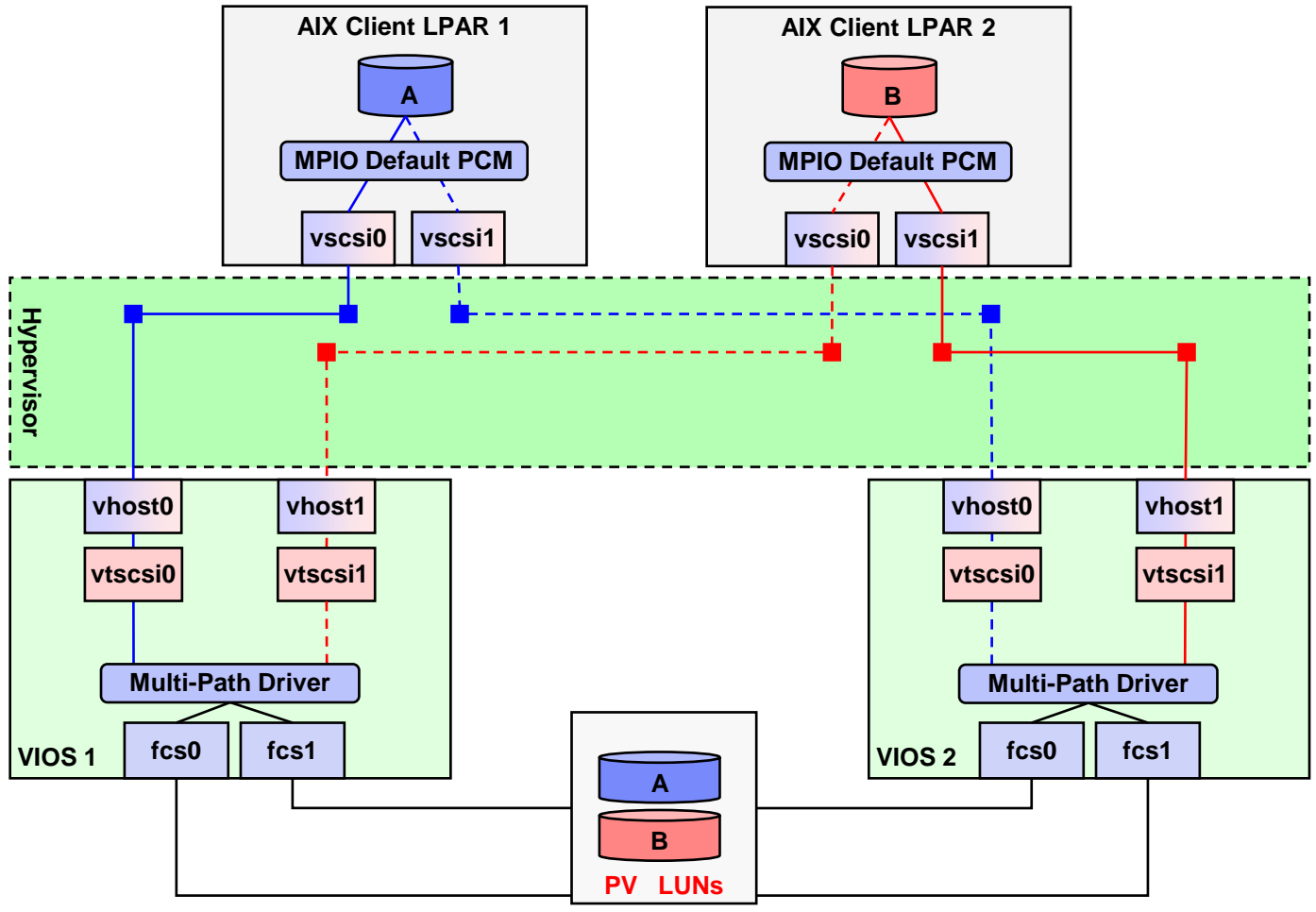
# 双VIO服务器, I/O单路径, LV或PV映射成虚拟SCSI盘。 分区内MPIO实现硬盘冗余



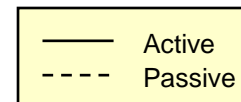
- **Complexity**
  - Simplest dual VIOS FC option
  - Requires MPIO to be setup on the client
- **Resilience**
  - Protection against failure of a single VIOS / FC adapter (or path)
- **Throughput / Scalability**
  - Primary LUNs can be split across multiple VIOS to help balance the I/O load
- **Notes**
  - Lowest port costs of dual VIOS FC options
  - Must be PV VSCSI disks.

Note: A multi-path code (such as RDAC) may be required in the VIOS even if there is a single FC adapter.

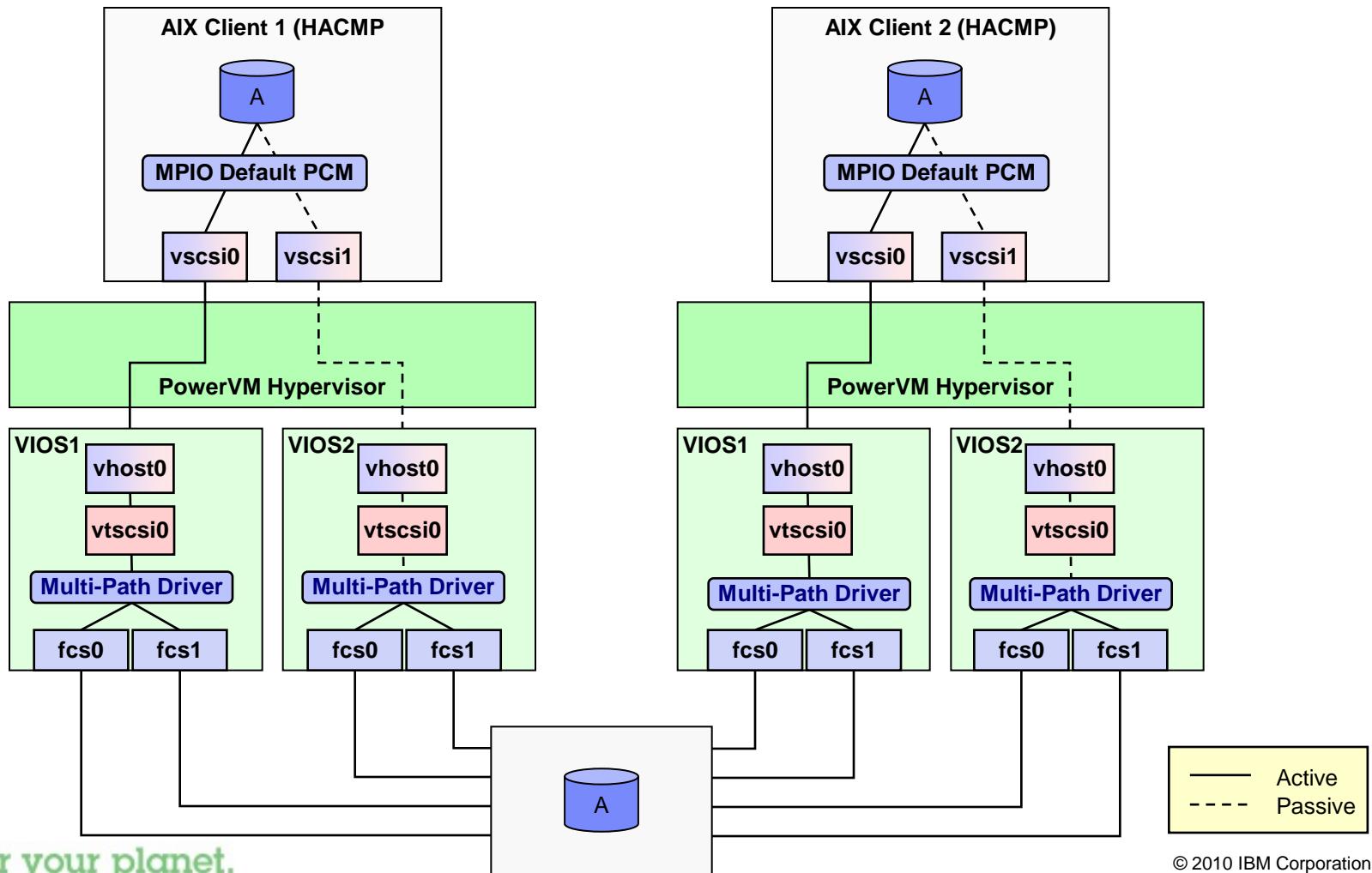
# 双VIO服务器, I/O多路径, LV或PV映射成虚拟SCSI盘。 分区内MPIO实现硬盘冗余



- **Complexity**
  - Requires MPIO to be setup on the client
  - Requires Multi-Path I/O setup on the VIOS
- **Resilience**
  - Protection against failure of a single VIOS, FC adapter, or path.
  - Protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to Multi-Path I/O
  - Primary LUNs can be split across multiple VIOS to help balance the I/O load.
- **Notes**
  - Must be PV VSCSI disks.



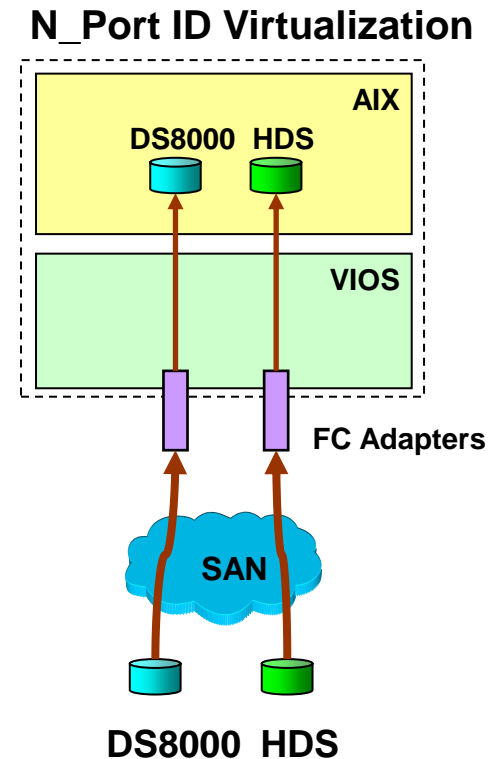
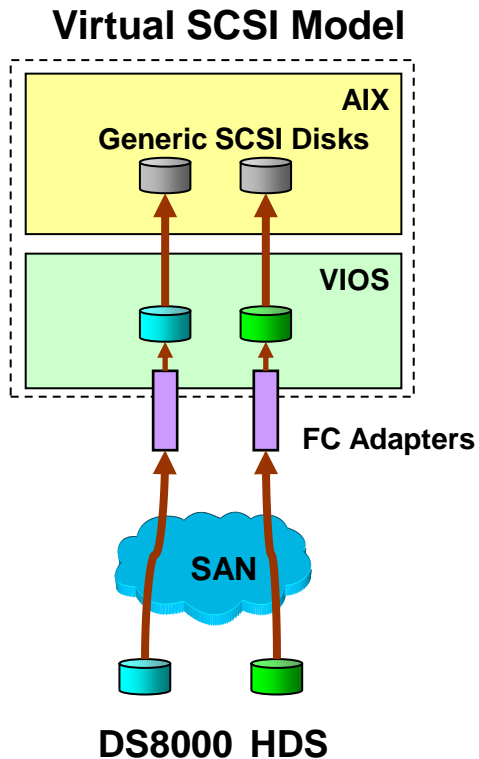
# 双VIO服务器, I/O多路径, LV或PV映射成虚拟SCSI盘。 分区内MPIO实现硬盘冗余 双机PowerHA (HACMP)



# NPIV简化SAN管理 (N\_Port ID Virtualization)

## ■ N\_Port ID Virtualization

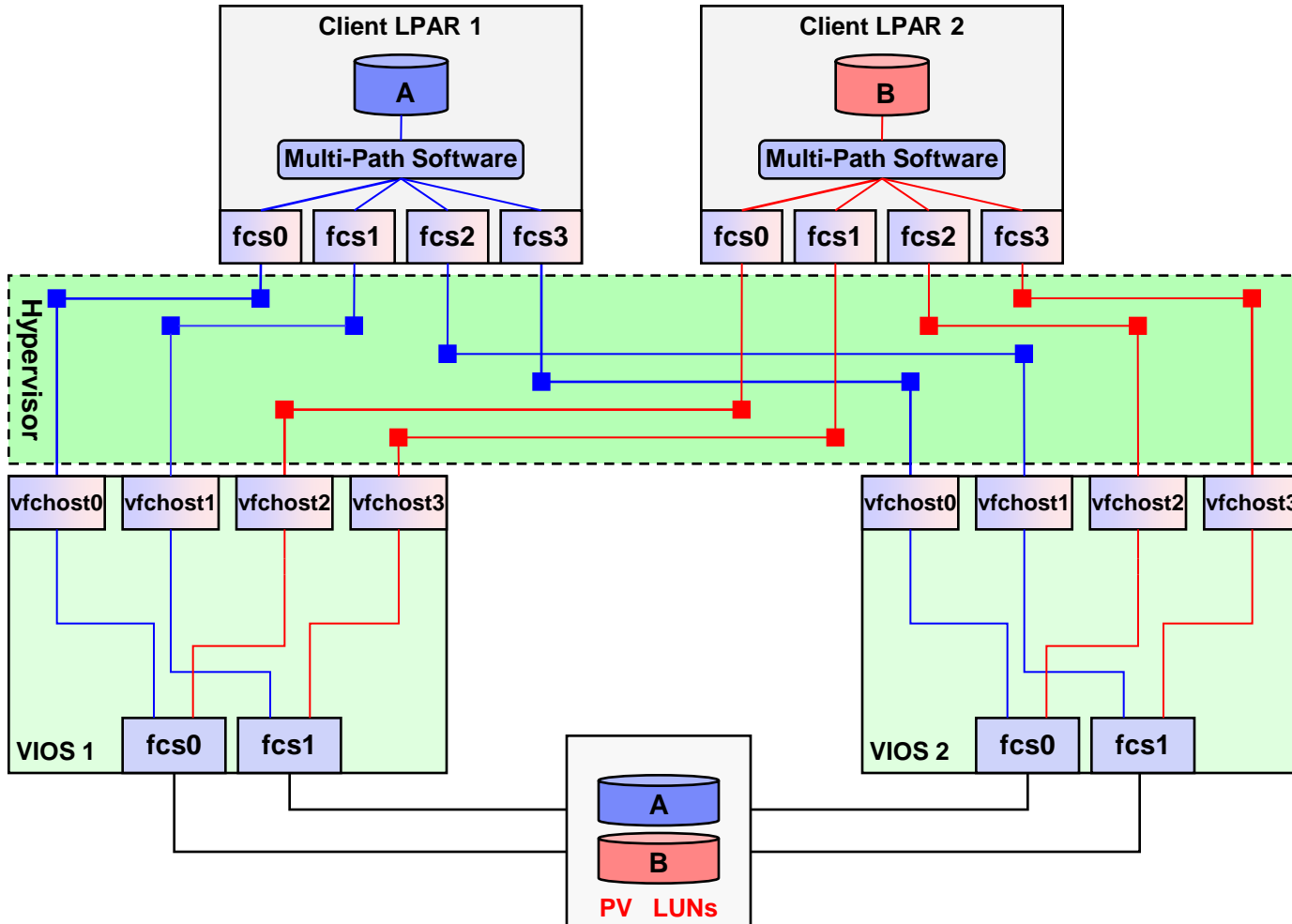
- Multiple Virtual World Wide Port Names per FC port – PCIe 8 Gb adapter
- LPARs have direct visibility on SAN (Zoning/Masking)
- I/O Virtualization configuration effort is reduced





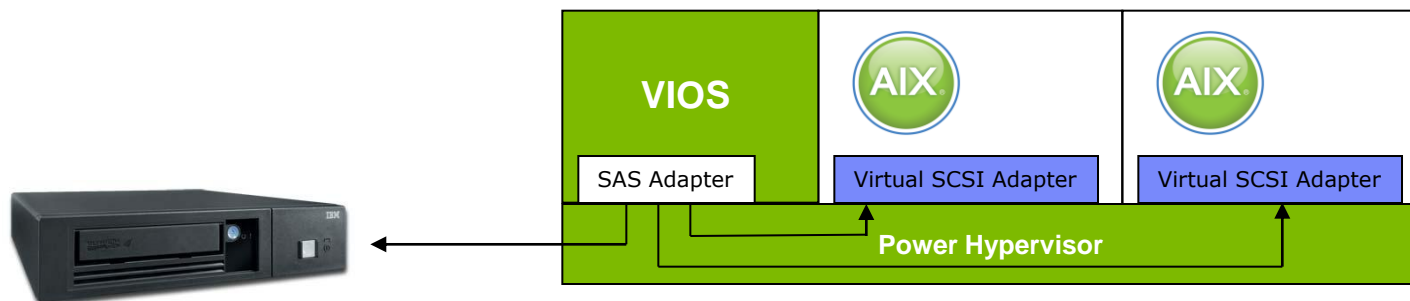
# NPIV虚拟化

## 分区中直接使用虚拟的FC卡



## VIOS Virtual Tape Support

- Enables client partitions to directly access selected SAS tape devices, sharing resources and simplifying backup & restore operations
  - SAS adapter is owned by VIOS partition
  - Included with PowerVM Express, Standard, or Enterprise Edition
  - Supports AIX 5.3 & 6.1 partitions
  - POWER6 processor-based systems



### Tape drives supported

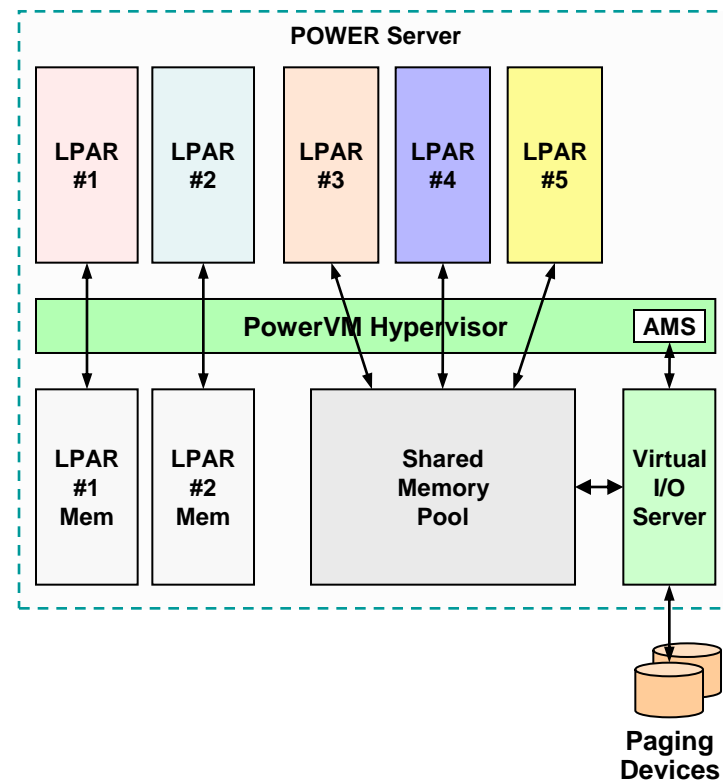
- DAT72: Feature Code 5907
- DAT160: Feature Code 5619
- HH LTO4: Feature Code 5746

- Statement of Direction
  - IBM intends to support VIOS virtual tape capabilities with IBM i and Linux environments in 2009.

## 内存虚拟化 — 动态内存共享(Active Memory Sharing)

AMS可以动态地把内存从一个分区移动到另一个分区，从而提高内存的使用率和灵活性。

- Power系统上的内存虚拟化技术
  - 内存根据分区工作负载自动分配
  - 分区使用VIOS的paging设备
  - 提高内存的使用率
- 适合三种不同的内存使用场景
  - 平均内存使用量很少
  - 激活/非激活分区
  - 分区负载的高峰时间在不同的时段

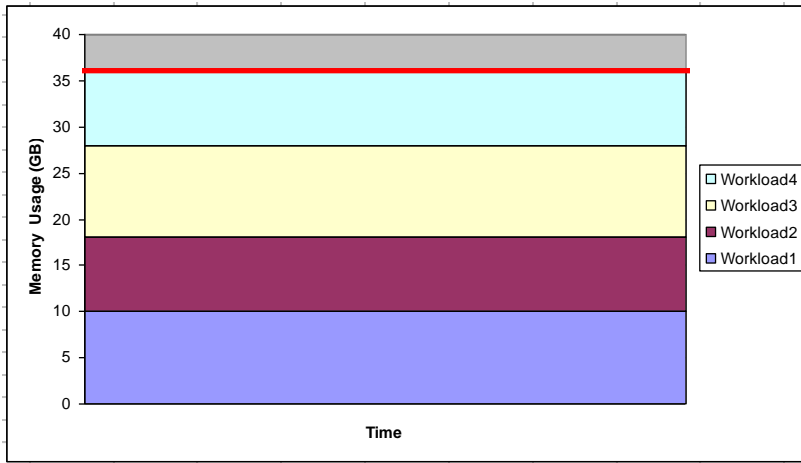


# AMS技术可减少使用成本

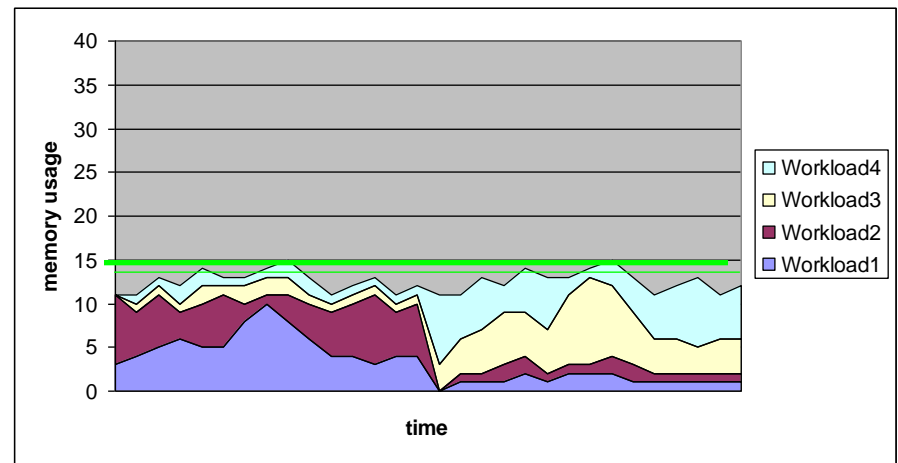
- 根据分区工作负载自动分配内存
- 减少内存的需求总量

Active / Inactive Partitions	Required Physical Memory
Dedicated Memory <span style="color: red;">—</span>	<b>36 GB</b>
Shared Memory <span style="color: green;">—</span>	<b>14 GB</b>

**4 Partitions with Dedicated Memory**

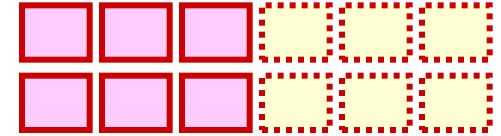


**4 Partitions with Shared Memory**



# 内存虚拟化 — 动态内存扩展(Active Memory Expansion)

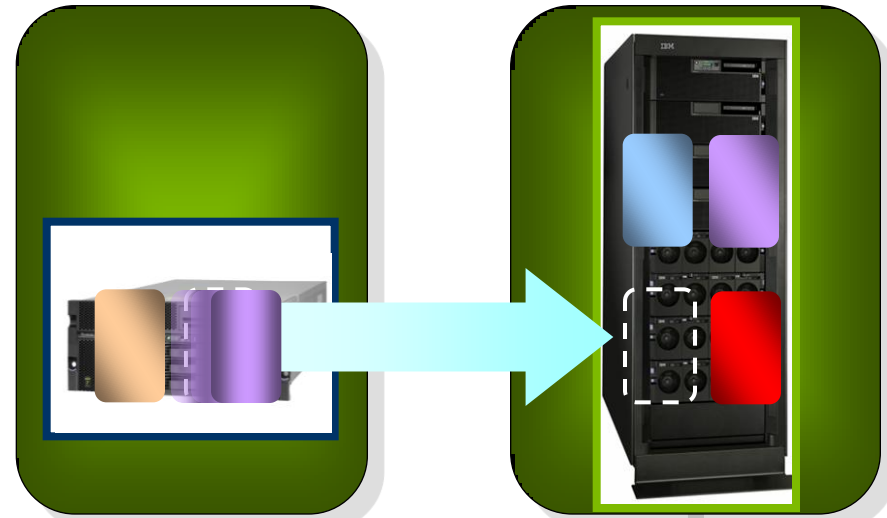
- Innovative POWER7 technology
  - For AIX 6.1 or later
  - For POWER7 servers
- Uses compression/decompression to effectively expand the true physical memory available for client workloads
- Often a small amount of processor resource provides a significant increase in the effective memory maximum
  - Processor resource part of AIX partition's resource and licensing
- Actual expansion results dependent upon how “compressible” the data being used in the application
  - A SAP ERP sample workload shows up to 100% expansion,
  - Your results will vary
  - Estimator tool and free trial available
  - Will not compress AIX kernel
  - Will not compress pinned memory pages (pinned pages to see)
  - Will not compress File memory pages incl. code, cache & mmap (numperm to see)
  - Already compressed data (e.g. compress database objects) will not compress well
  - Computational memory good target (data:heap/stack, but not the code)
  - Localized memory access pattern works well with AME



## 分区在线迁移(Live LPAR mobility)

可以将一个分区从一台服务器上在线迁移到另外一台服务器上，不丢失一笔交易!

- ✓ 降低计划的停机时间
- ✓ 在不同的服务器之间，重新部署负载。
- ✓ 在迁移期间，业务继续运行。

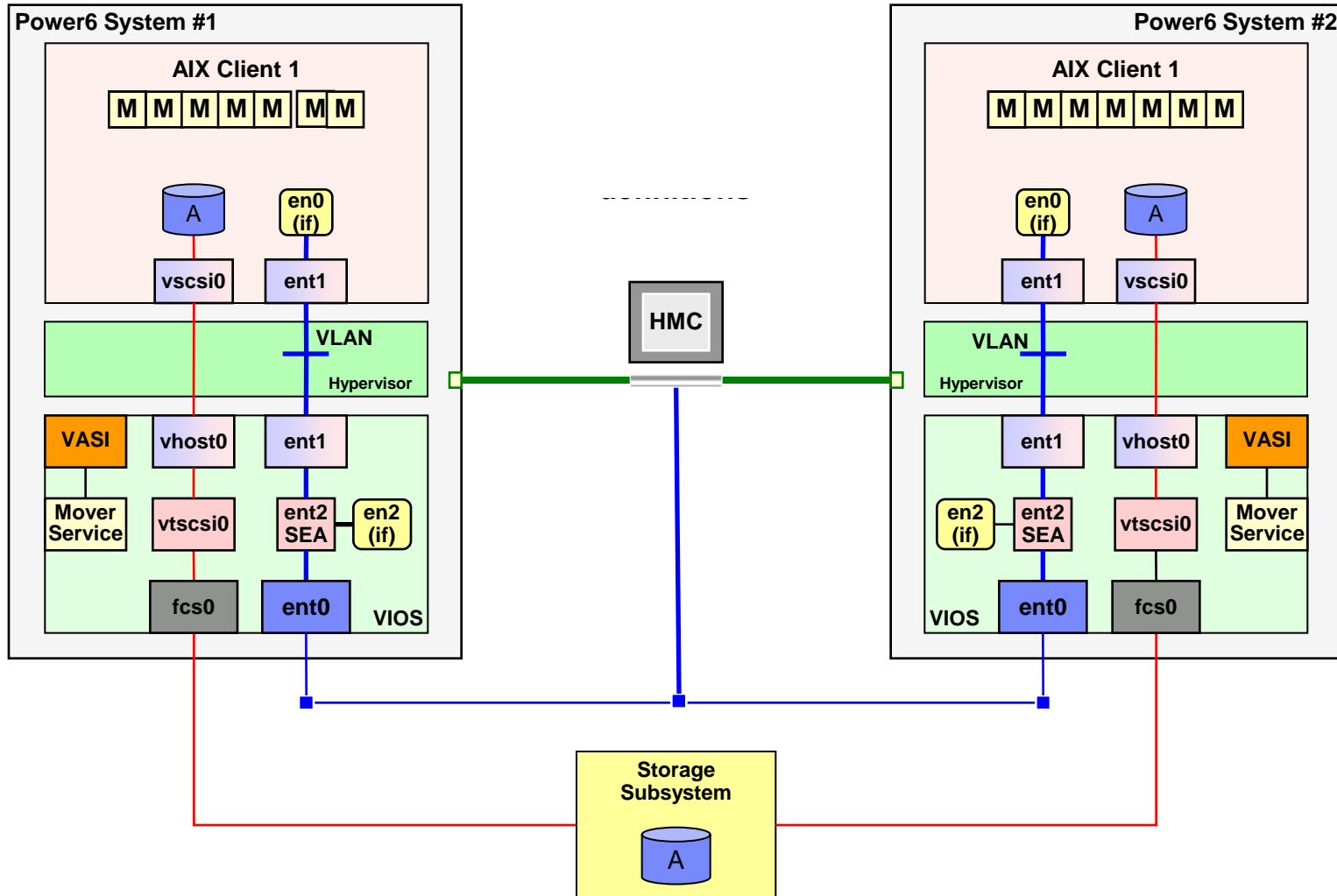


**Virtualized SAN and Network Infrastructure**

在线分区迁移功能需要购买PowerVM企业版许可证,支持POWER6服务器和AIX 5.3, AIX 6.1以及Linux操作系统。

# Live Partition Mobility

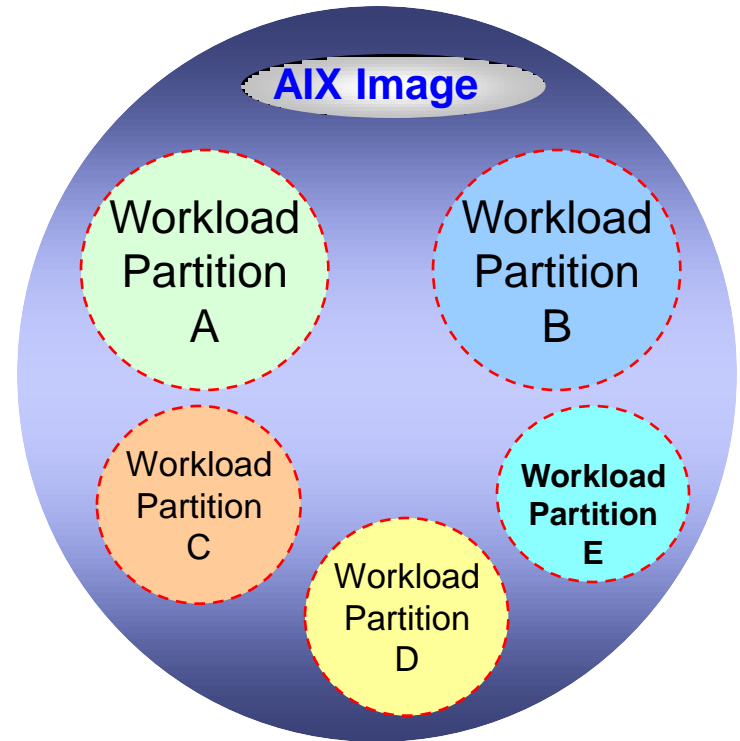
*Partition Mobility supported on POWER6  
AIX 5.3, AIX 6.1 and Linux*



# 工作负载分区(Workload Partitions — WPARs)

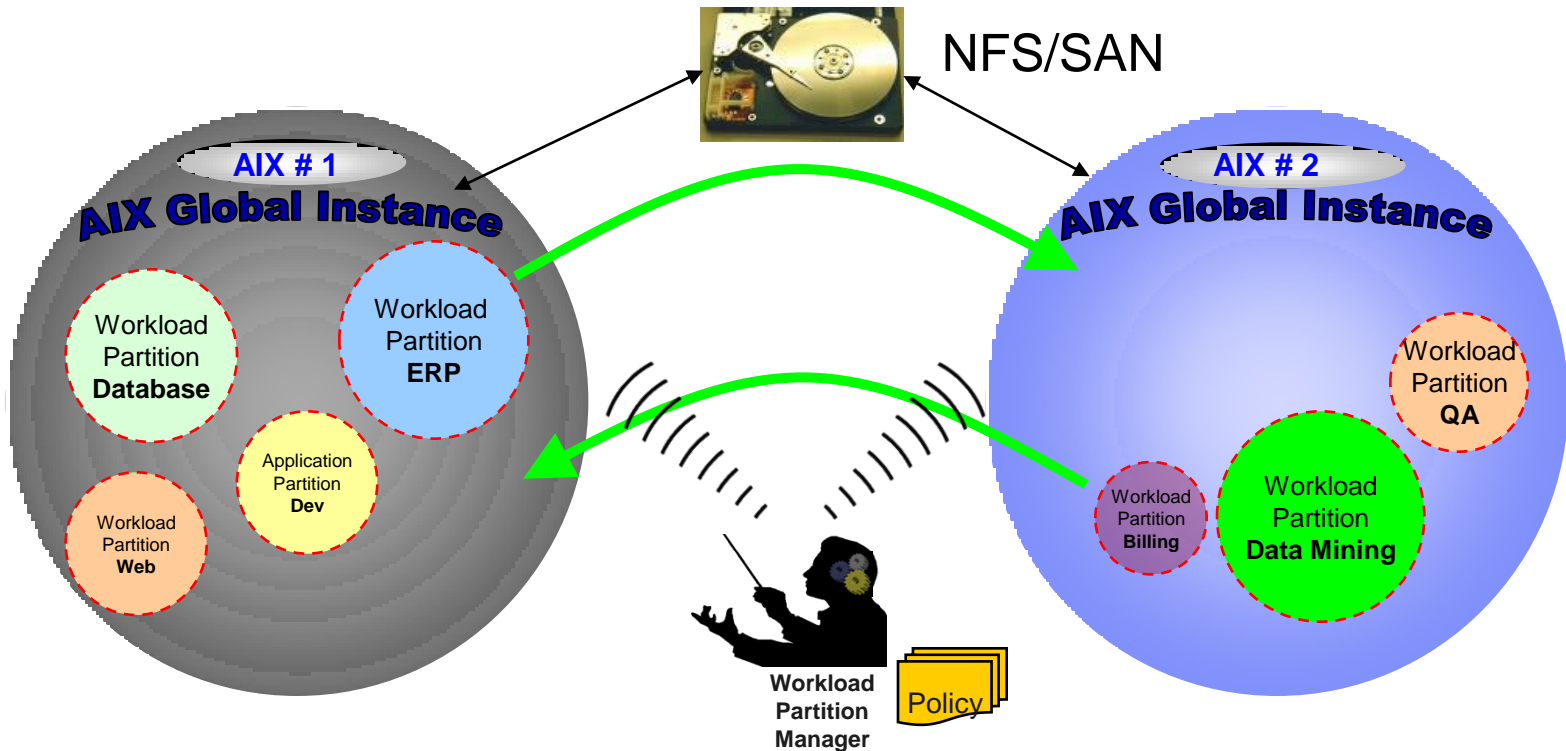
在单一AIX操作系统下，为客户应用提供独立的安全空间。

- **是一种软件虚拟化技术**
  - 运行在AIX V6.1环境下
  - 支持POWER4,POWER5,POWER6 , POWER7 服务器
- **两种类型的应用负载分区**
  - **System WPAR**: 拥有独立、安全、隔离的环境，类似一个完全独立的操作系统
  - **Application WPAR**: 为单一应用程序提供的隔离的运行环境
  - 可以对WPAR的资源进行控制
    - 包括CPU, memory, paging space, threads和processes的数量
- **共享系统资源**
  - 操作系统/共享库和文本等
  - 处理器，内存，I/O设备。





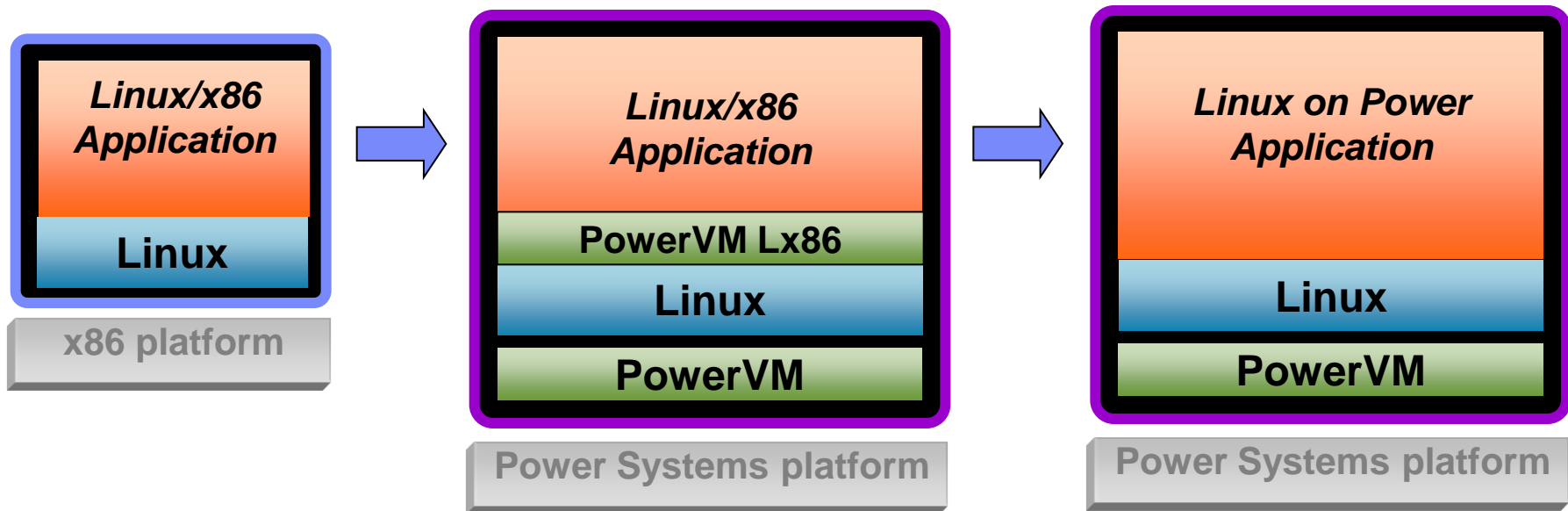
# AIX V6实时应用迁移(Live Application Mobility)



- 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

# PowerVM Lx86 Accelerates Linux Workload Migration

- PowerVM Lx86 cross-platform virtualization runs unmodified Linux/x86 applications within VMs using Linux on Power
  - Copy x86 application binaries and run them – no rewriting necessary
  - Run Linux/x86 workloads with AIX, IBM i and Linux on Power workloads
- Simplifies migration and virtualization of workloads from x86-based platforms to higher-performance Power Systems servers running Linux
- Provides a convenient ‘bridge’ to native compilation of Linux workloads
- Included with all PowerVM Editions



# Integrated Virtualization Manager

*A virtualization solution for small and mid-size companies*

- **Simplifies Management**  
*Browser-based tool for creating and managing partitions*
- **Reduces Costs**  
*Eliminates the need to purchase a dedicated hardware console*
- **Included with all PowerVM Editions**  
*Runs in the Virtual I/O Server partition*

Select	ID	Name	State	Uptime	Memory	Processors	Processing units	Processor utilization	Reference Code
<input type="checkbox"/>	1	10-0044A	Running	9.91 Days	1 GB	2	0.2	7.5%	
<input type="checkbox"/>	2	lp2-aix	Running	8.13 Days	1.98 GB	2	0.2	1.9%	
<input checked="" type="checkbox"/>	3	lp3-linux	Running	8.16 Days	2 GB	1	0.1	16.8%	SuSE Linux ppc64

# PowerVM虚拟化功能汇总

	AIX 5.3	AIX 6.1	IBM i 6.1.1	RHEL 5.4	SLES V10 SP3	SLES V11
DLPAR Processor add/remove	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O adapter add/remove	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR Memory add	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR Memory remove	Yes	Yes	Yes	No	No	Yes
Micro Partitioning	Yes	Yes	Yes	Yes	Yes	Yes
Capacity Upgrade on Demand	Yes	Yes	Yes	Yes	Yes	Yes
Virtual I/O Server	Yes	Yes	Yes	Yes	Yes	Yes
Integrated Virtualization Manager	Yes	Yes	Yes	Yes	Yes	Yes
Virtual SCSI (VIO Server)	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Ethernet (VIO Server)	Yes	Yes	Yes	Yes	Yes	Yes
Workload Partitions (AIX only)	No	Yes	No	No	No	No
Application Mobility (AIX only)	No	Yes	No	No	No	No
Lx86 (Statement of direction for POWER7)	No	No	No	Yes	Yes	Yes
Shared Dedicated Capacity (POWER6)	Yes	Yes	Yes	Yes	Yes	Yes
Multiple Processor Pools (POWER6)	Yes	Yes	Yes	Yes	Yes	Yes
Live Partition Mobility (POWER6)	Yes	Yes	No	Yes	Yes	Yes
N_Port ID Virtualization (POWER6)	Yes	Yes	Yes	Yes	No	Yes
Active Memory Expansion (POWER7)	No	Yes	No	No	No	No
Active Memory Sharing (POWER6)	Yes	Yes	Yes	No	No	Yes
Integrated Virtual Ethernet (POWER6)	Yes	Yes	Yes	Yes	Yes	Yes
Active Memory Expansion (POWER7 / AIX 6.1)	No	Yes	No	No	No	No

**Note:** Some features are supported with specific fix levels or only on specific hardware models. Check the appropriate documentation.

# PowerVM 的三种版本

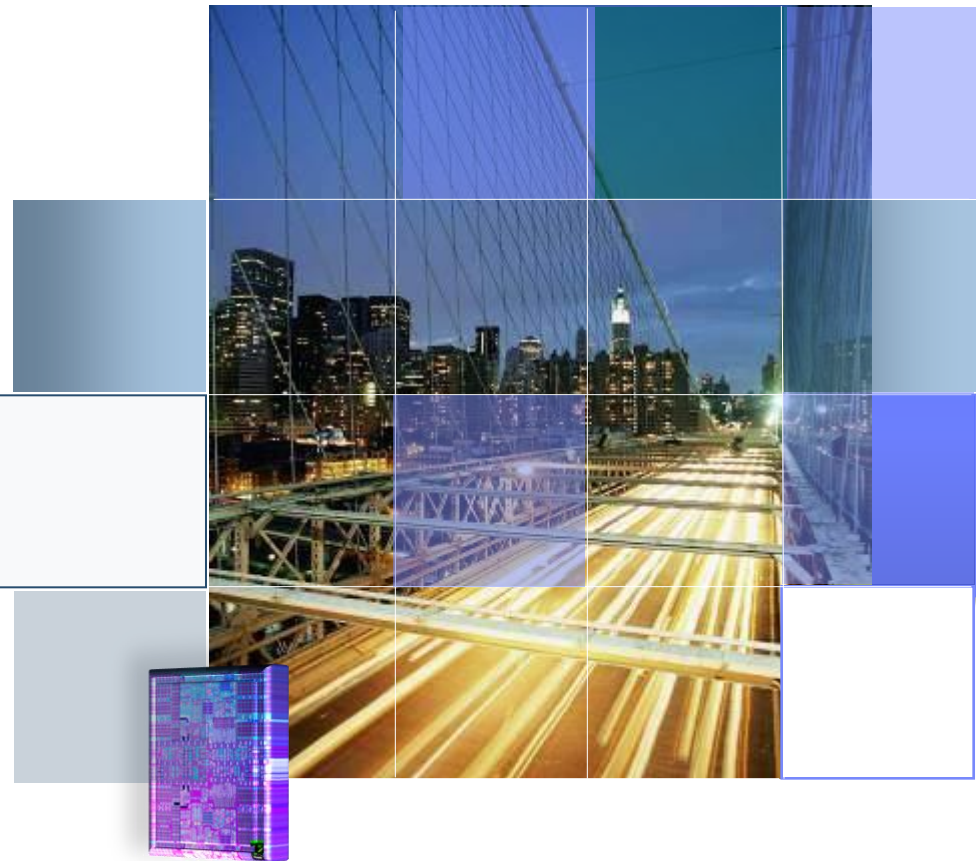
- **PowerVM Express Edition**
  - Evaluations, pilots, PoCs
  - Single-server projects
  
- **PowerVM Standard Edition**
  - Production deployments
  - Server consolidation
  
- **PowerVM Enterprise Edition**
  - Multi-server deployments
  - Cloud infrastructure

PowerVM Edition	Express	Standard	Enterprise
服务器	Power Blade, 520, 550	Power Blade, Power System*	Power Blade Power System*
最大分区数	2 DLPARS +1 VIOS per Server	10 / Core (up to 1000)	10 / Core (up to 1000)
管理平台	VMControl IVM	VMControl IVM, HMC	VMControl IVM, HMC
虚拟IO服务器	✓	✓	✓
分区在线迁移			✓
动态内存共享			✓
多CPU共享池		✓	✓
共享专有分区	✓	✓	✓
客户分区支持	AIX / Linux / i	AIX / Linux / i	AIX / Linux / i
Lx86	✓	✓	✓

**PowerVM Editions** offer a unified virtualization solution for any Power workloads

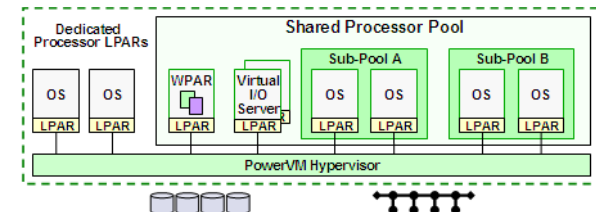


## 实施服务器虚拟化的考虑

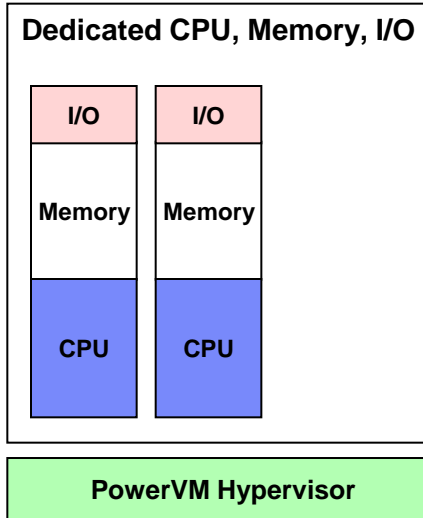


# 实施虚拟化的步骤

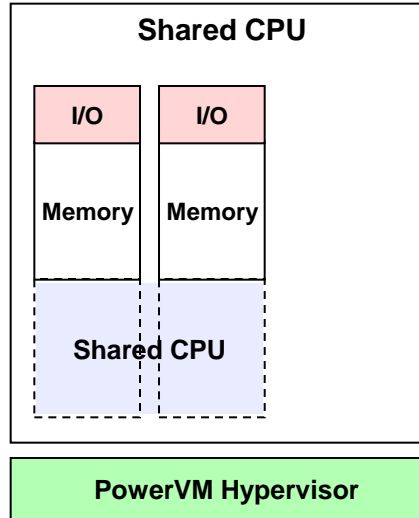
1. 明确需求
  - 组织架构、服务器或应用部署方式等
2. CPU虚拟化
  - Its easy and is appropriate for almost all workloads. Understand multiple shared pools
3. 操作系统引导盘(rootvg)虚拟化
  - Internal disks can be the first limiting factor in number of LPARs
  - Need to understand multi-path requirements
4. 以太网虚拟化
  - Many NICs run at very low utilization
  - Understand VLAN, link aggregation, and connectivity requirements
5. 非rootvg虚拟化 - virtual SCSI, virtual FC or both
6. 内存虚拟化 – requires virtual CPU and I/O
7. 迁移性 – requires virtual CPU and I/O



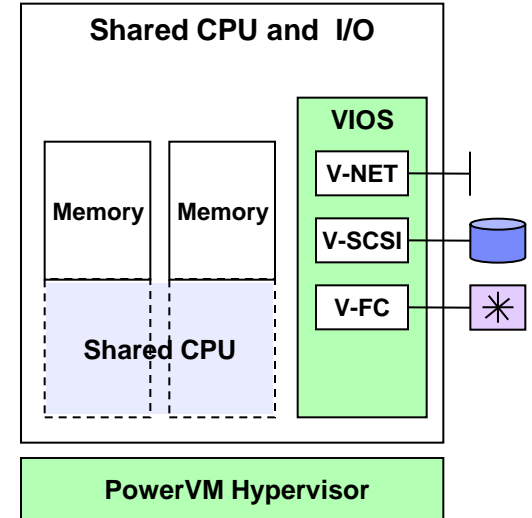
# PowerVM 虚拟化技术的选择



- **Dedicated LPAR Benefits**
  - More flexible & granular than stand alone servers
  - Good for high peak to average CPU utilization
- **Considerations**
  - Considerable unused CPU I/O, and memory capacity
  - Potentially higher software costs



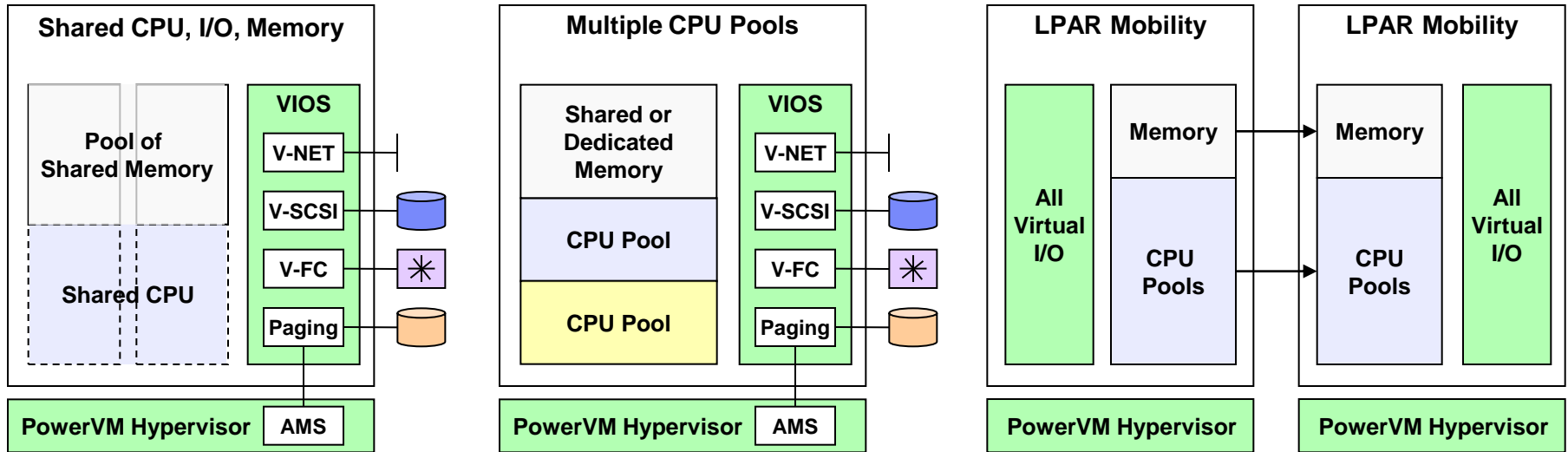
- **Shared CPU Benefits**
  - Much better use of CPU resources
  - Lowers software costs
- **Considerations**
  - Boot drives can limit number of configurable LPARs
  - Significant low use I/O cards
  - Memory dedicated to low use environments



- **Virtual I/O Benefits**
  - No limitation on # of boot devices – more LPARs
  - Sufficient I/O for capacity availability, & connectivity
  - Good use of CPU resource
- **Considerations**
  - Memory dedicated to low use environments may still limit consolidation



## PowerVM虚拟化技术的选择 (续)



### Shared Memory Benefits

- Increase memory utilization especially for low utilization environments
- Can mix shared and dedicated memory LPARs on the same frame
- May allow for increased CPU utilization where previously limited by memory
- Requires virtual I/O & CPU

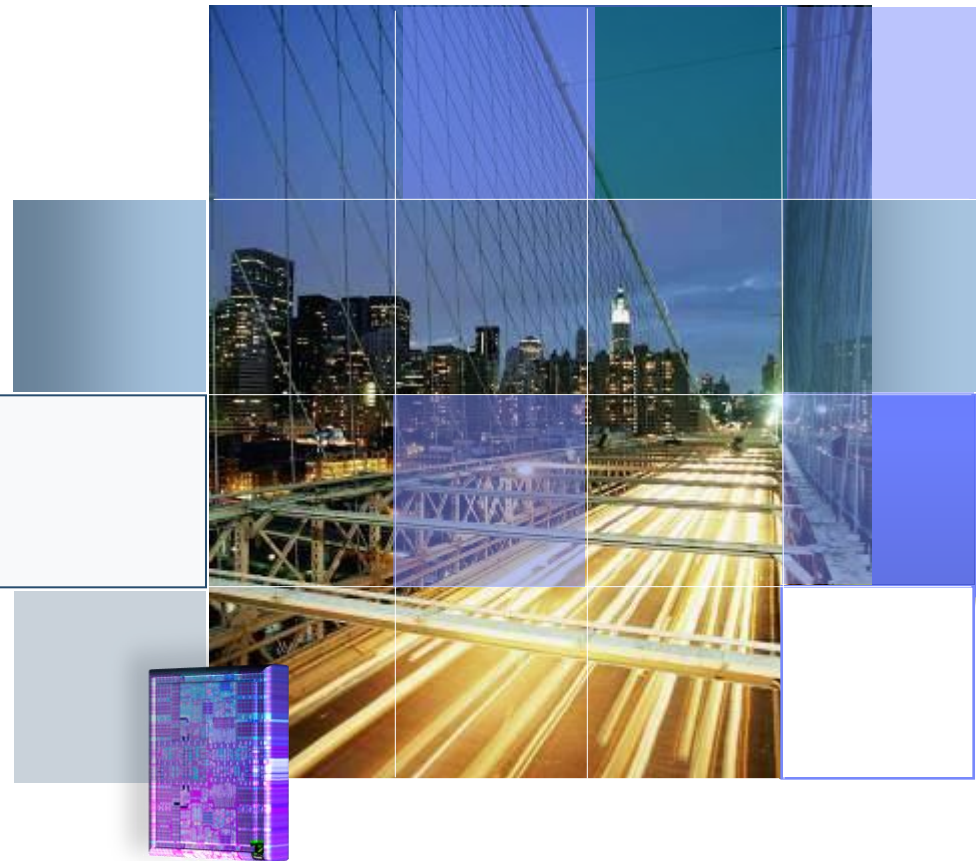
### Multiple CPU Pools

- Allows CPU resources to be grouped by middleware, department, etc
- May lower core based software licensing
- Can guarantee a group of LPARs a given amount of CPU capacity

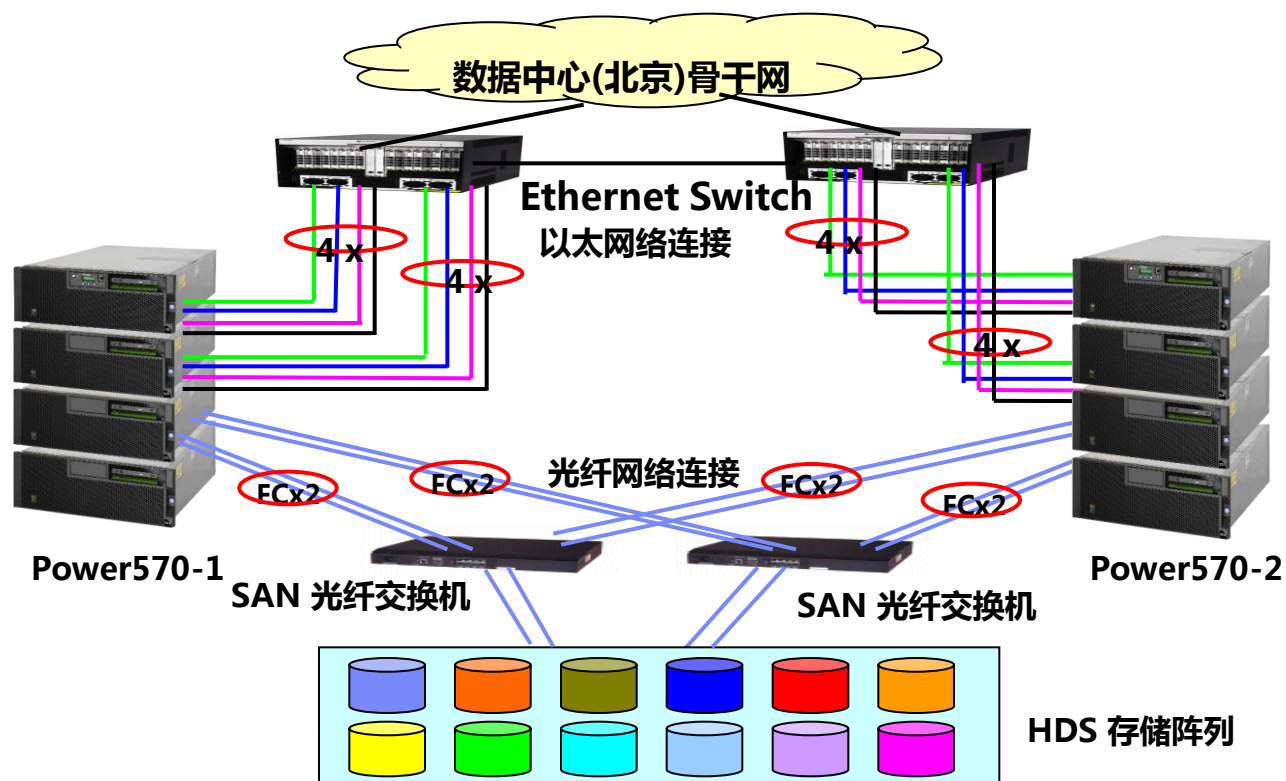
### LPAR Mobility

- Move running LPARs between servers to improve availability
- Move dedicated memory to dedicated memory and shared to shared
- Can move LPARs to specific CPU pools for licensing purposes
- Requires virtual I/O

## 服务器虚拟化的案例



# 某行数据中心(北京) 测试环境PowerVM方案



## 解决方案描述

- 采用PowerVM虚拟化技术，利用有限的硬件资源创建了20个分区满足测试需要；
- 采用PowerVM微分区技术，使用较少的CPU满足大量测试分区需要；
- 采用PowerVM高可用技术，整个系统无单点故障；

## 解决方案优势

- 采用PowerVM虚拟化技术，在节约成本的同时能够快捷、轻松地满足以后测试要求；
- 采用PowerVM微分区技术，能够提高系统整机利用率；
- 整合服务器和虚拟化资源，减少硬件，软件，管理和能源成本；

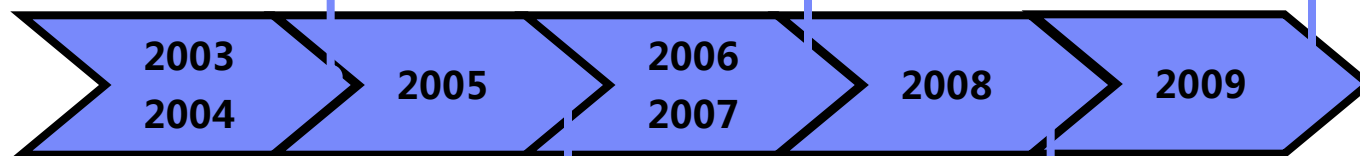
# IBM POWER服务器虚拟化在交行

- 2003年 - 2004年5月，大集中一期生产环境，使用 LPAR 技术

- 2006~2007, 在数据中心生产环境，全面部署DLPAR方案

- 2009年6月后，新上线应用系统启用微分区，实现资源自动调整功能

- 2009年10月，活动分区迁移LPM在数据中心的测试环境和准生产环境投产应用



- 2005年3月，大集中二期 UAT 环境，尝试P5 570 + 微分区 + VIO技术

- 2008年，微分区和VIOS 技术在开发测试环境全面推广

- 2008年5月起，生产系统Web层部署微分区和 VIOS

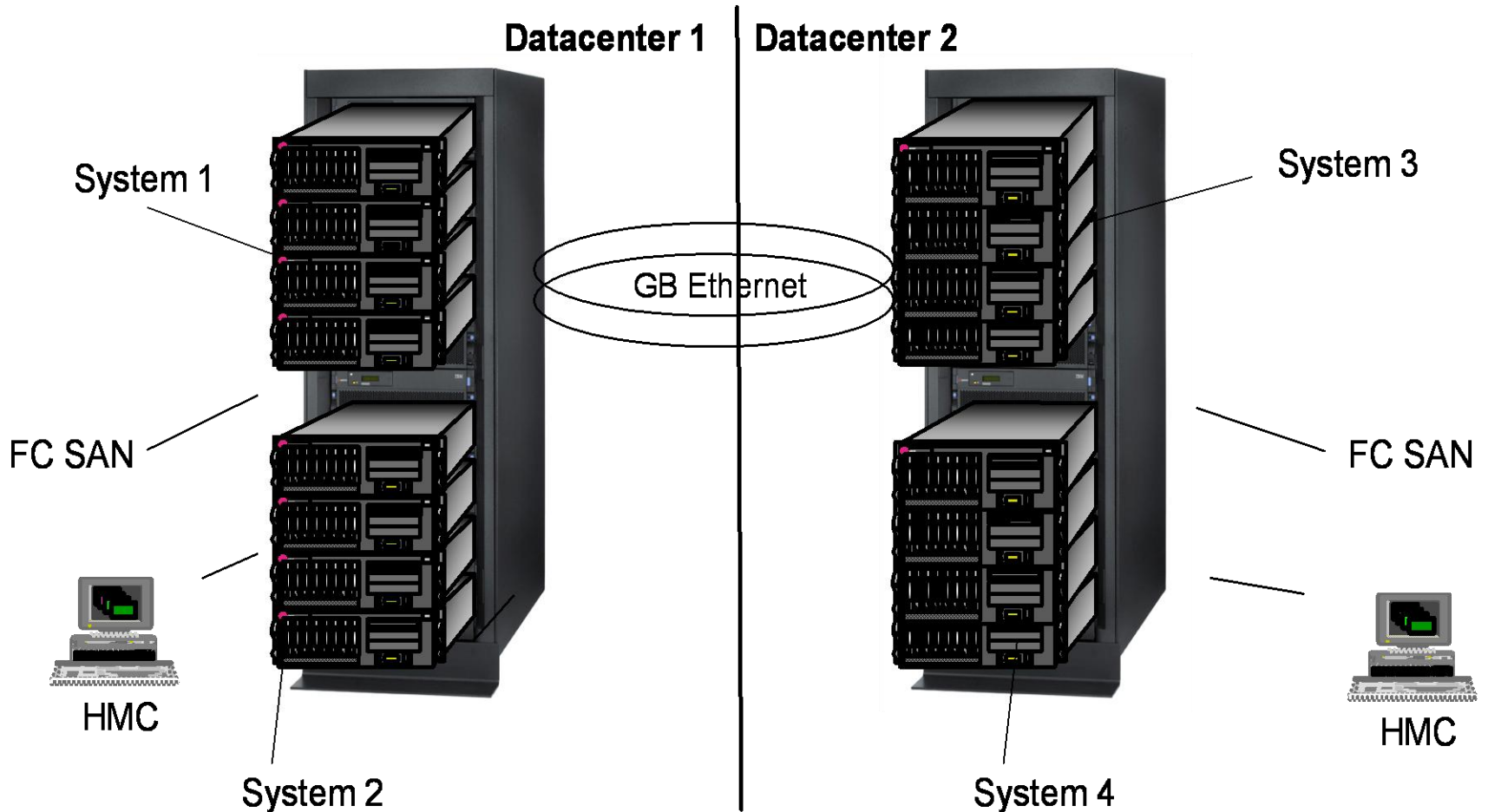
# SAP客户虚拟化案例

## Customer environment exploiting virtualization capabilities on POWER5 systems

- Total number of SAP systems: 77 (PROD + DEV + QAS + “idle HA” partitions)
- **Using SAP Quicksizer**
  - Total number of CPU = **189** (from 0.25 core up to 8 cores)
  - Total amount of memory = **1.5 TB** (assuming 8 GB/p5 core)
  
- Implemented on 4x p570 12-way (16-way o/w 4 CUoD)
  - Total number of active CPU = **48** (= 4 x 12)
  - Total amount of memory = **672 GB** (Total = 1024 GB o/w 352 for CUoD)
  
- **SAP productive partitions can expand up to 12 CPUs**
- All LPARs are **shared and uncapped** configured
- Main memory
  - **Less in total, but more per CPU**

# SAP客户虚拟化案例 (续)

Customer environment exploiting virtualization capabilities on Power5 systems  
(4x p570, each with 12 active CPU + 4 CPU CUoD)



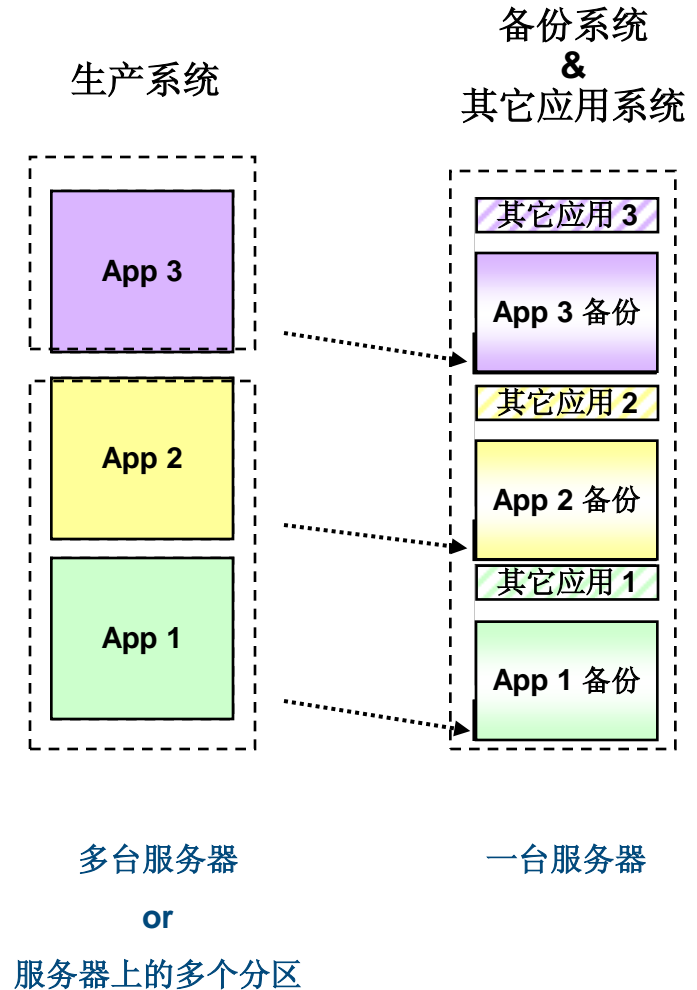
## Example: Distributing 21 SAP instances on 23 LPARs within one single p570 system (12 active CPUs, 4 CoD)

SAP-System	LPAR	Memory (in G)			CPU-Entitlement			vCPU			Weighting
		Minimum	Desired	Maximum	Desired	Minimum	Maximum	desired	min	max	
VIO-Server1	2	1	2	4	0,4	0,2	2	2	1	4	128
VIO-Server2	3	1	2	4	0,4	0,2	2	2	1	4	128
P33	4	4	6	12	0,4	0,2	1,2	2	1	2	64
P23	5	4	9	18	0,5	0,3	2	3	1	8	64
CS-Test	6	1	2	4	0,1	0,1	0,4	1	1	4	32
C03	7	2	6	12	0,3	0,2	0,8	2	1	2	32
C17	8	1	2	4	0,2	0,2	0,5	2	1	2	32
C21	9	2	3	6	0,2	0,2	0,5	2	1	2	32
C07	10	2	6	12	0,4	0,2	0,9	2	1	4	32
P07_idle	11	1/4	16	16	0,4	0,1	4	4	1	8	64
P31_idle	12	4	6	12	0,4	0,2	0,8	2	1	8	64
I03	13	1	2	4	0,1	0,1	0,3	1	1	3	32
C04	14	3	8	16	0,1	0,1	0,5	1	1	5	32
T98	15	2	8	16	0,1	0,1	0,5	1	1	5	16
C91	16	2	6	12	0,1	0,1	0,5	1	1	5	32
IDMS-X	17	1	4	8	0,1	0,1	0,5	1	1	5	32
A41	18	1	4	8	0,1	0,1	0,5	1	1	5	32
K07	19	1	4	8	0,1	0,1	0,5	1	1	5	16
IDMSXP2	20	1	4	8	0,2	0,1	0,5	1	1	5	64
sideprod	21	4	5	6	0,1	0,1	0,5	1	1	5	64
Q04	22	4	8	16	0,1	0,1	0,5	1	1	5	32
P04_idle	23	1/4	8	16	0,2	0,1	0,5	2	1	5	64
Q72	24	4	20	20	1,6	0,4	2	8	2	8	32
<b>ME RZ2</b>			<b>141</b>	<b>242</b>	<b>4,7</b>	<b>3,0</b>	<b>19,9</b>	<b>44</b>			

Production
Qualityassurance
Development  
K-Backup (idle LPAR)  
Virtual-I/O Server

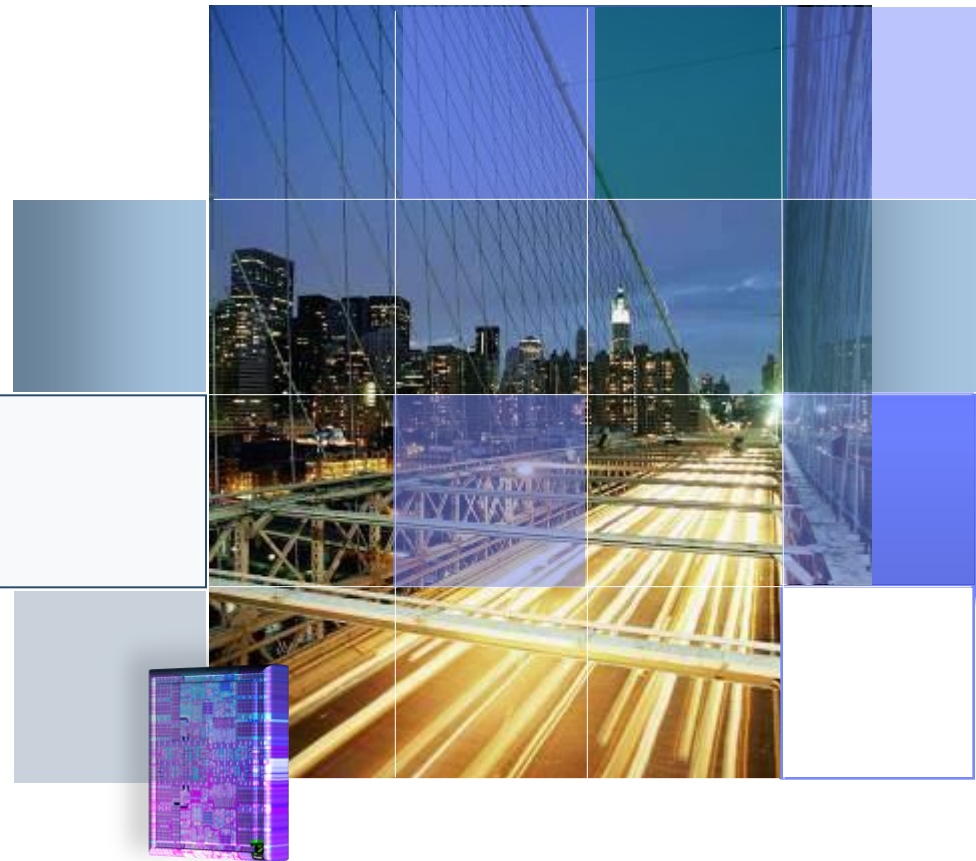
# PowerHA 与 PowerVM的结合

- 备份机实现HA或DR功能
- 处理能力共享
- 提高服务器整体使用率



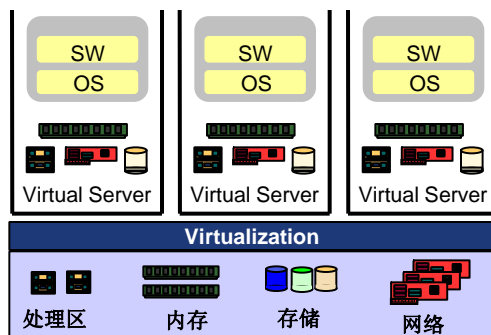
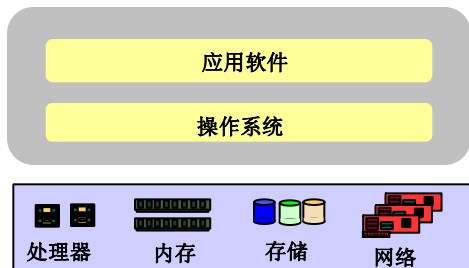


# 虚拟化的管理

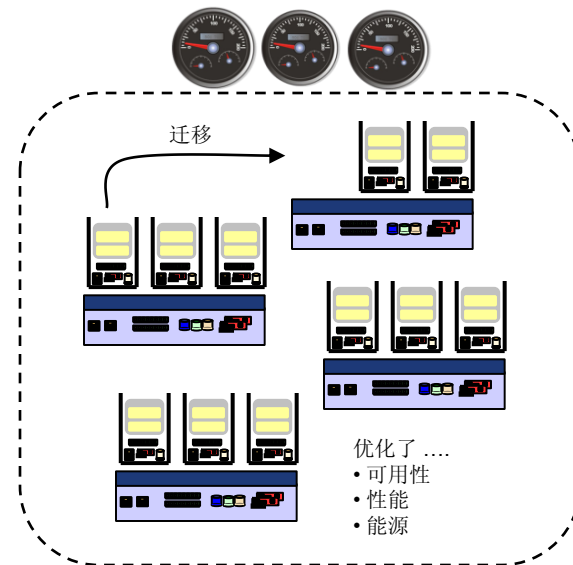


# 平台的演化

## Systems Director VMControl 像管理单一系统一样来管理系统池



更多的虚拟，更少的管理



服务器  
Servers

虚拟服务器  
Virtual Servers

系统池  
System Pools

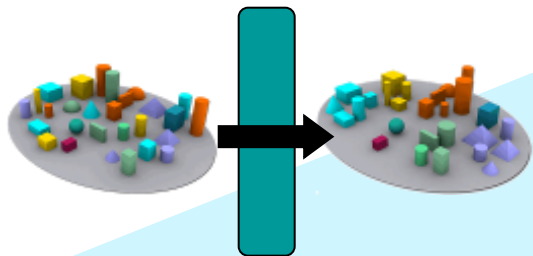


# 对虚拟资源进行最有效的控制和管理



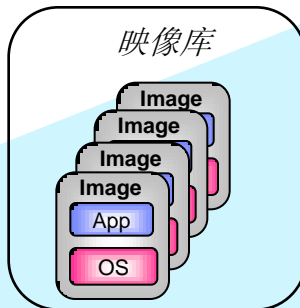
## Systems Director VMControl

易捷版  
-虚拟机



- 基于业务需求选择最佳实践匹配
- 创建像构建块一样的标准化的虚拟机
- 选择默认的属性

标准版  
- 管理虚拟映像



- 捕获并收编数据中心中使用的虚拟映像
- 标准化虚拟映像构建块
- 定制虚拟环境运行时需求
- 通过虚拟映像简化部署

企业版  
- 优化系统池



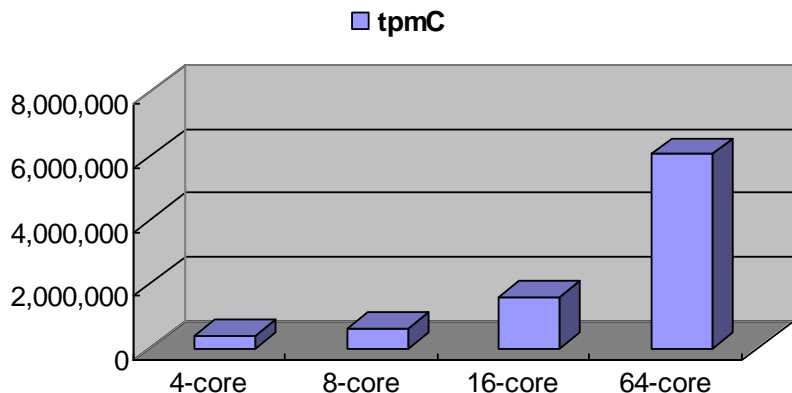
- 像管理单一系统一样管理系统池
- 把多系统一体化管理
- 自动安排新的工作负载
- 统一为固件更新管理配置
- 跨硬件代的持久的、即插即用的能力

云计算-就绪



- 工作负载基于服务水平目标集中管理
- 确保达到服务水平承诺
- 完整的IT过程的虚拟化管理
- 持续可用
- 弹性扩充
- 自动供给

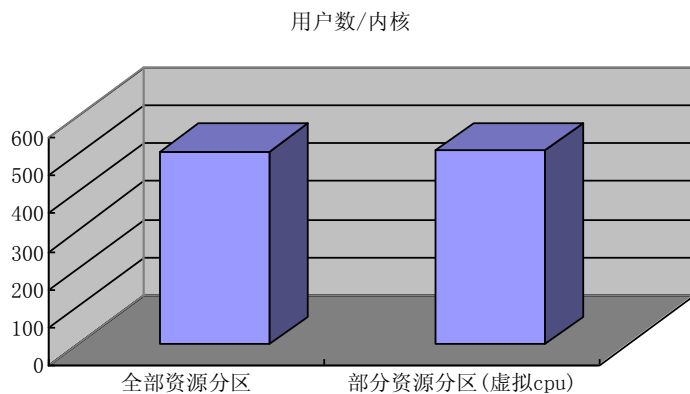
# Power虚拟化不损失性能 — 所有发布指标均在虚拟环境下测试



IBM TPC-C 测试结果

IBM 是目前公布带 PowerVM 等虚拟技术 TPC-C 测试结果的唯一供应商，测试结果表明Power具有更强的线性扩展能力。

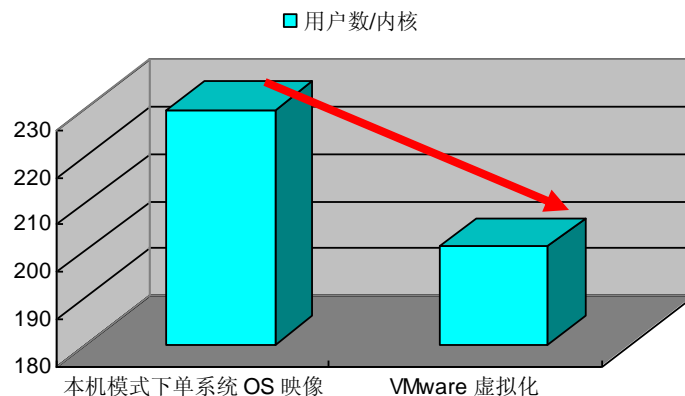
信息来源: <http://www.tpc.org> 表中所示为10/16/09测试结果



IBM SAP SD 2-Tier “全部资源分区”与“部分资源分区 (虚拟cpu)”对比

IBM “全部资源分区”与采用“部分资源分区(虚拟 CPU)”的 SAP SD 2-tier测试结果对比，两种测试结果每个内核用户数大致相同，说明IBM Power虚拟化没有损失任何性能

信息来源: <http://www.sap.com/benchmark>



SAP SD 2-tier“本机模式下单系统 OS 映像”与“VMware 虚拟化”对比

“本机模式下单系统OS映像”与采用“VMware 虚拟化”的测试结果对比，两种结果之间存在很大差别。“虚拟化”测试结果每个内核用户数更少，说明Vmware虚拟化对性能带来了非常大的损失

信息来源: <http://www.sap.com/benchmark>

# 智慧的地球需要智慧的系统

