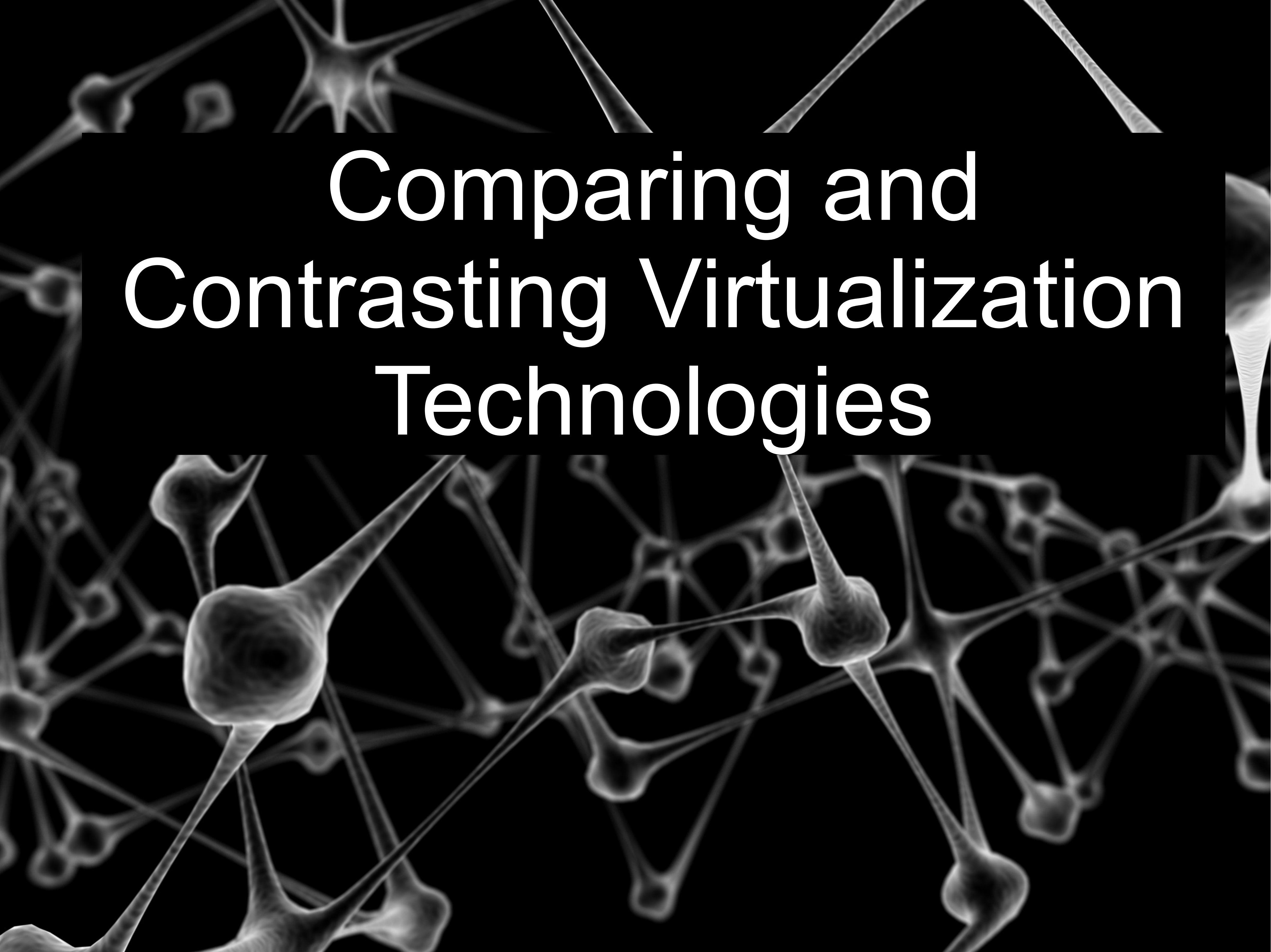
Comparing and Contrasting Virtualization Technologies



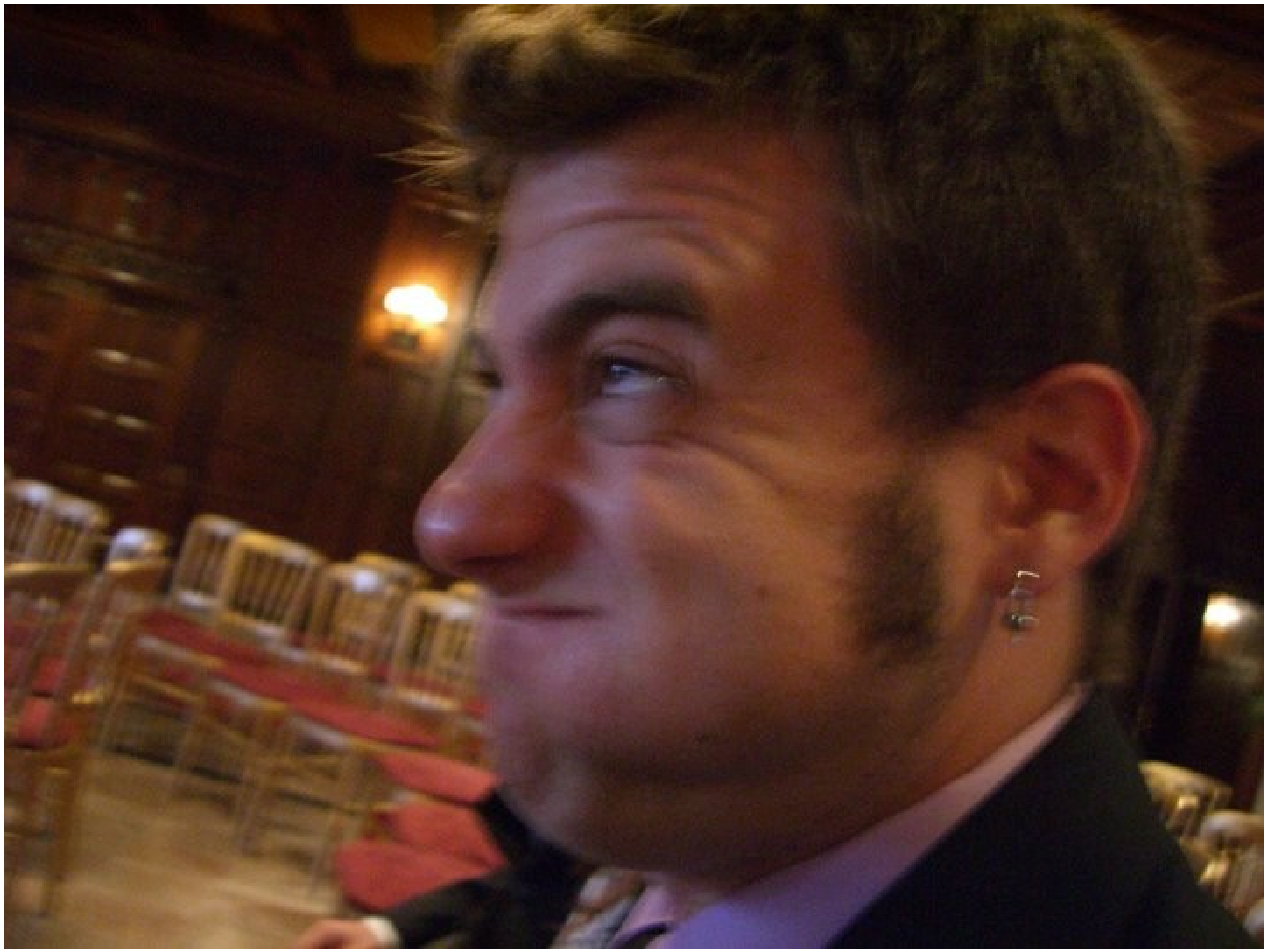
Abstract

Do you know the differences between Xen and VMware?

Do you know when it is more advantageous to use one over the other?

Virtualization can be a complicated subject with many different facets, and it is not always easy to choose the strategy that best fits your needs. This session will explore the various virtualization options that apply to System x, System p and System z. You will learn about the commonalities between each offering and also how they differ.

buzzetti@us.ibm.com



open source and Linux

Clarkson
UNIVERSITY
®

python™

The Design Center

IBM Design Centers for IT Optimization & Business Flexibility

Our mission – "To design and architect innovative end-to-end solutions with select worldwide clients that leverage appropriate leading-edge technologies and accelerate their deployment."

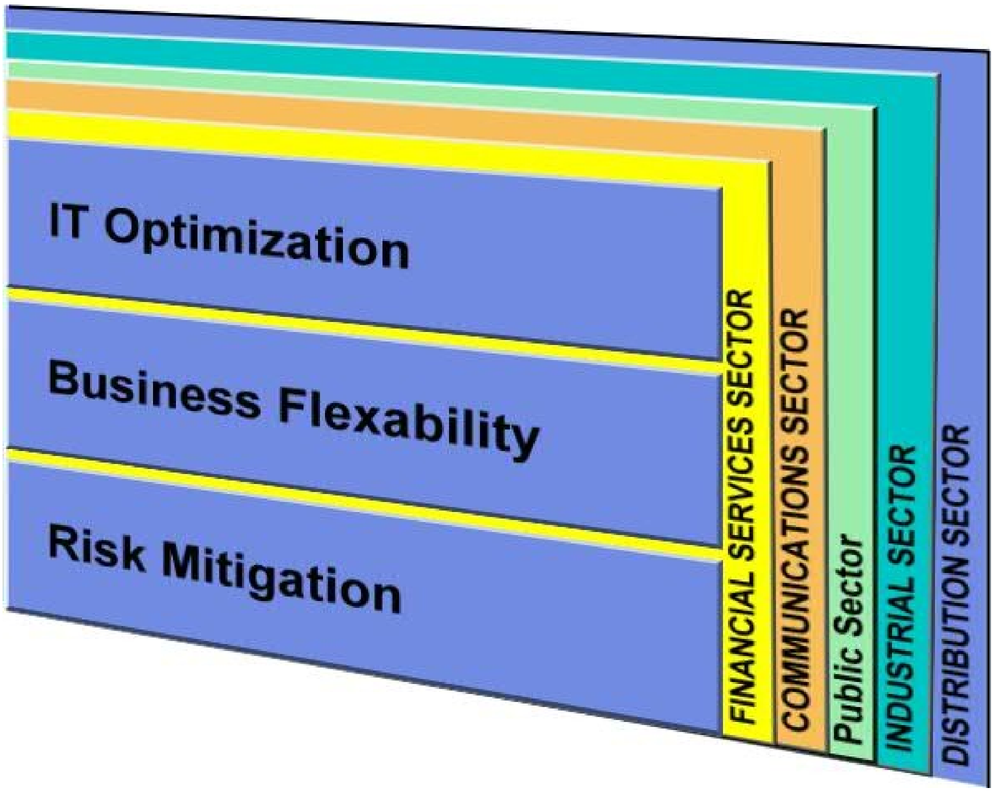
Engagement Offerings:

3 __flexible__ engagement models

that can be tailored by industry

to suit the individual client's needs

- ✓ *Whiteboarding Session*
- ✓ *Architectural Review*
- ✓ *Design Workshop*



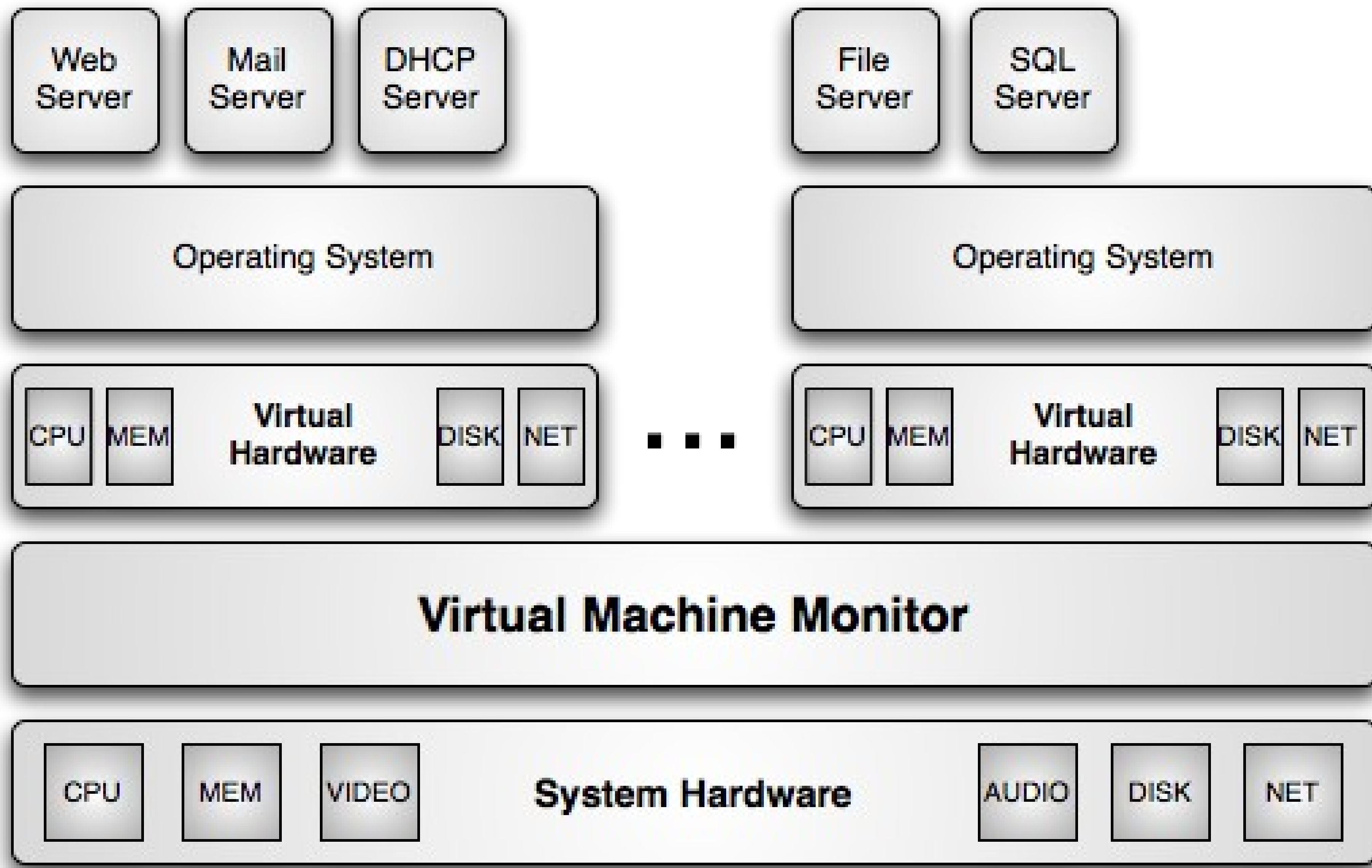
There is no free lunch

Virtualization

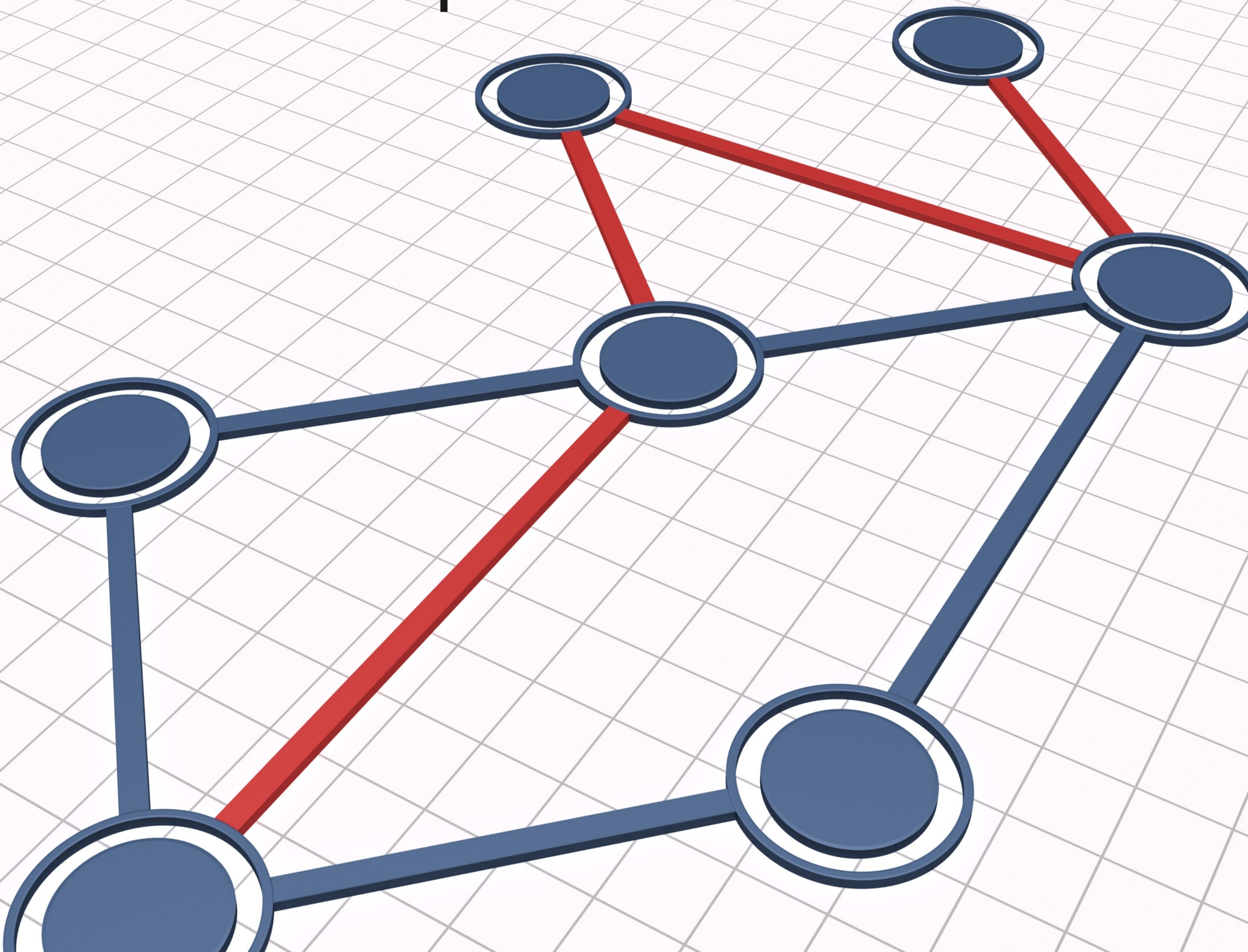
Popek and Goldberg

In 1974, Gerald Popek and Robert Goldberg released an article titled ***"Formal Requirements for Virtualizable Third Generation Architectures"***

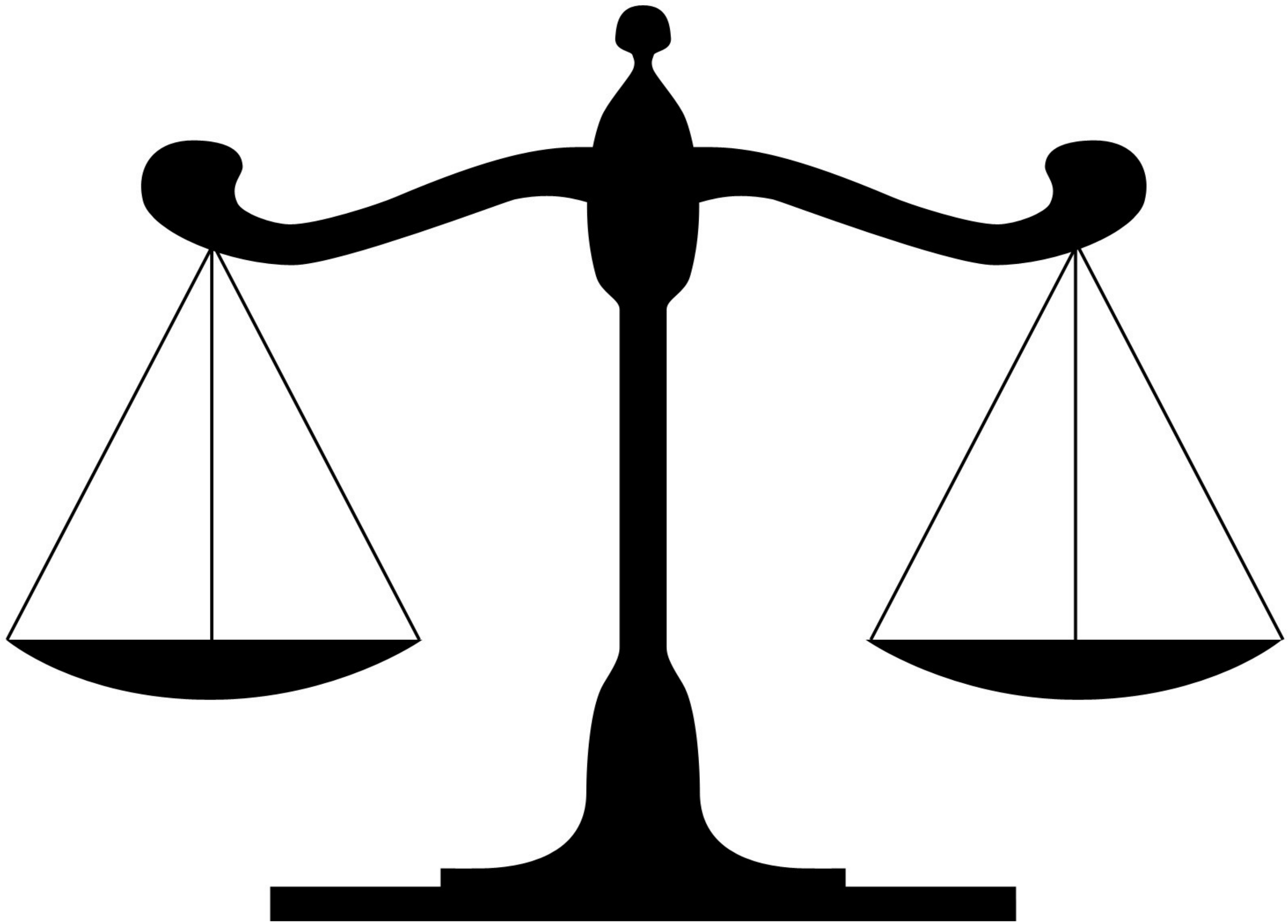
Virtual Machine Monitors



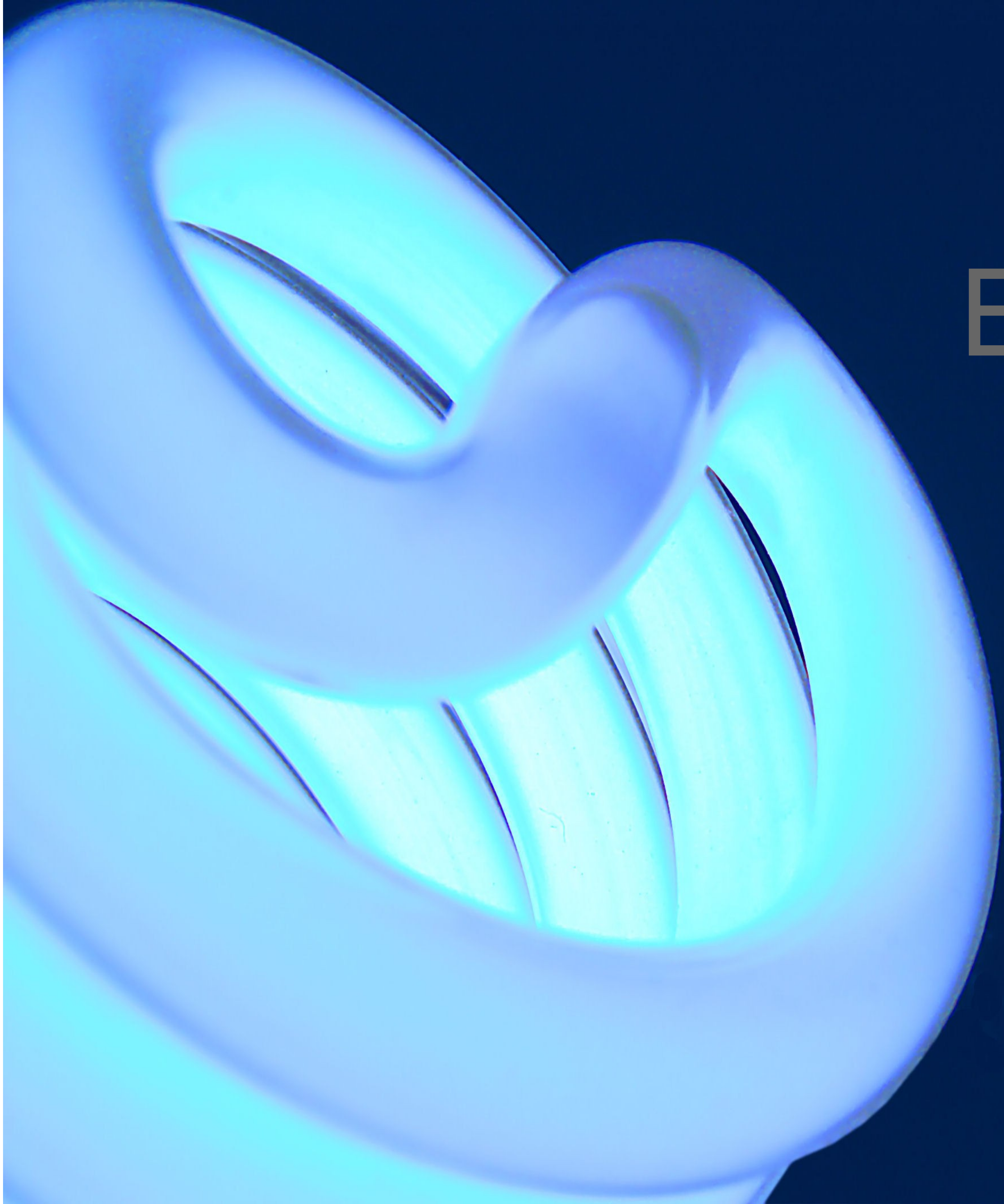
VMM Properties



Equivalence



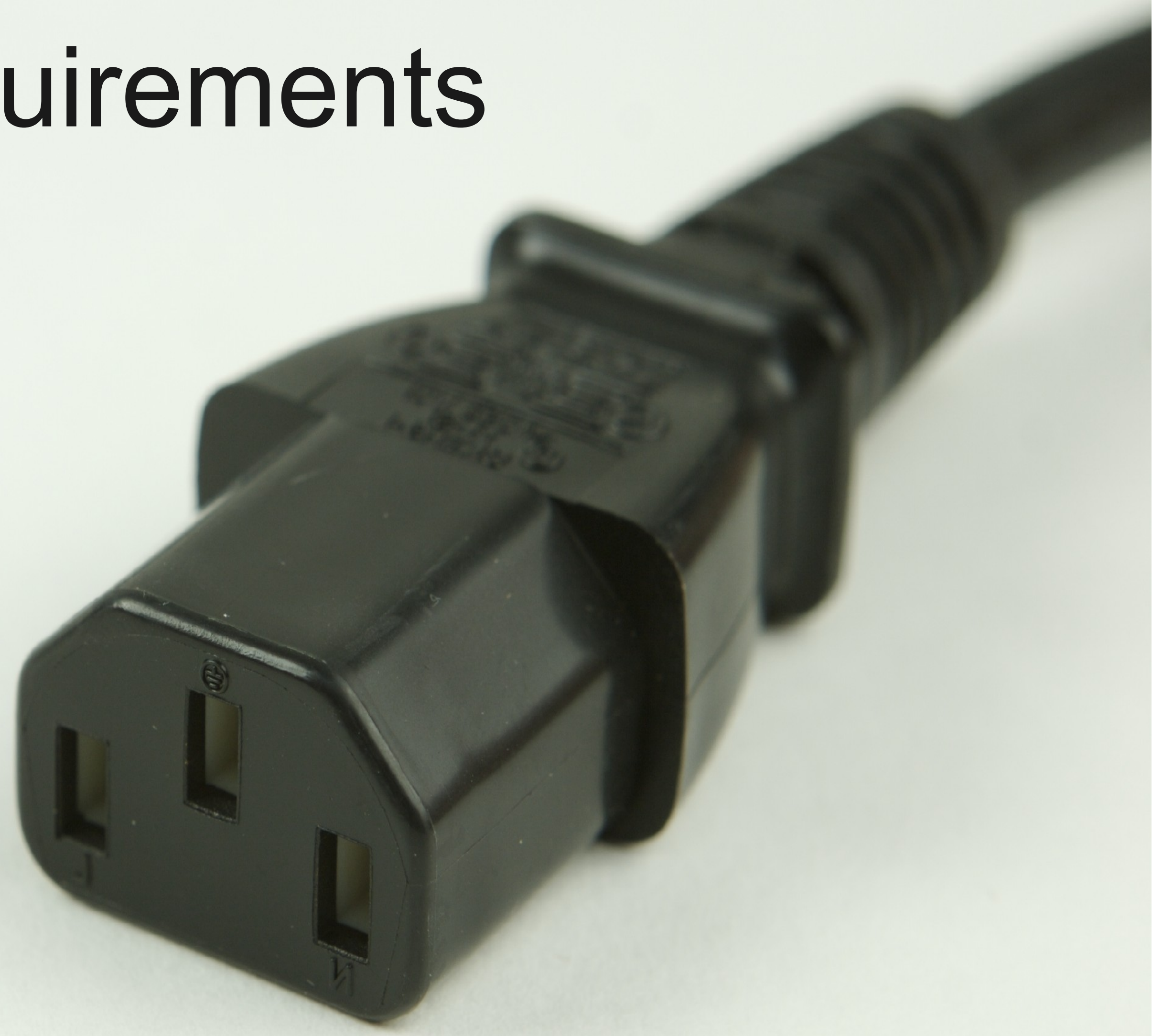

START
Resource
Control

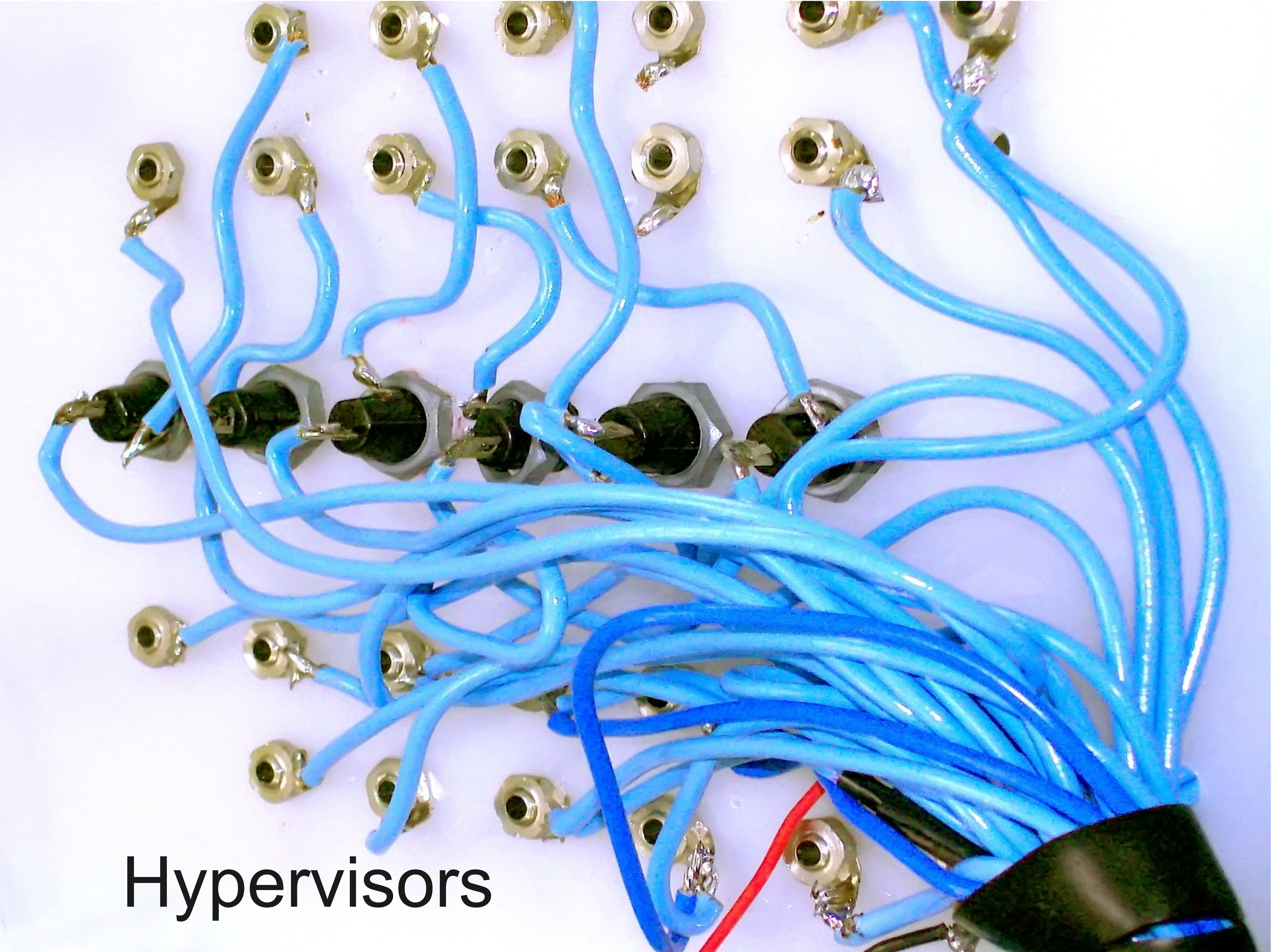


Efficiency

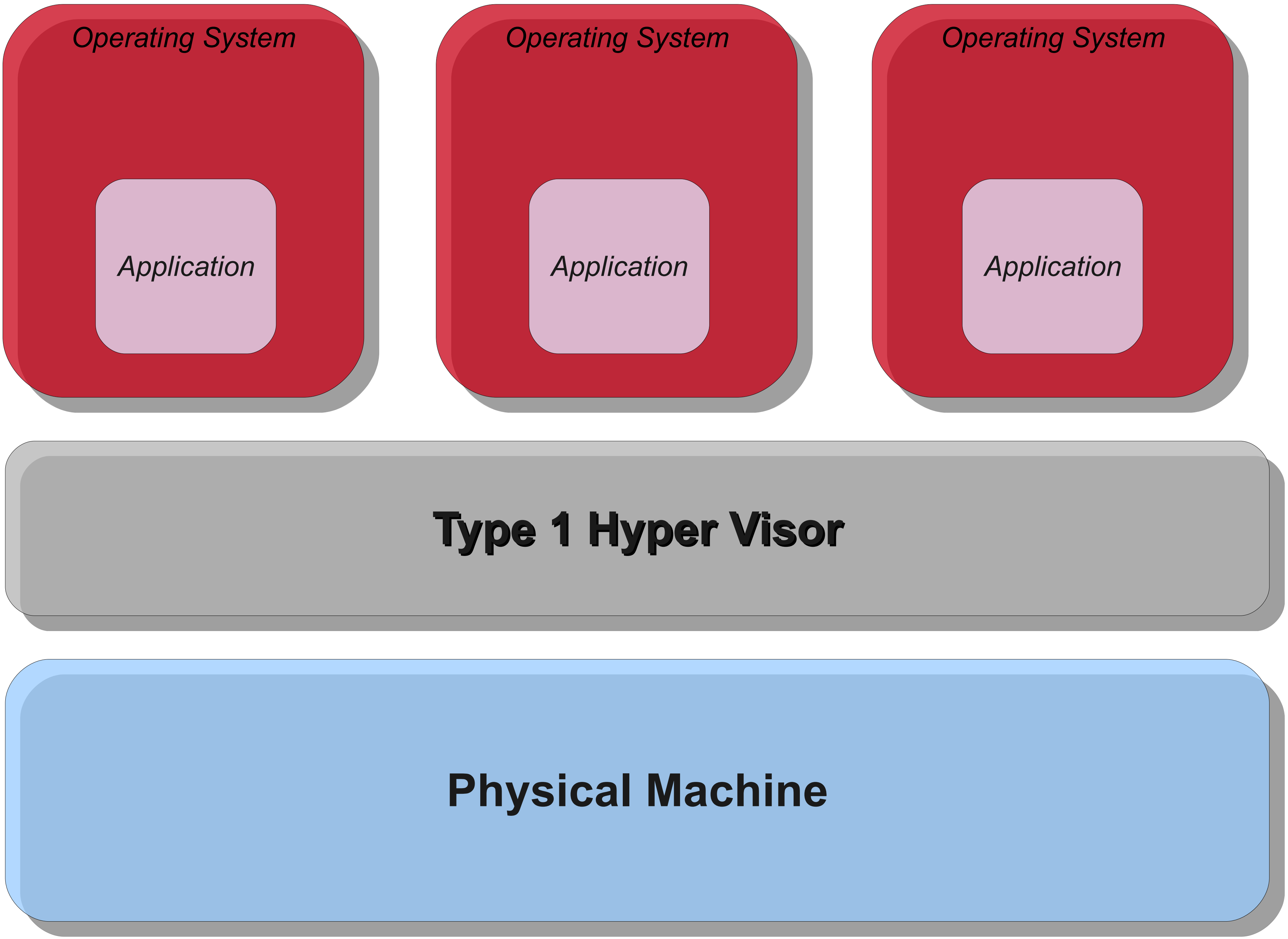
Requirements

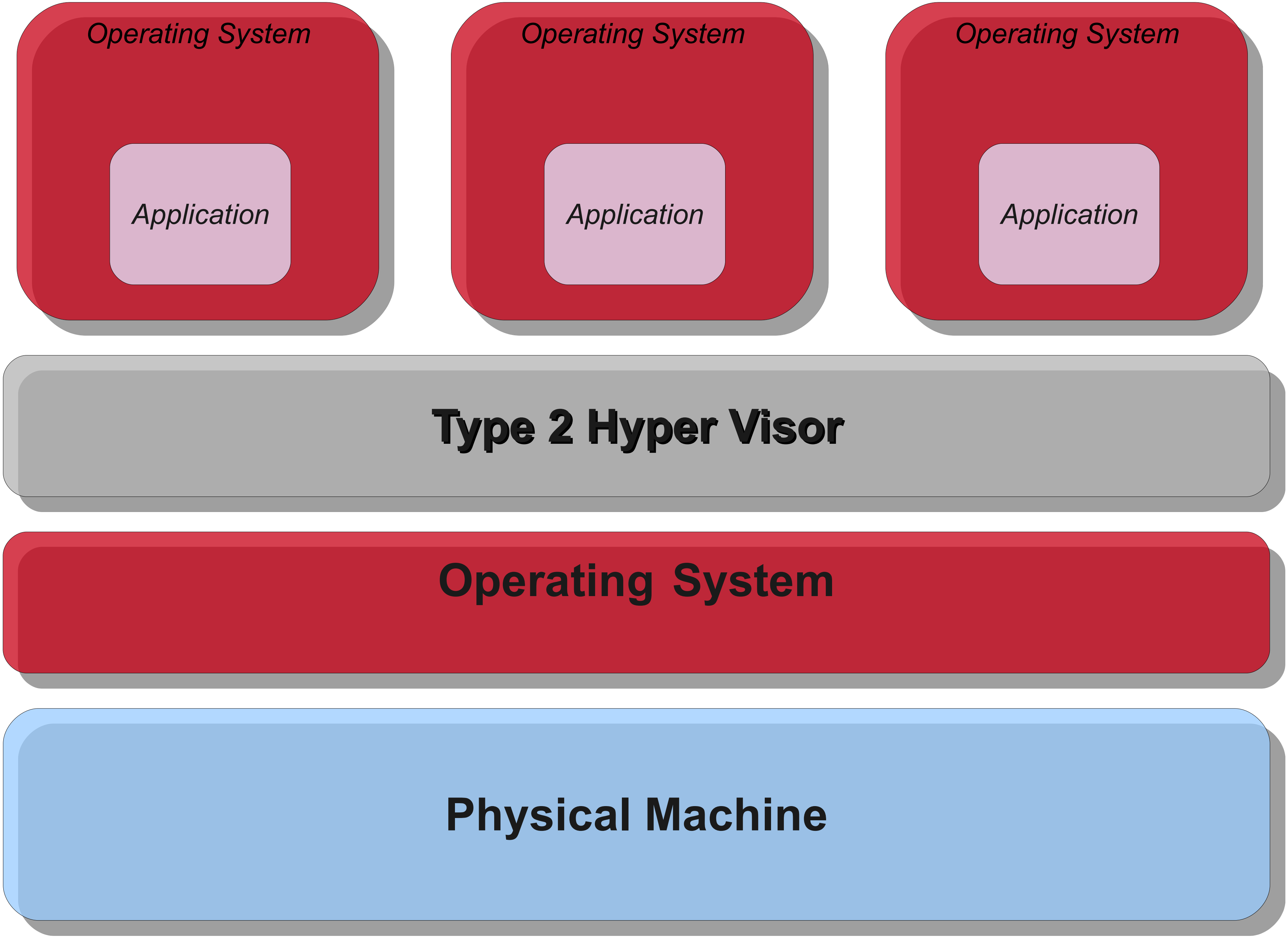
- Privileged instructions
- Control sensitive instructions
- Behavior sensitive instructions



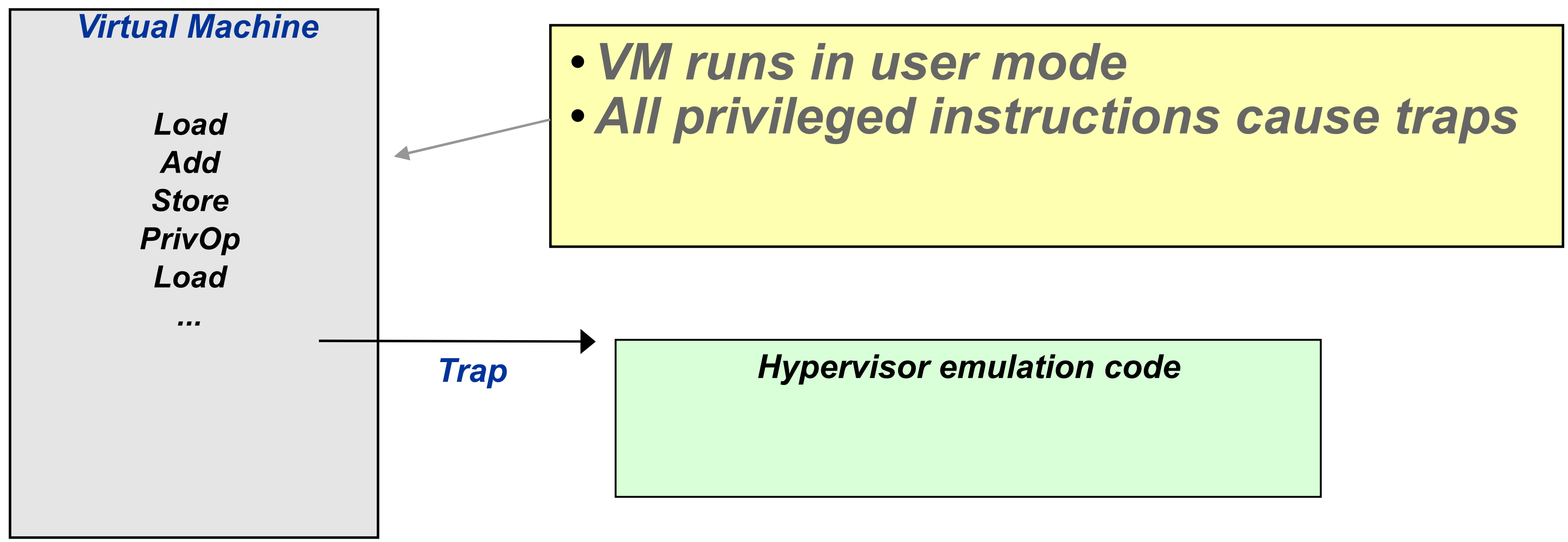


Hypervisors


Operating System
Application
Operating System
Application
Operating System
Application
Type 1 Hyper Visor
Physical Machine


Operating System
Application
Operating System
Application
Operating System
Application
Type 2 Hyper Visor
Operating System
Physical Machine

Trap and Emulate

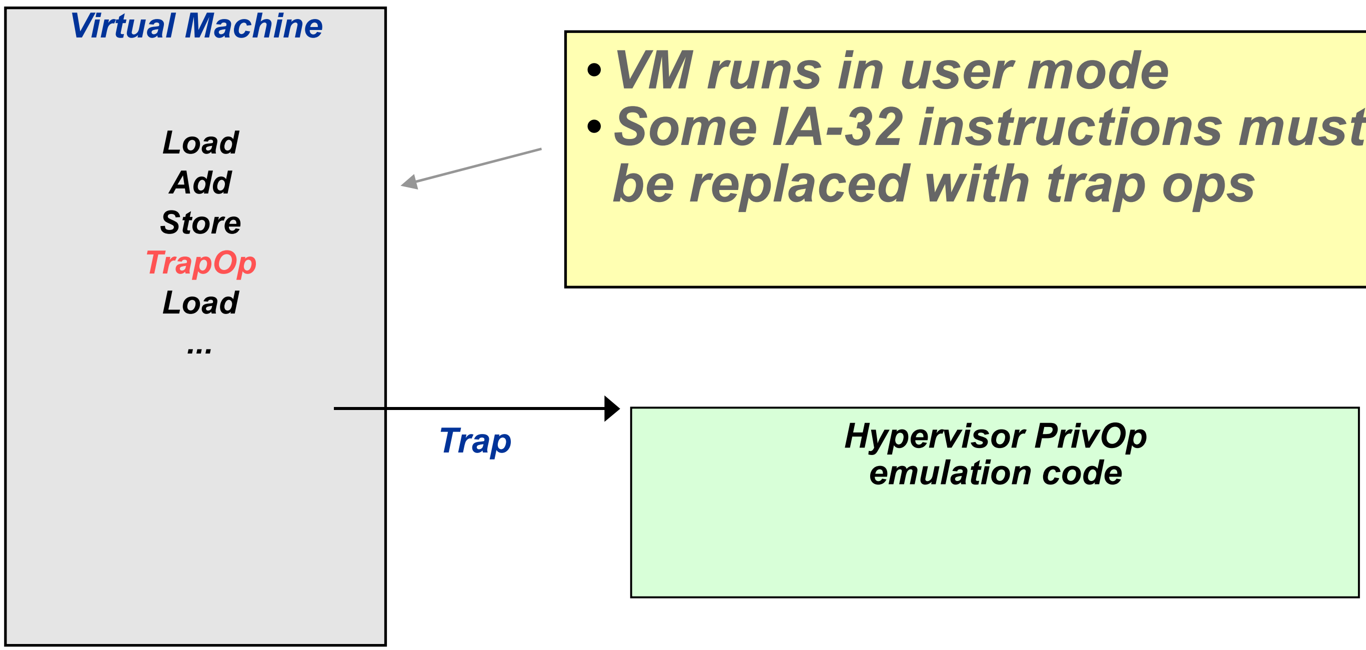


Examples: CP-67, VM/370

Benefits: Runs unmodified OS

Issues: Substantial overhead

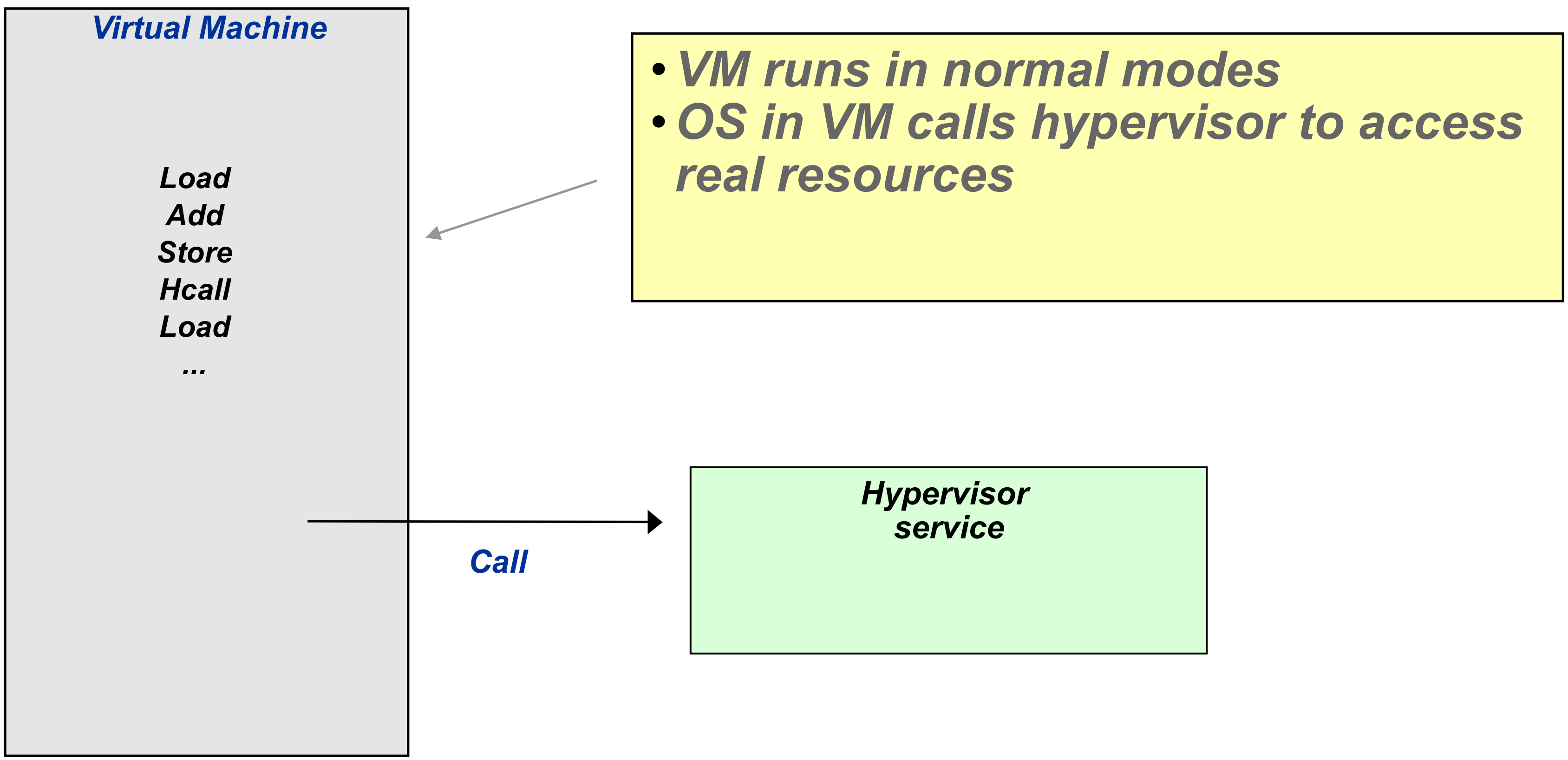
Translate, Trap, and Emulate



- *VM runs in user mode*
- *Some IA-32 instructions must be replaced with trap ops*

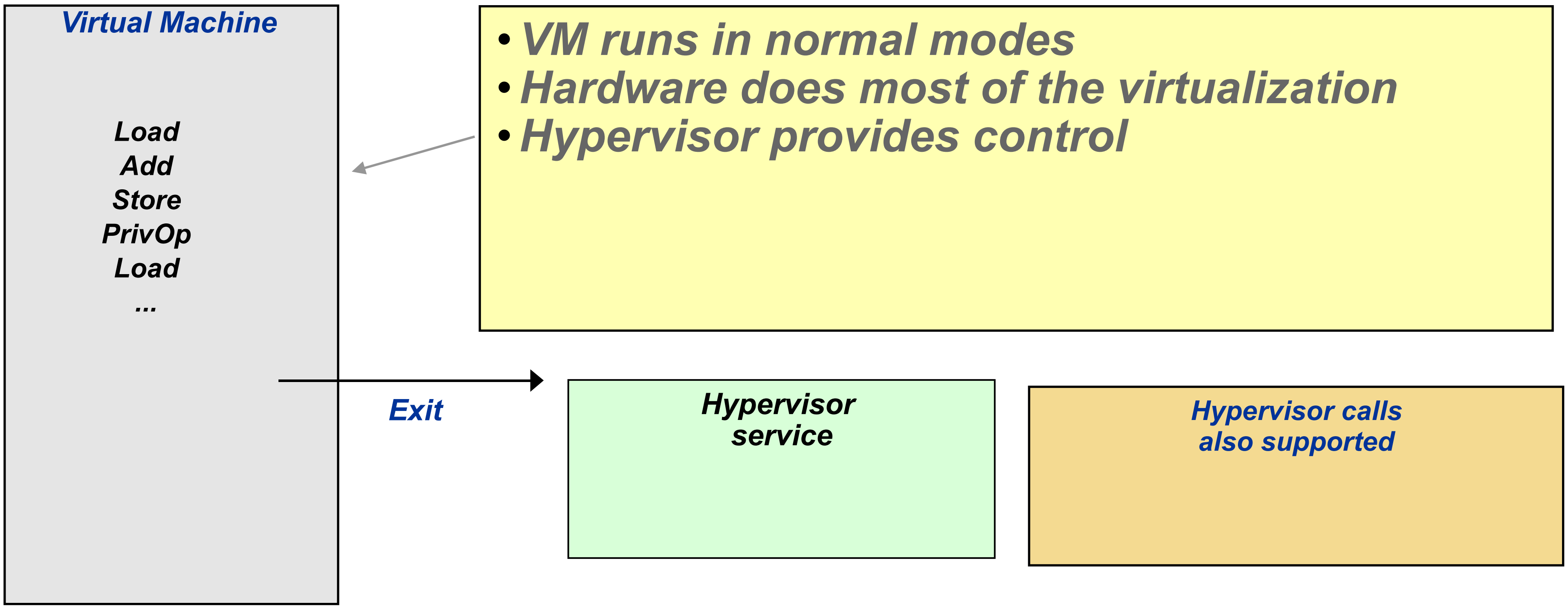
Examples: VMware, Microsoft VS
Benefits: Runs unmodified, translated OS
Issues: Substantial overhead

Paravirtualization



Examples: POWER VM, Xen
Benefits: High efficiency
Issues: OS must be modified

Direct Hardware Virtualization



Examples: System z LPAR, z/VM
Benefits: High efficiency, runs unmodified OS
Issues: Requires underlying hardware support

System x

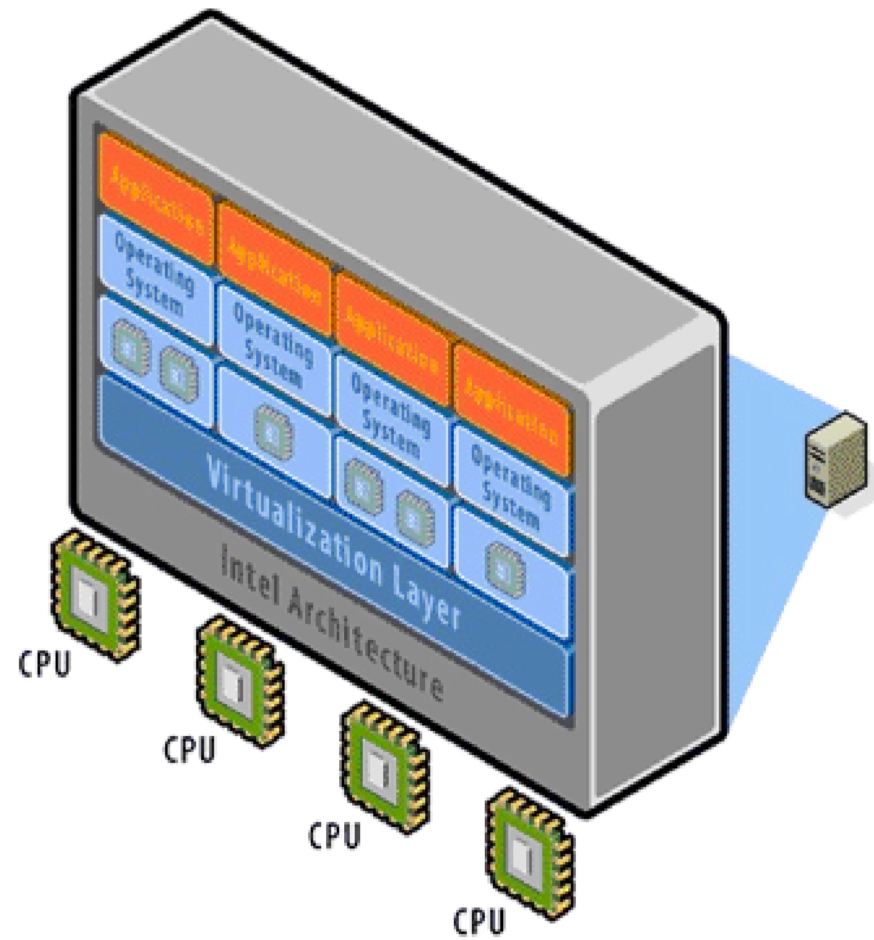
Virtualizing IA-32

- The IA-32 was not designed to be virtualized
- Many protected instructions are not required to be executed in protected mode
- There are a great deal of devices which must be supported

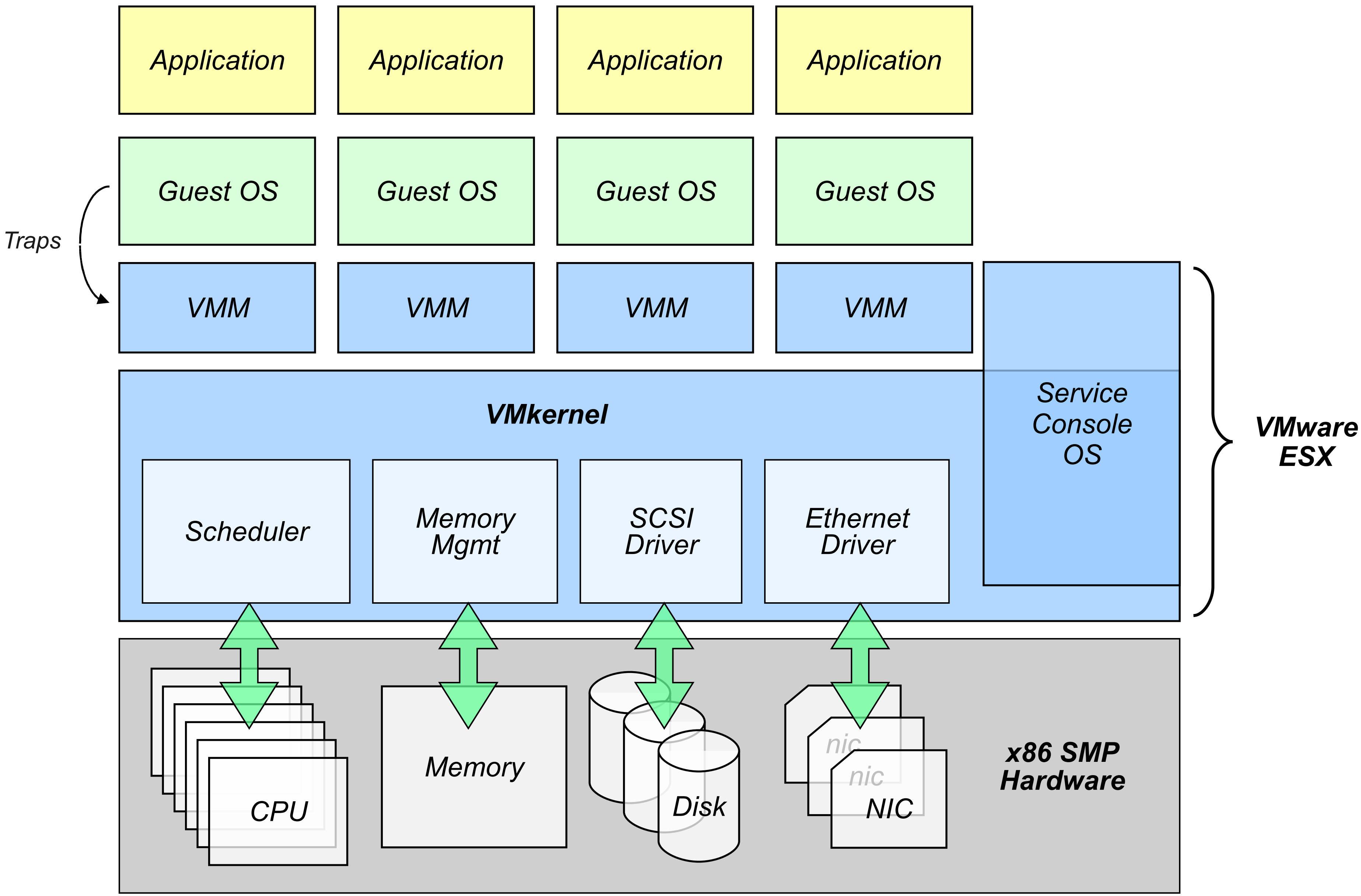
vmware®

VMWare Virtualization

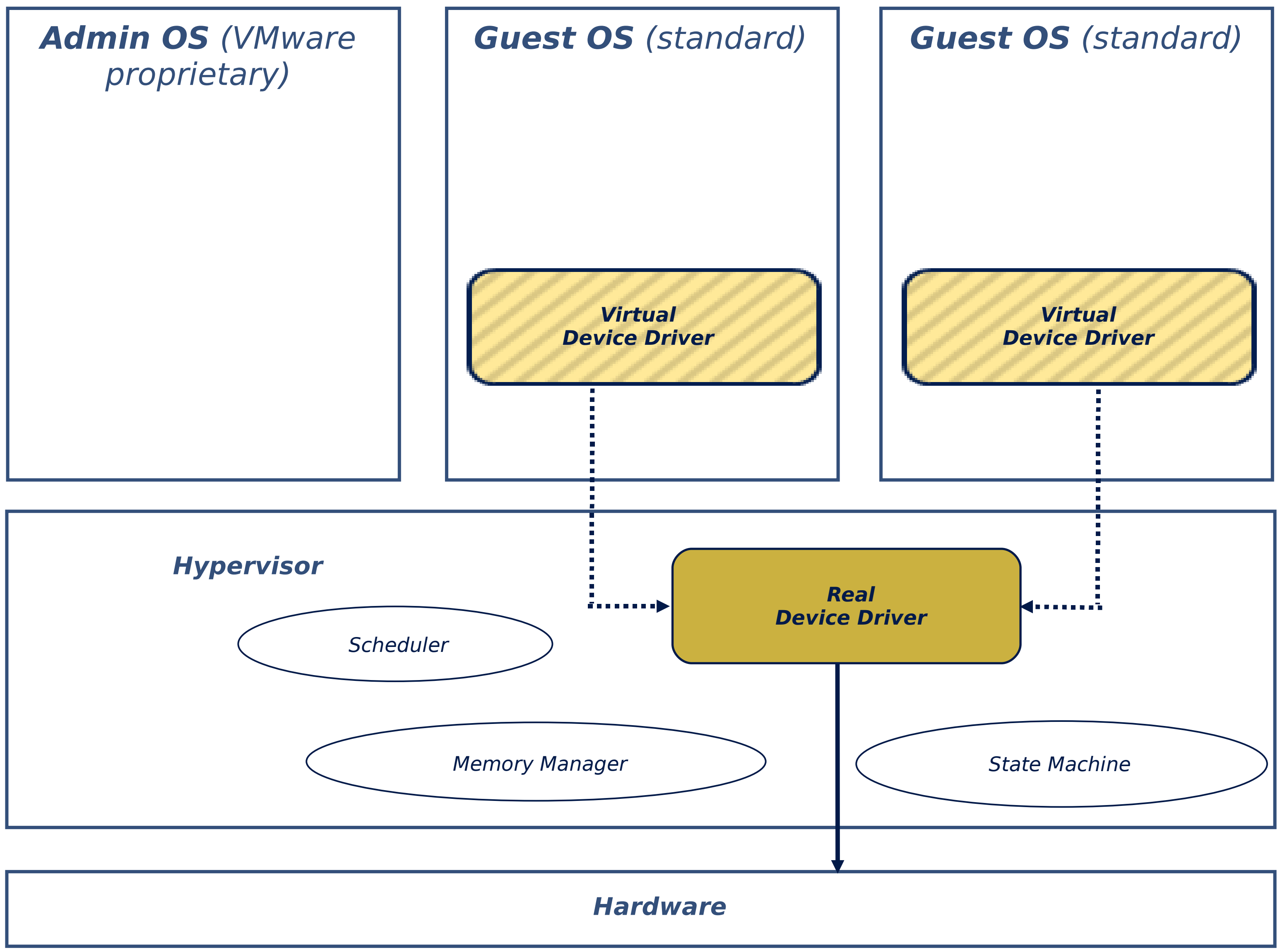
- CPU: Direct Execution w/ Binary Translation
- MEM: Shadow Table w/ Ballooning Driver
- I/O: Hosted Architecture or Limited Support



VMware ESX Server



Device Driver Model



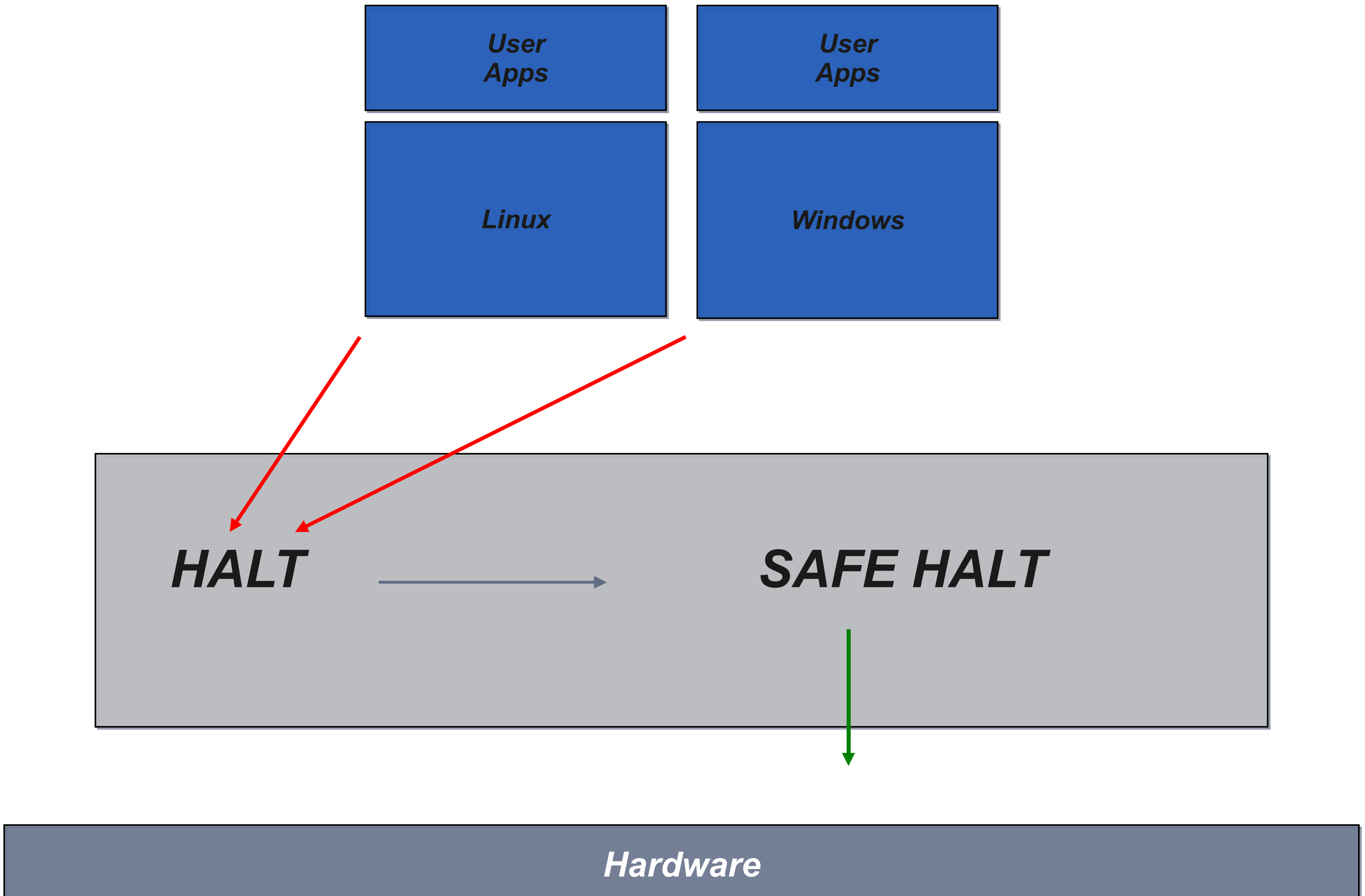
Vmware ESXi

ESXi VS ESX

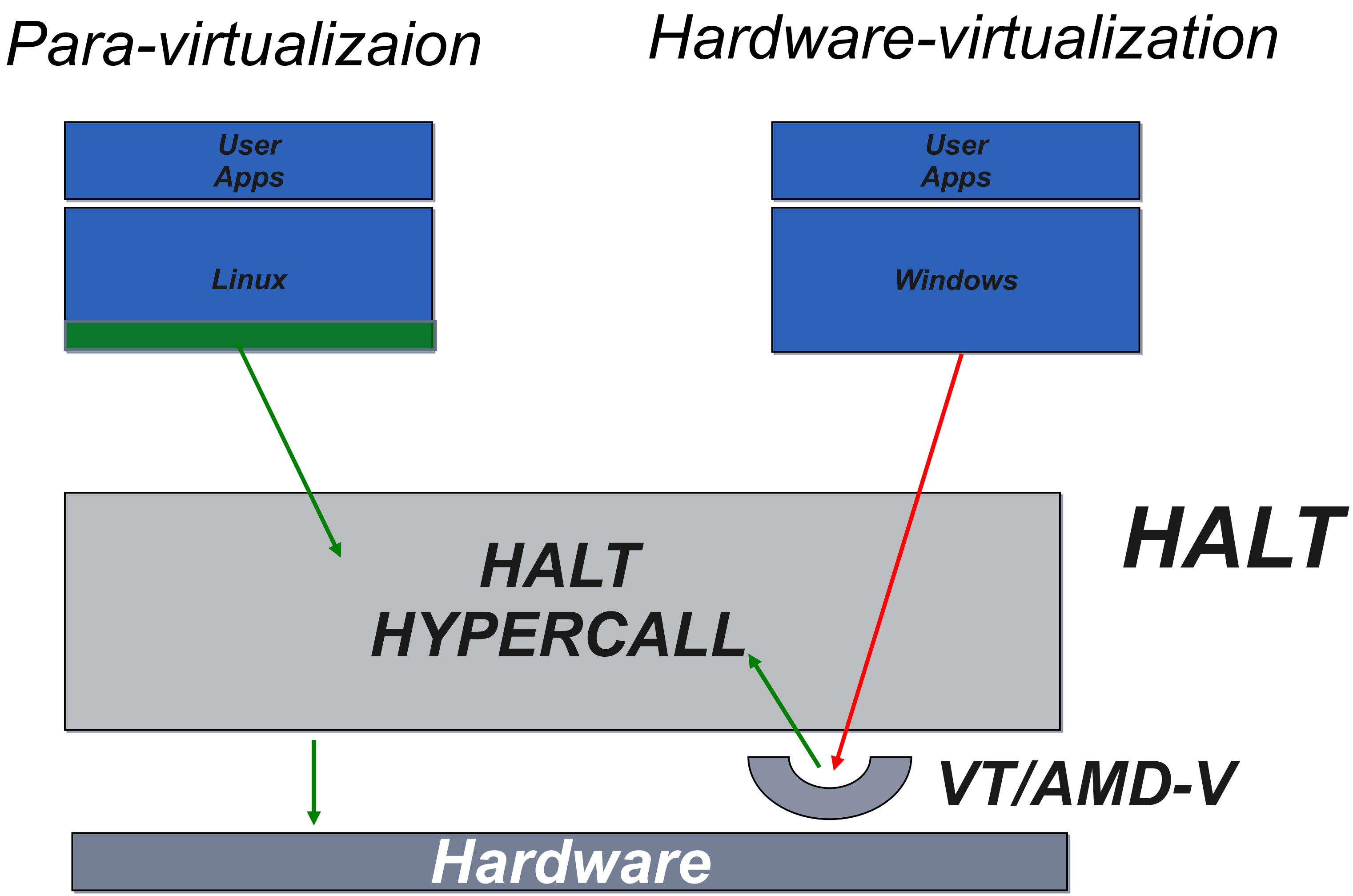
	ESXi	VI Foundation	VI Standard	VI Enterprise
Core hypervisor	Yes	Yes	Yes	Yes
Virtual SMP	Yes	Yes	Yes	Yes
VMFS	Yes	Yes	Yes	Yes
VirtualCenter Agent		Yes	Yes	Yes
Update Manager		Yes	Yes	Yes
Consolidated Backup		Yes	Yes	Yes
High Availability			Yes	Yes
VMotion				Yes
Storage VMotion				Yes
DRS				Yes
DPM				Yes

Xen Source™

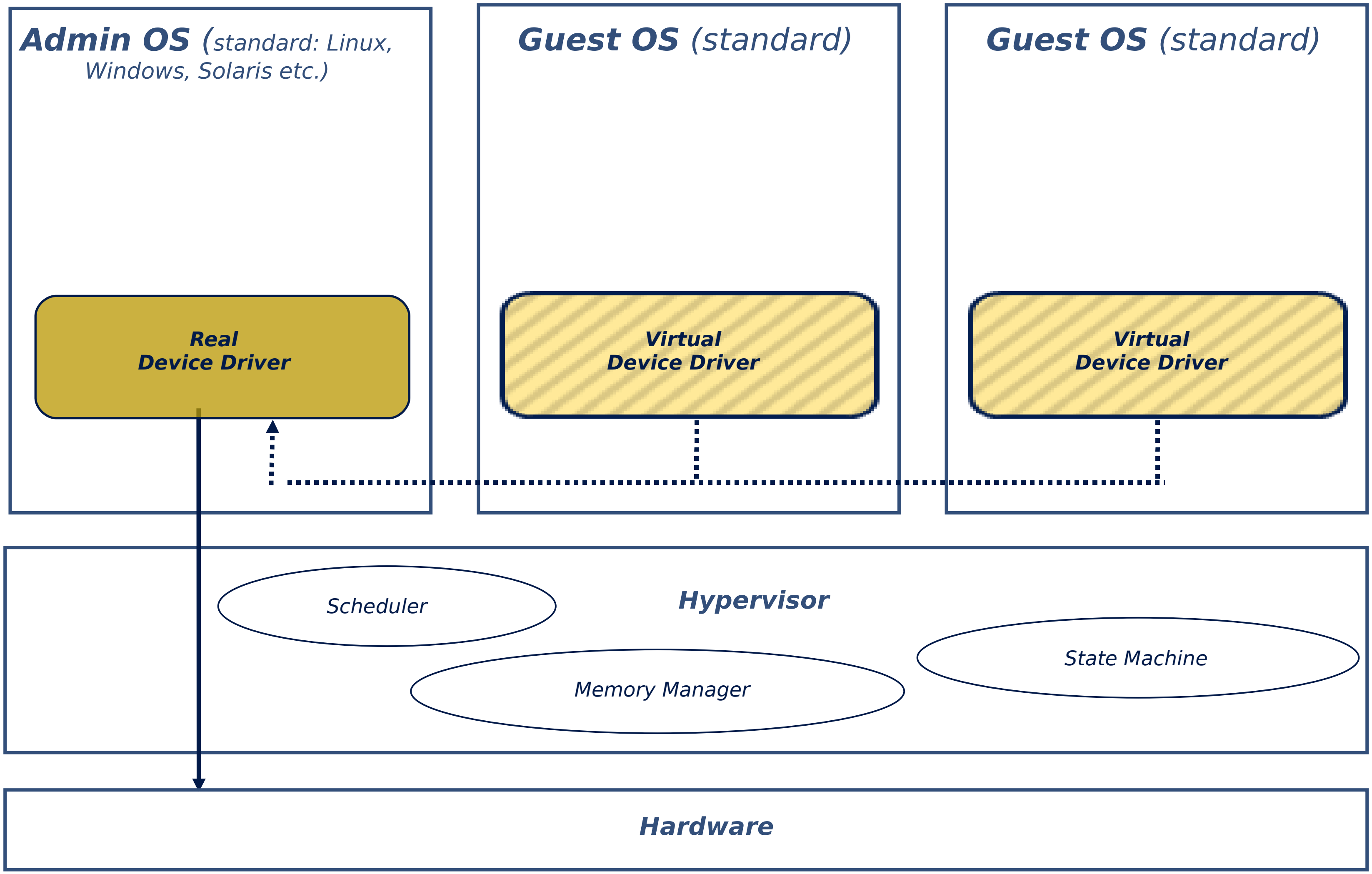
Xen full virtualization



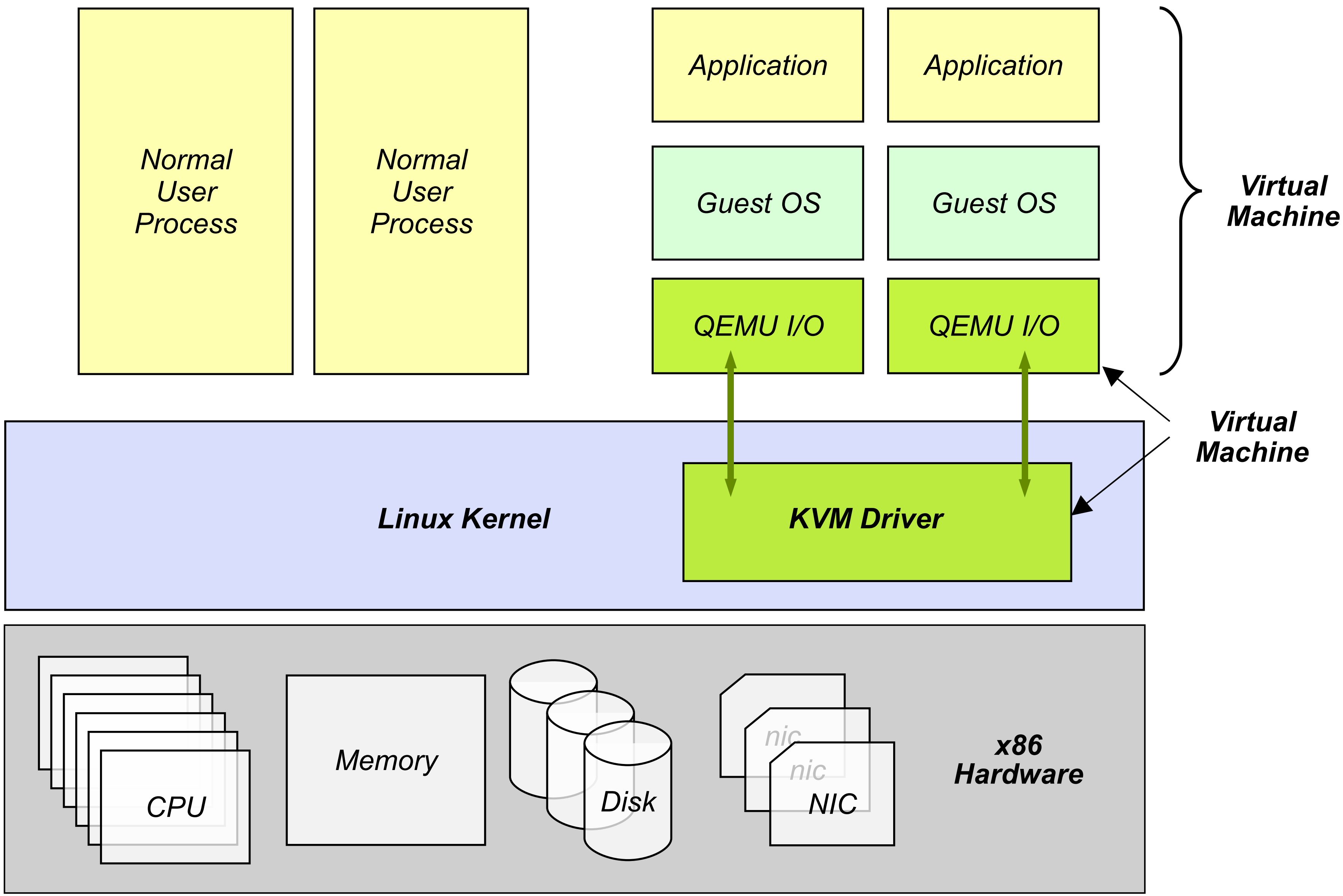
Xen para virtualization



Device Driver Model

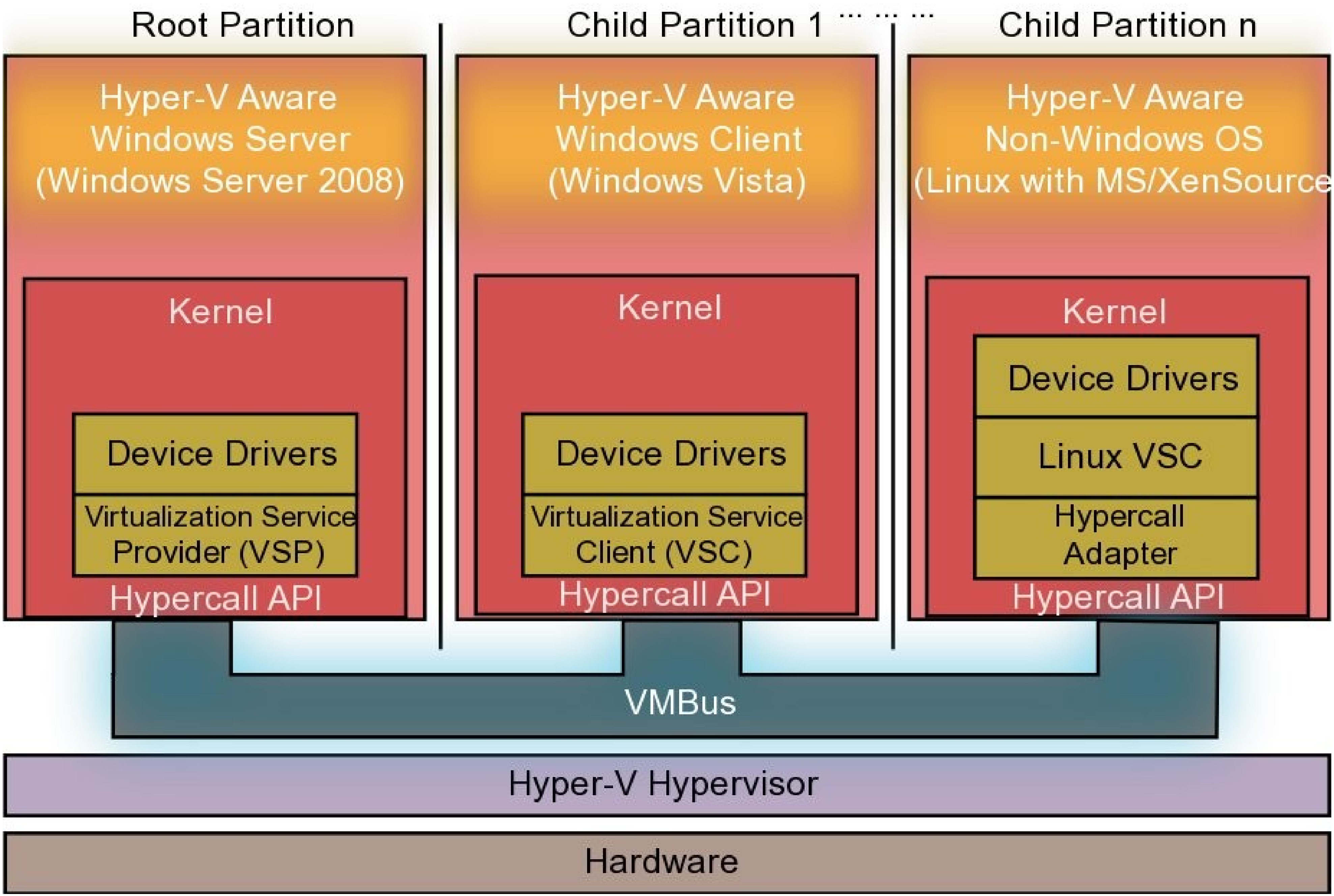


KVM


Normal
User
Process
Normal
User
Process
Application
Application
Guest OS
Guest OS
QEMU I/O
QEMU I/O
Virtual
Machine
Virtual
Machine
Linux Kernel
KVM Driver
CPU
Memory
Disk
nic
nic
NIC
x86
Hardware

Sun x/VM (Virtual Box)

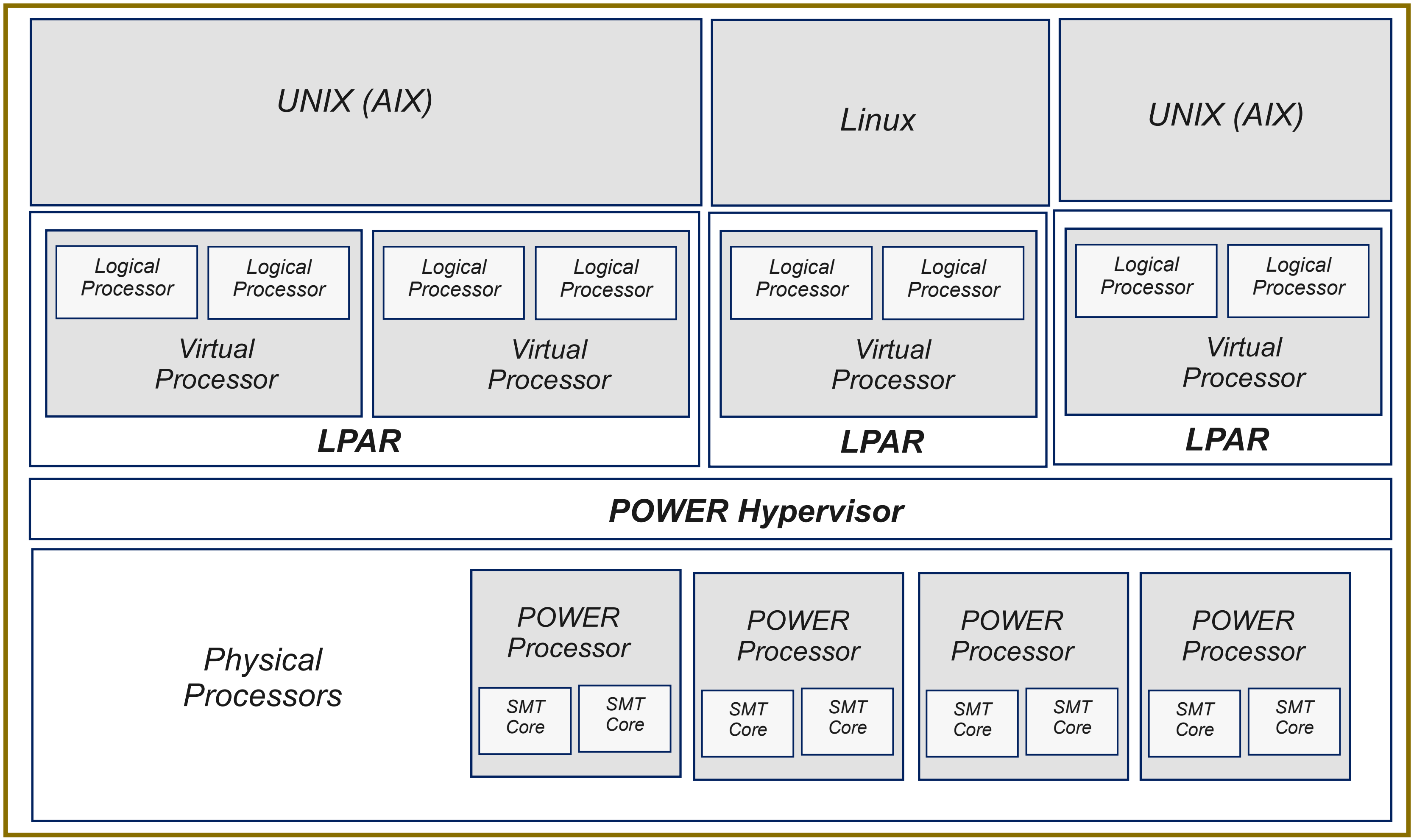
Microsoft Hyper V

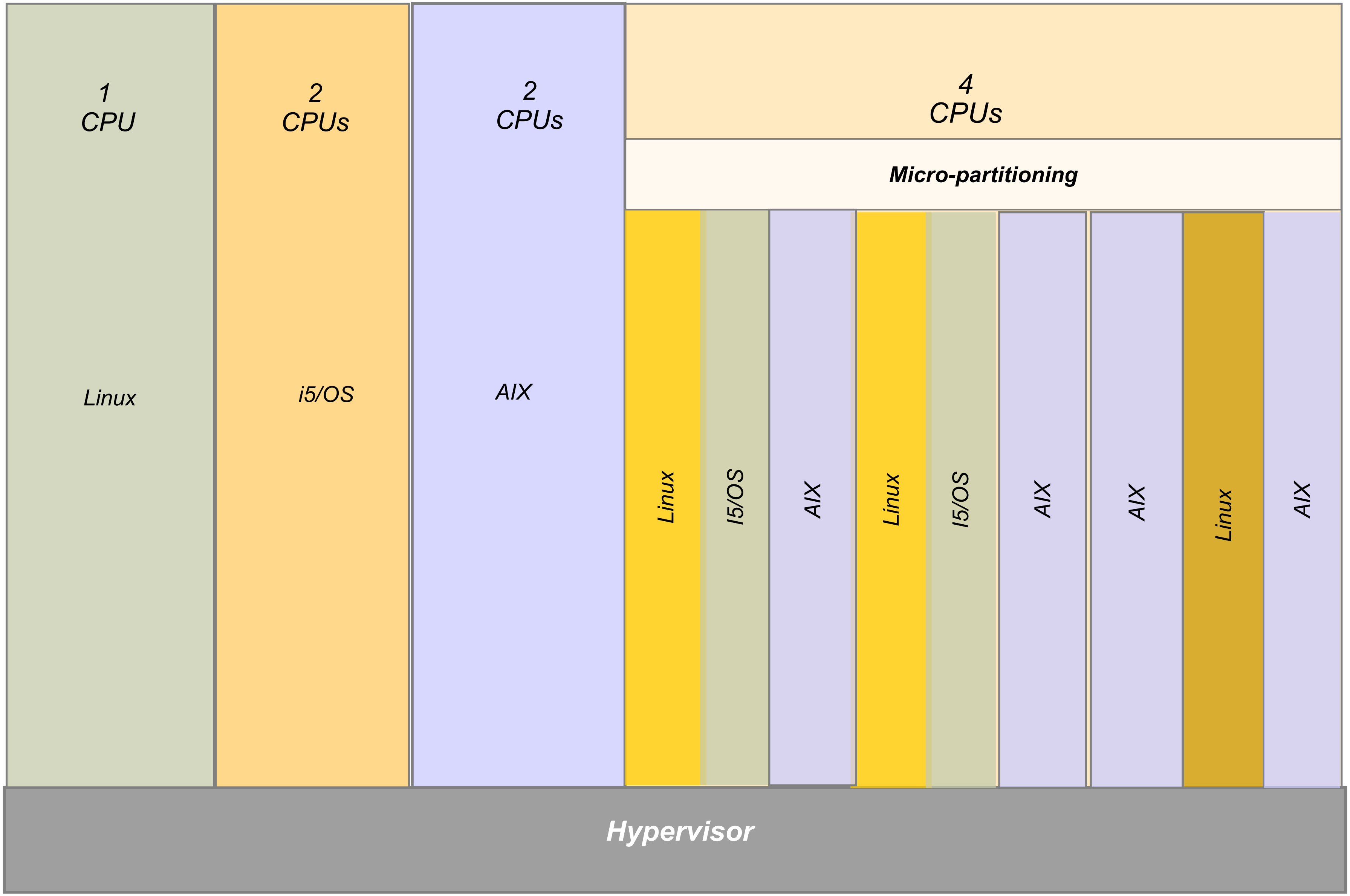


System p

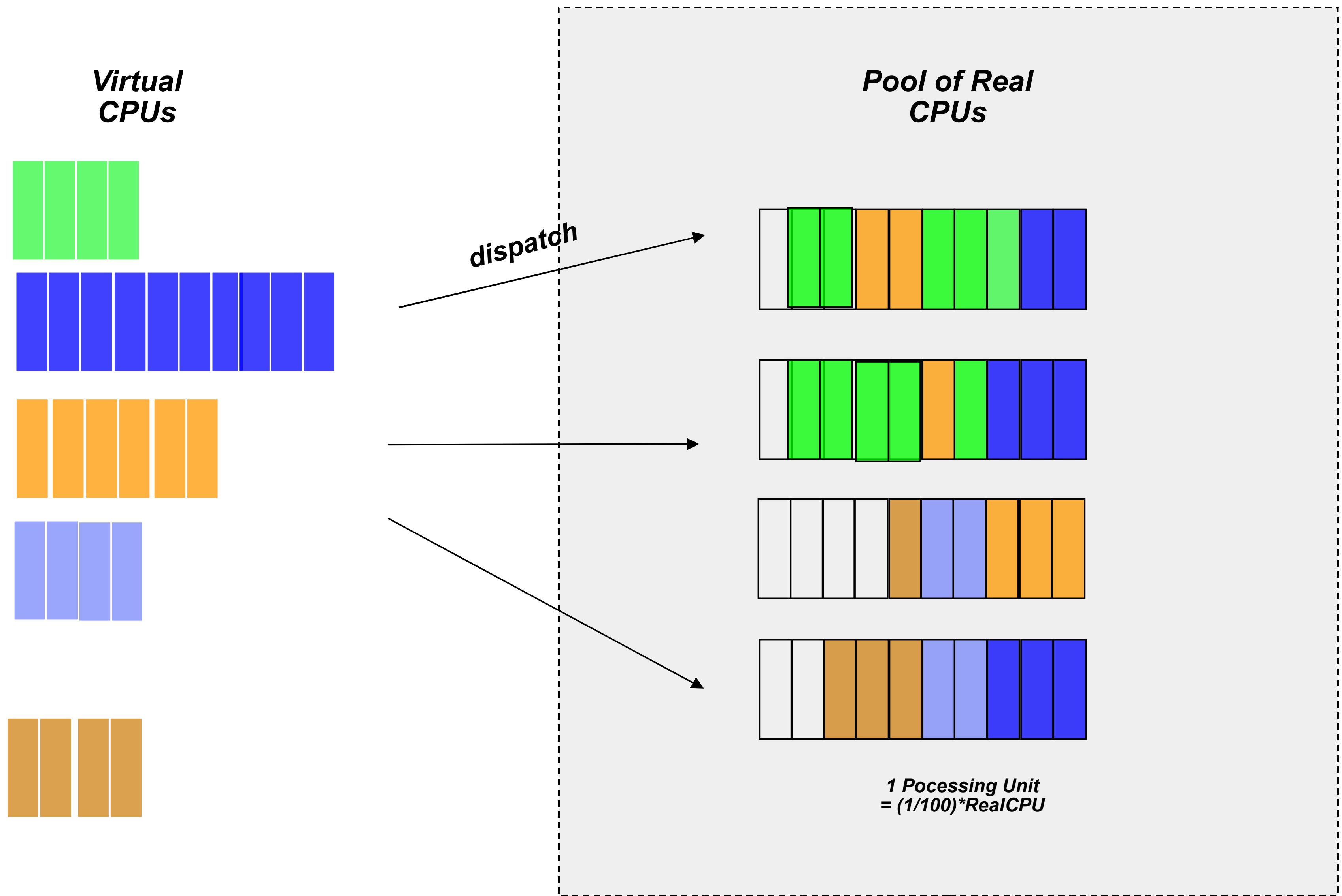
PowerVM™

LPARs

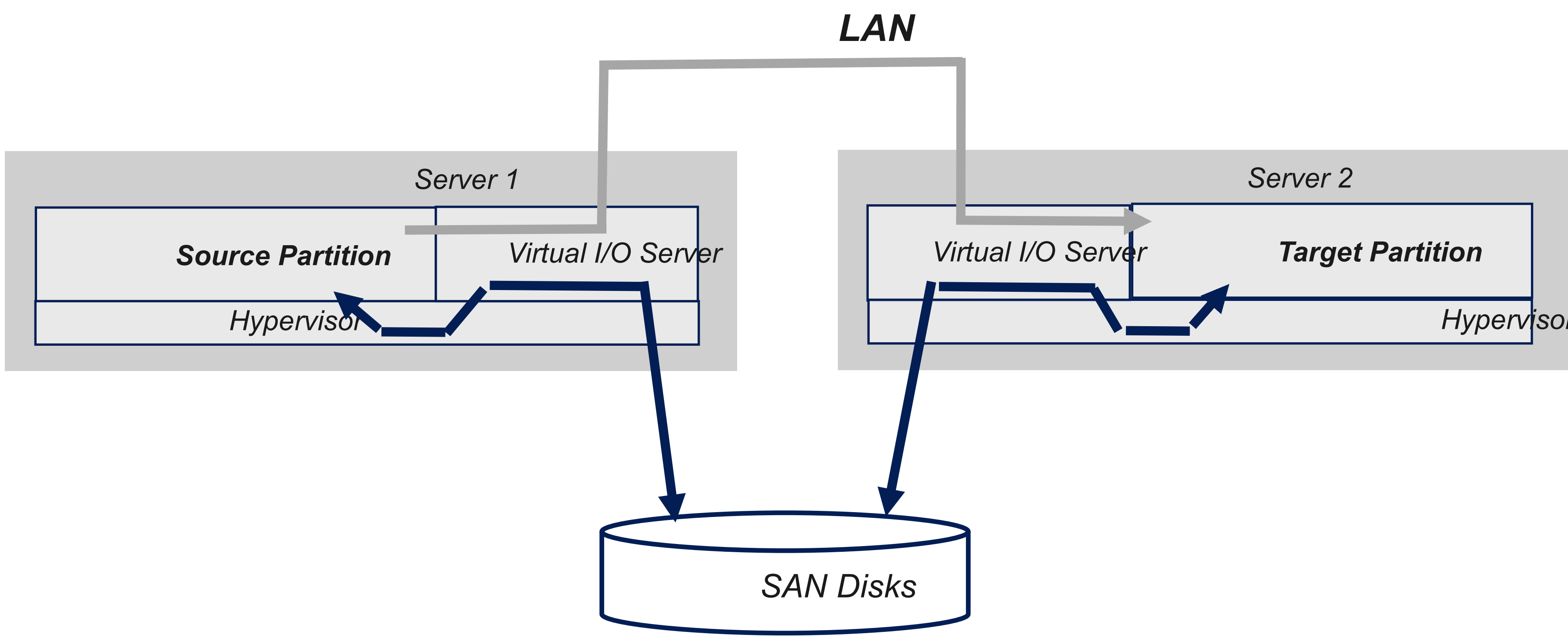

UNIX (AIX)
Linux
UNIX (AIX)
Logical Processor
Logical Processor
Virtual Processor
Logical Processor
Logical Processor
Virtual Processor
LPAR
Logical Processor
Logical Processor
Virtual Processor
LPAR
Logical Processor
Logical Processor
Virtual Processor
LPAR
POWER Hypervisor
Physical Processors
POWER Processor
SMT Core
SMT Core
POWER Processor
SMT Core
SMT Core
POWER Processor
SMT Core
SMT Core
POWER Processor
SMT Core
SMT Core


1
CPU
2
CPUs
2
CPUs
4
CPUs
Micro-partitioning
Linux
i5/OS
AIX
Linux
I5/OS
AIX
Linux
I5/OS
AIX
AIX
Linux
AIX
Hypervisor

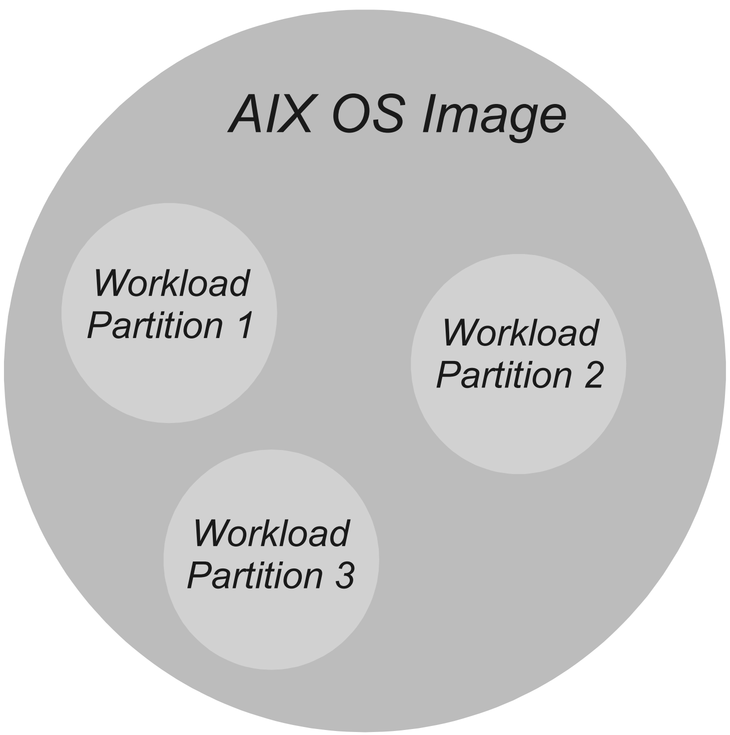
Micropartition



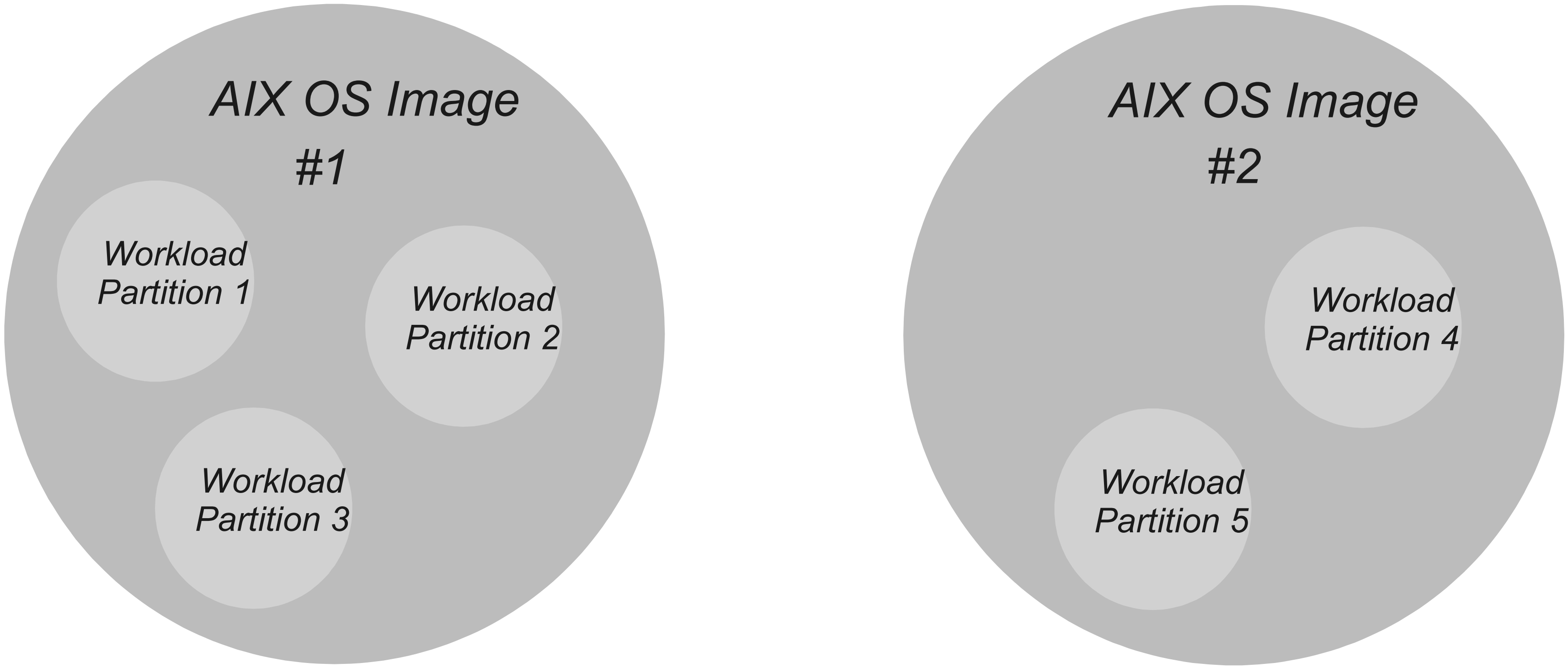
IBM System p Servers
Live Partition Mobility



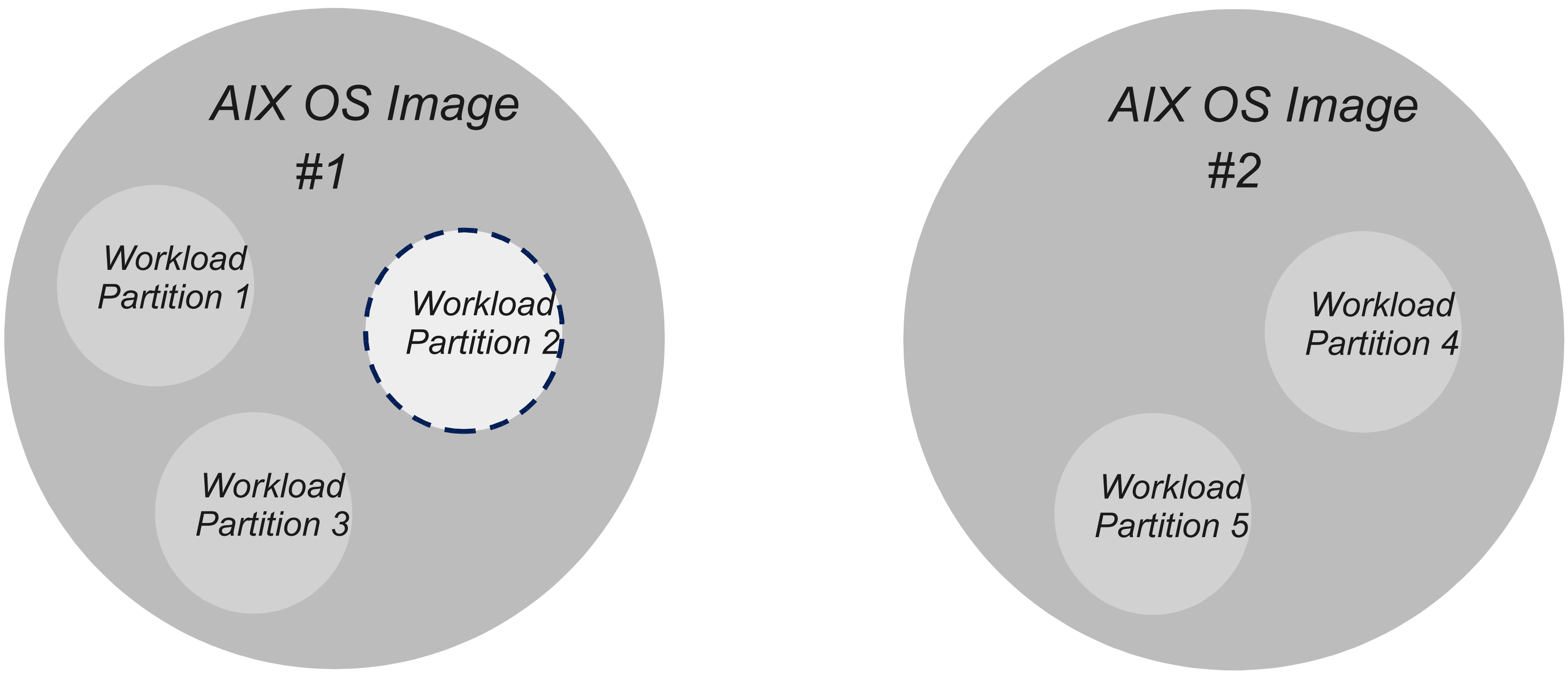
Workload Partitions (WPARs)



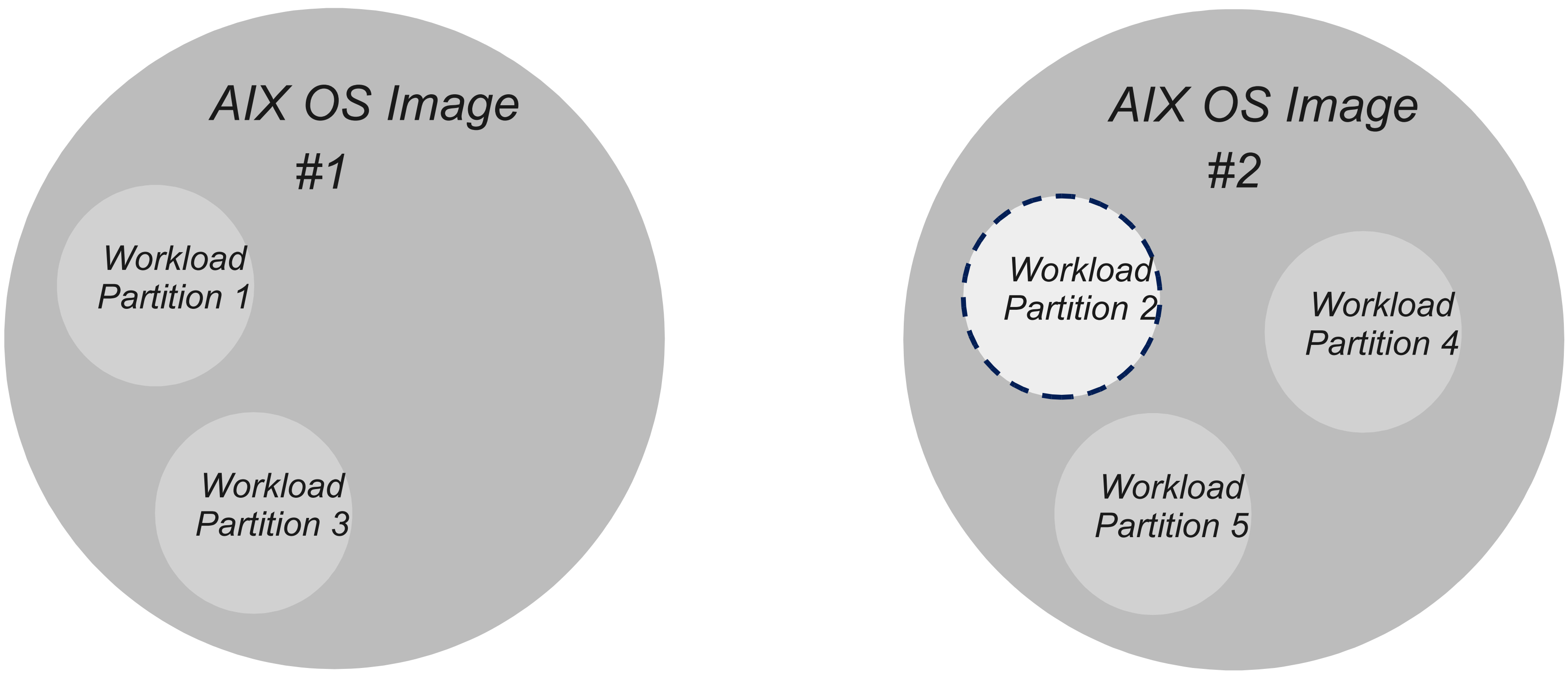
Live Application Mobility



Live Application Mobility



Live Application Mobility



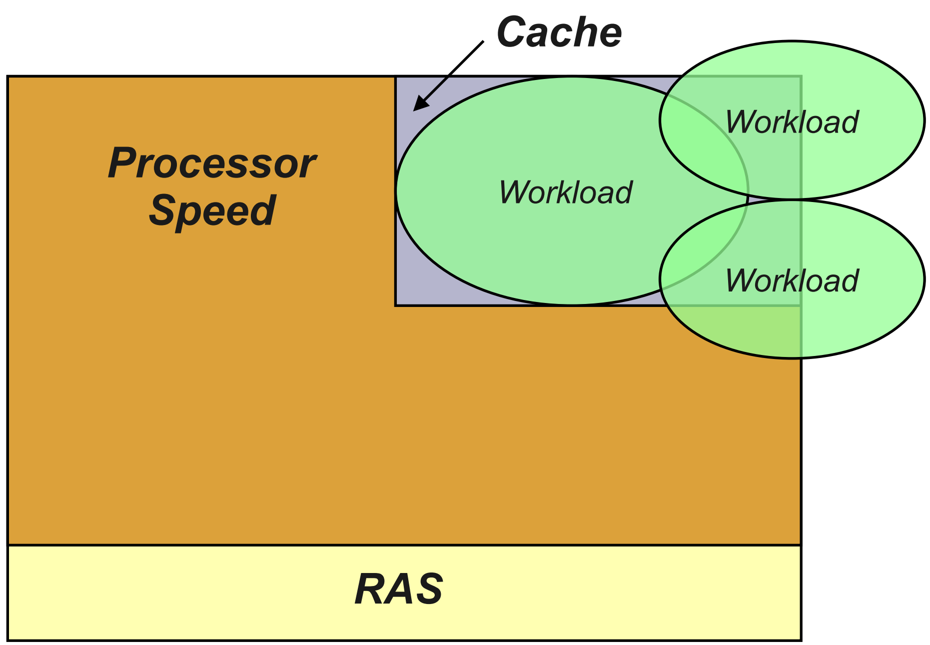
Live Application Mobility

System z

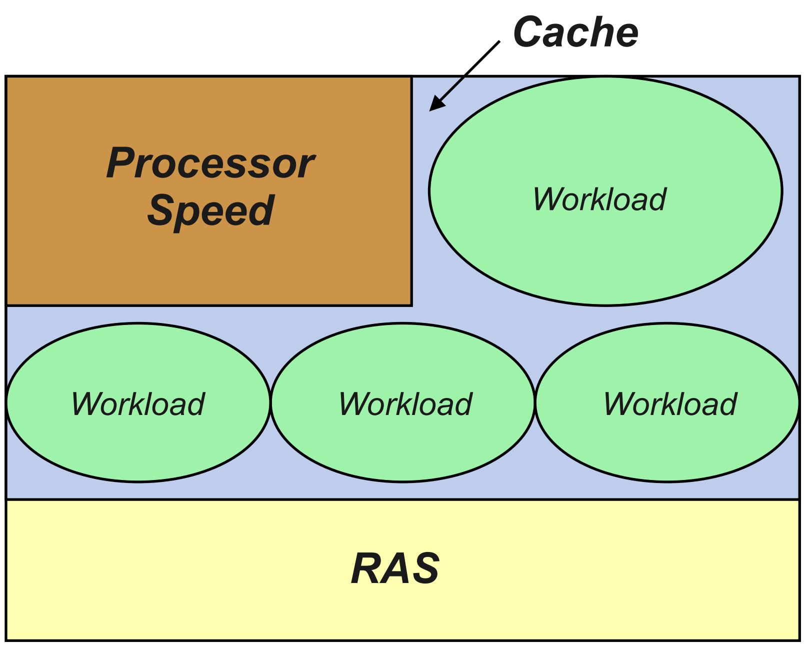
Virtualizing System/360

Chip Design Affects Virtualization Capabilities

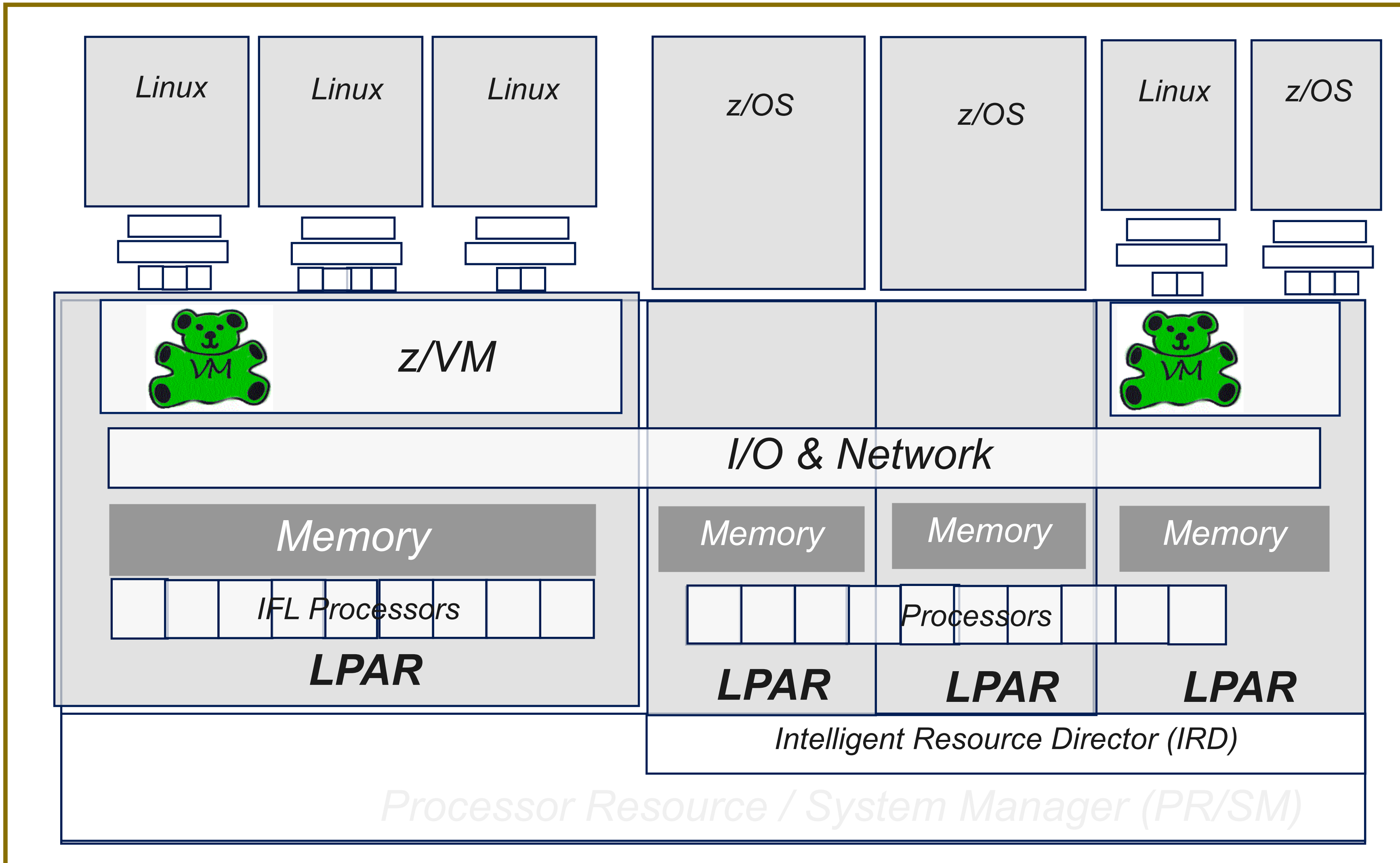
Replicated Server Chip Design



Consolidated Server Chip Design

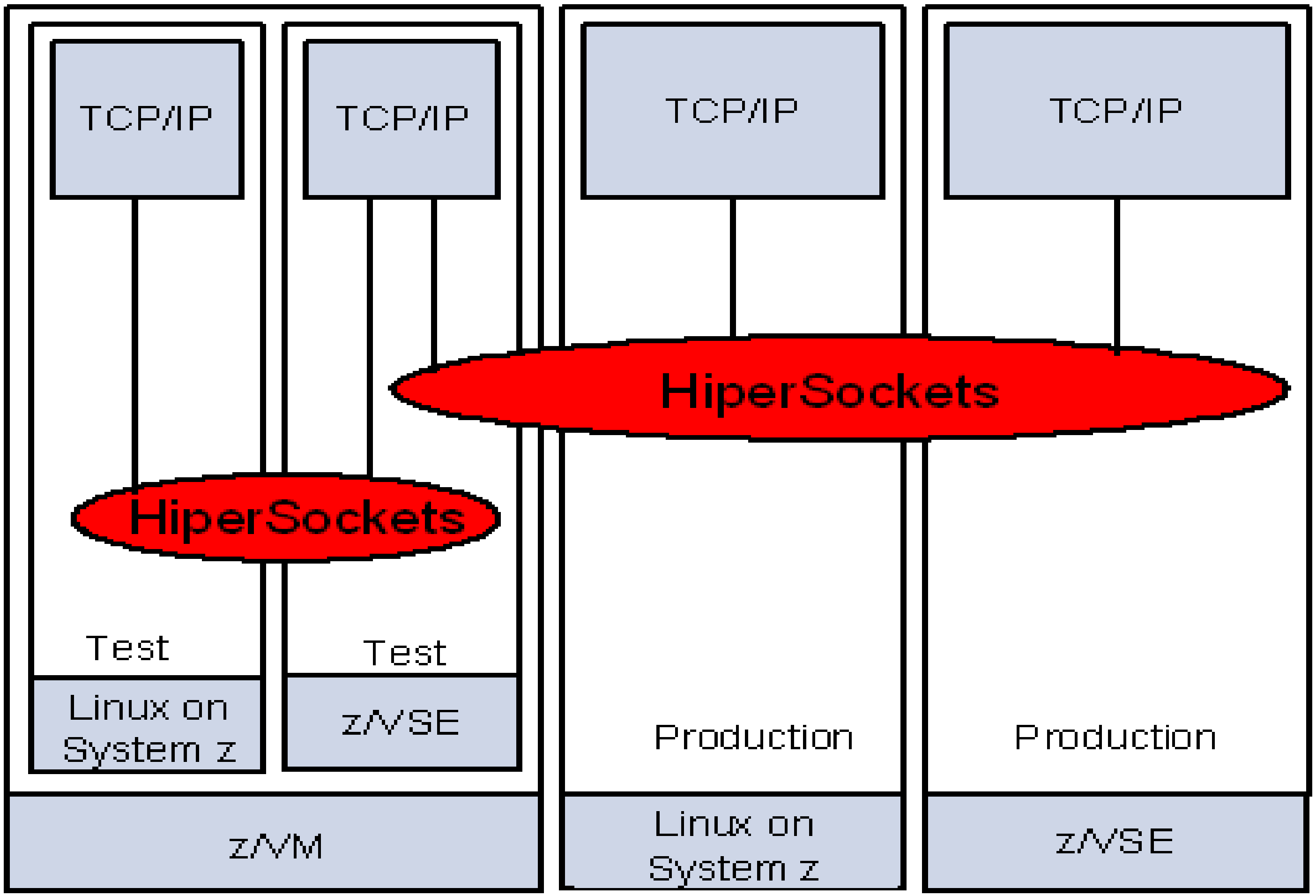


Z/VM

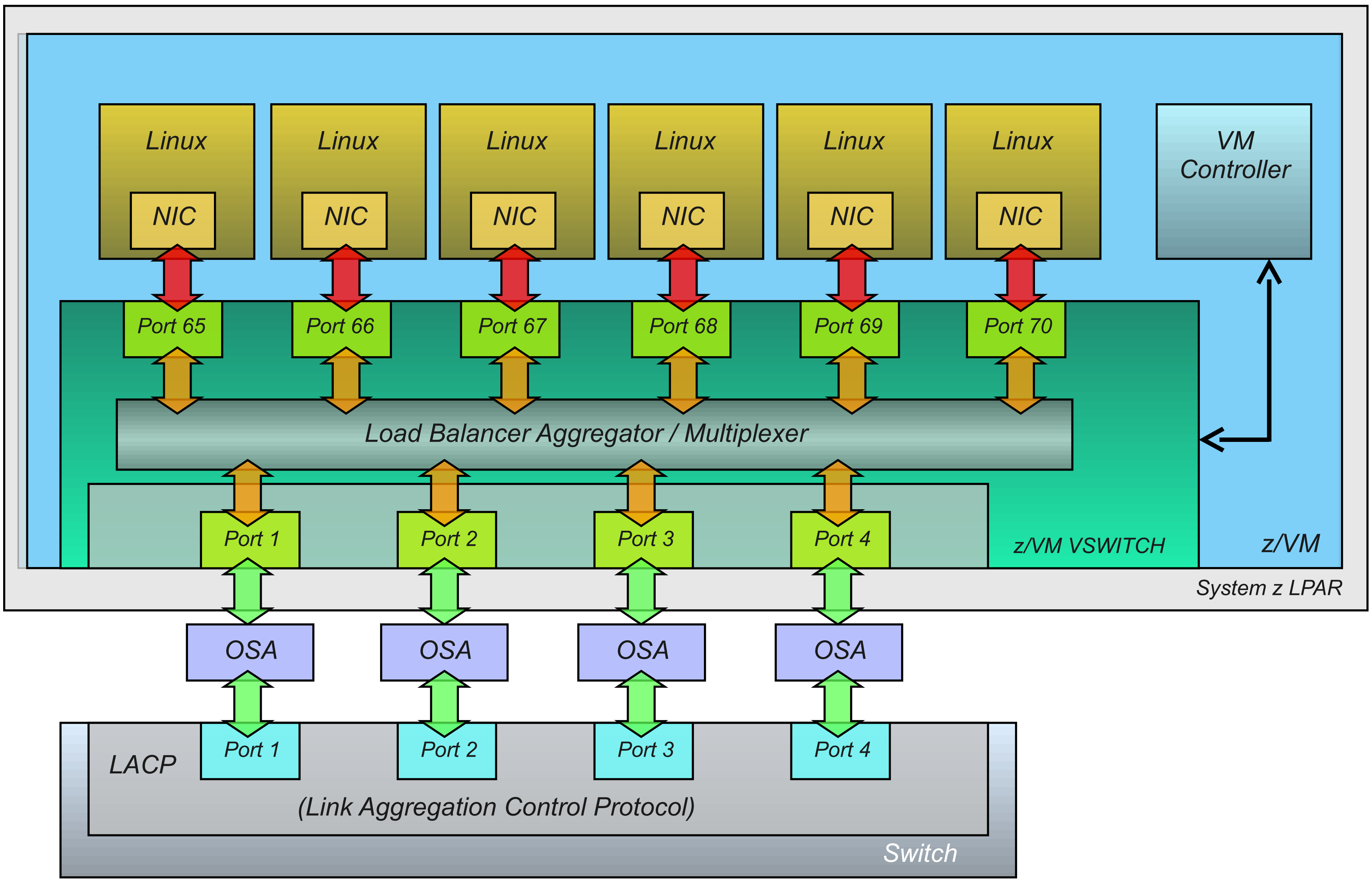


Linux
Linux
Linux
z/OS
z/OS
Linux
z/OS
z/VM
I/O & Network
Memory
Memory
Memory
Memory
IFL Processors
Processors
LPAR
LPAR
LPAR
LPAR
Intelligent Resource Director (IRD)
Processor Resource / System Manager (PR/SM)

HiperSockets



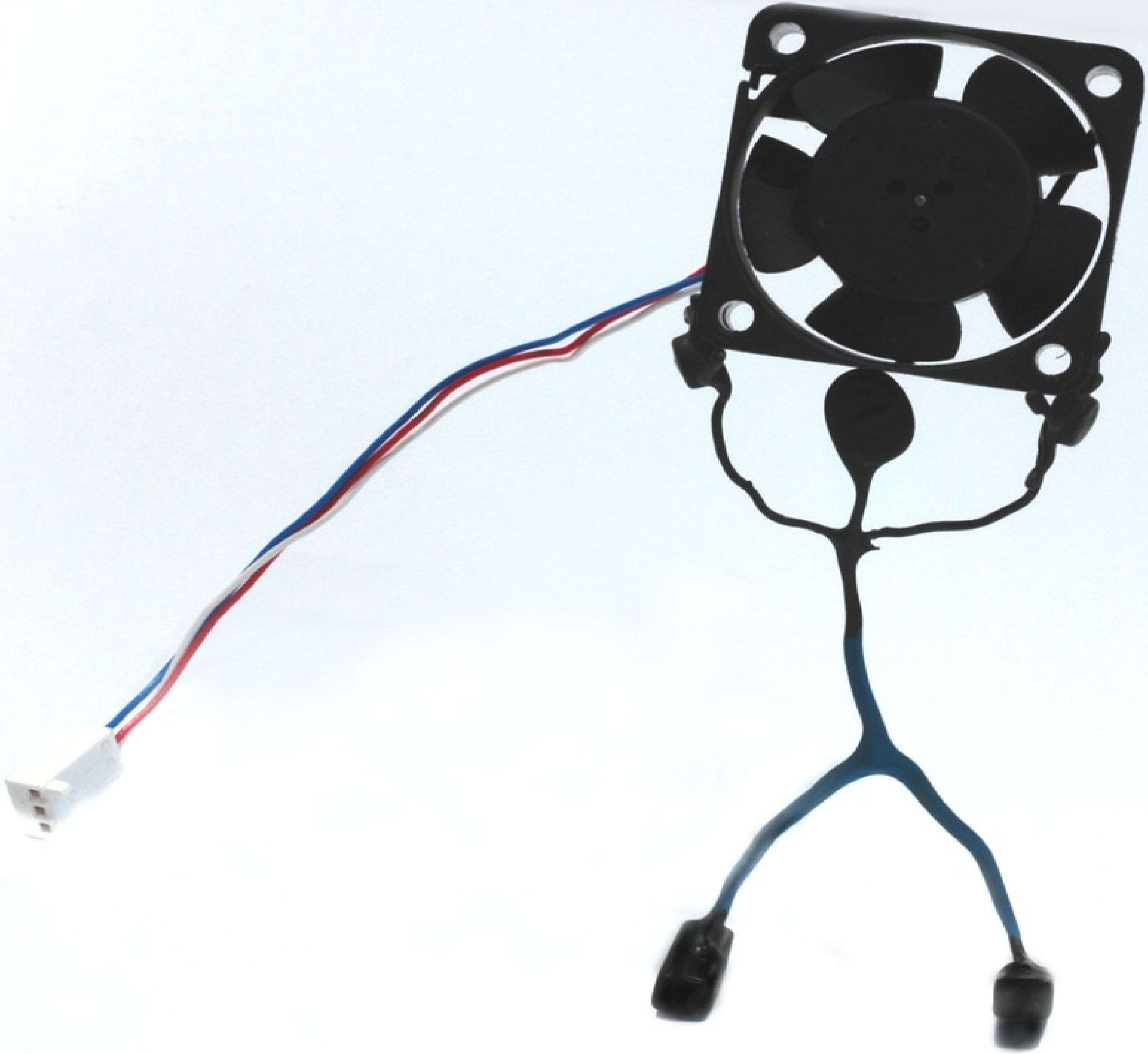
z/VM Virtual Switch



How do I choose?

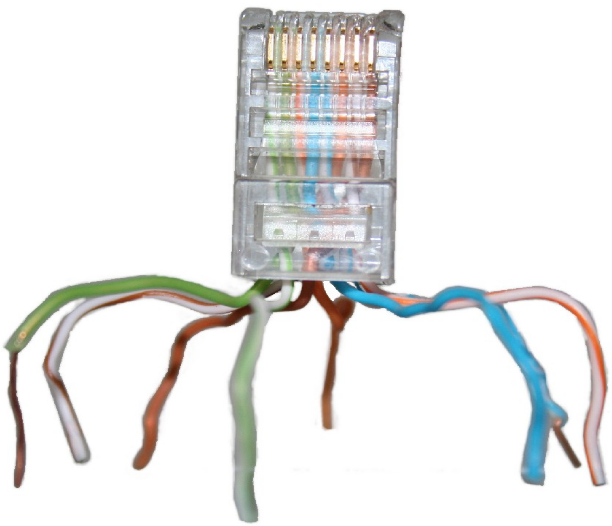

What are
the pain
points


HOTEL

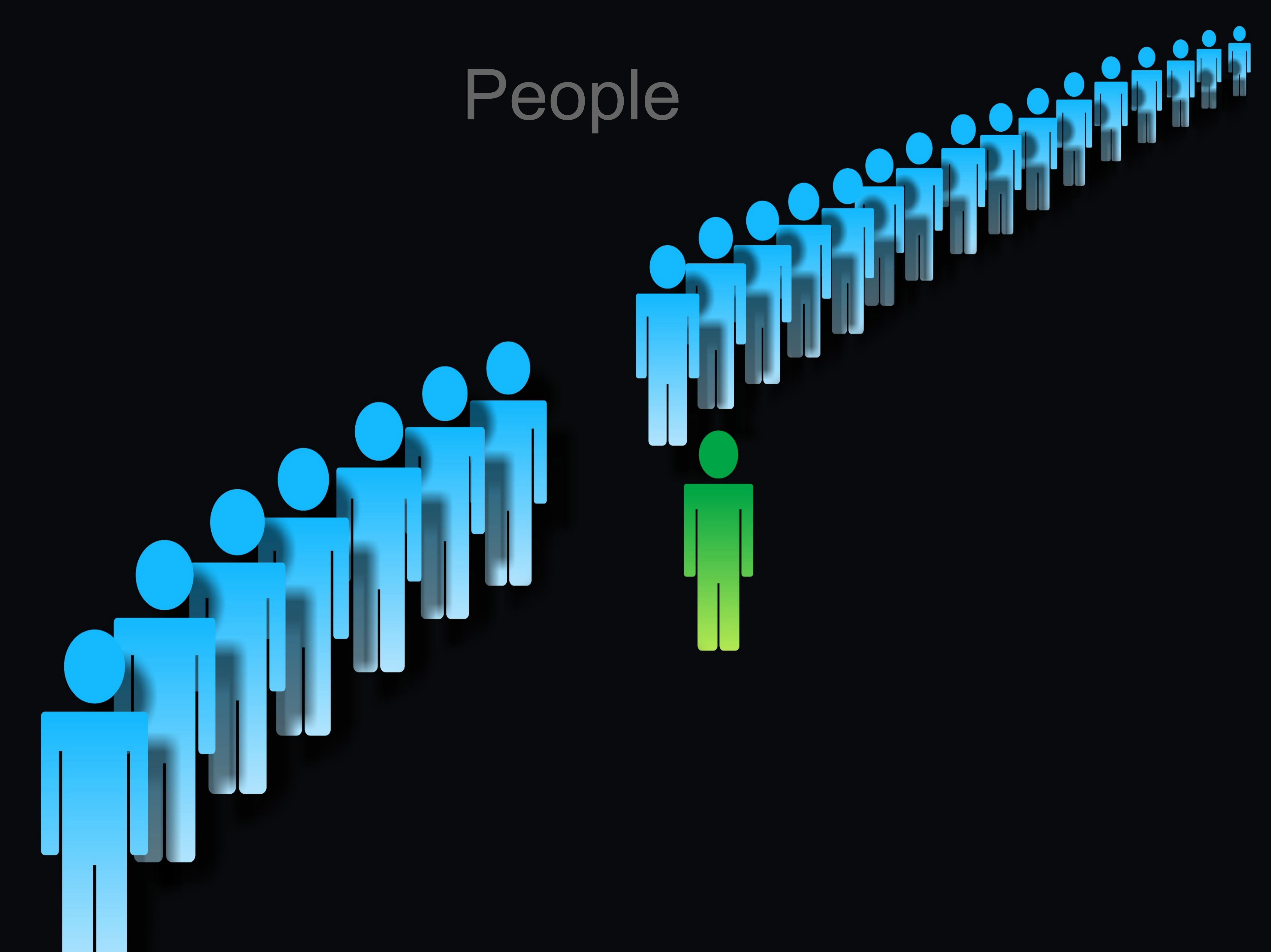


Cooling

Lan Spiders



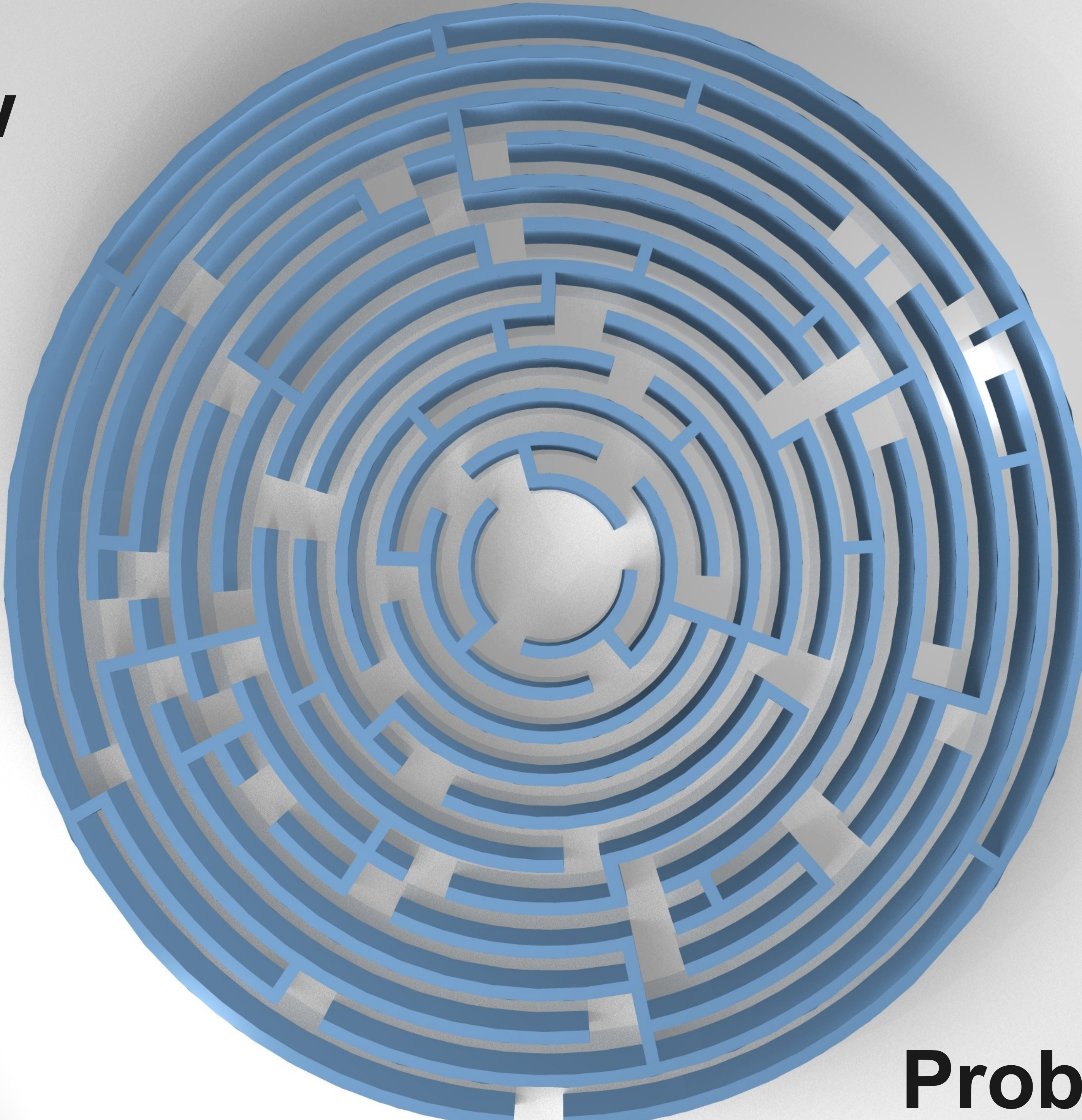
It's ok to laugh, its a joke.


People


DEAD
END
MINIMUM
MAINTENANCE
ROAD
TRAVEL AT YOUR OWN RISK

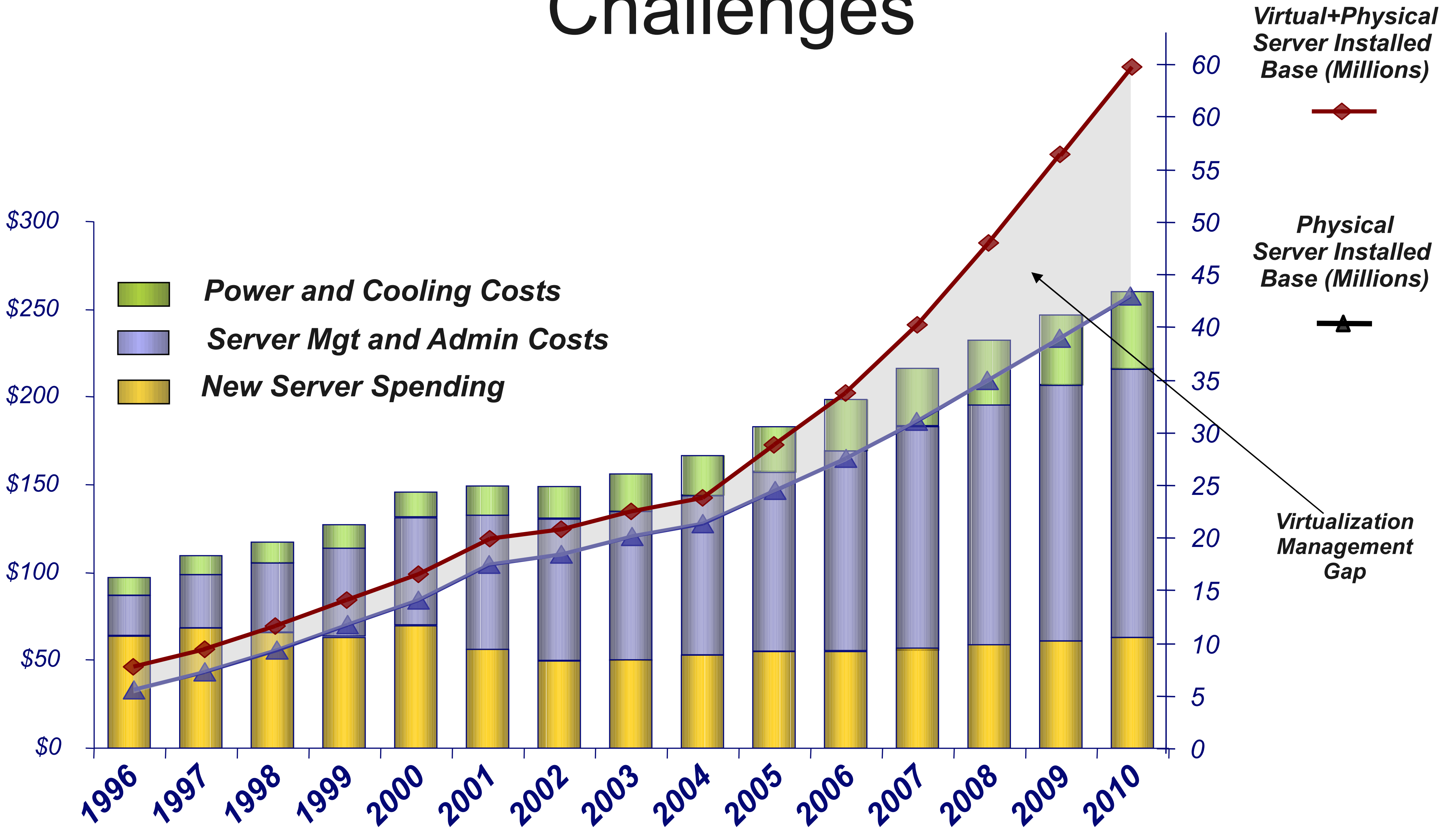
Know your workload

New



Problems

New Management Challenges



Source: IDC, May 2006

Agent Proliferation

Where is my data ?

Accountability

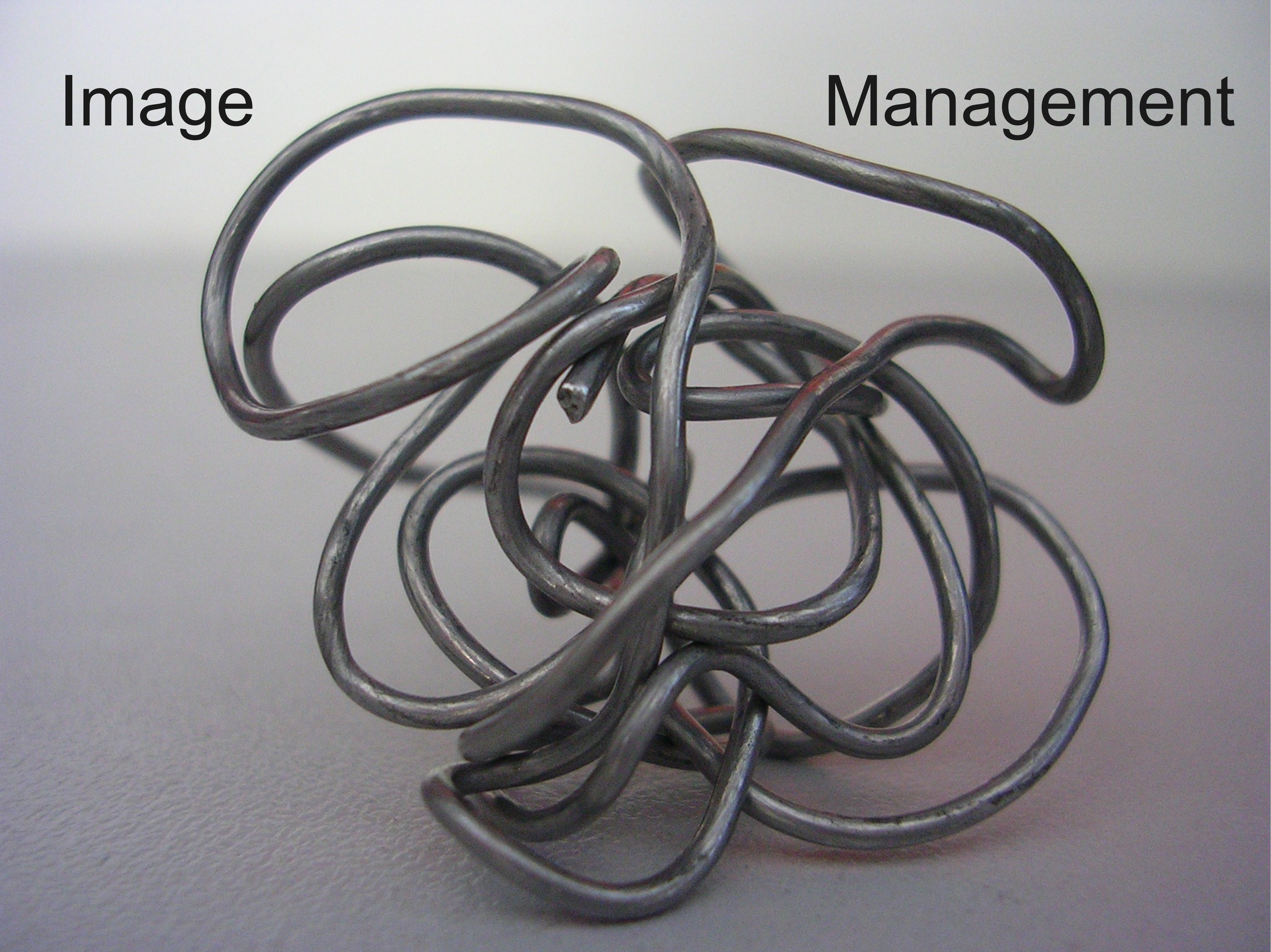
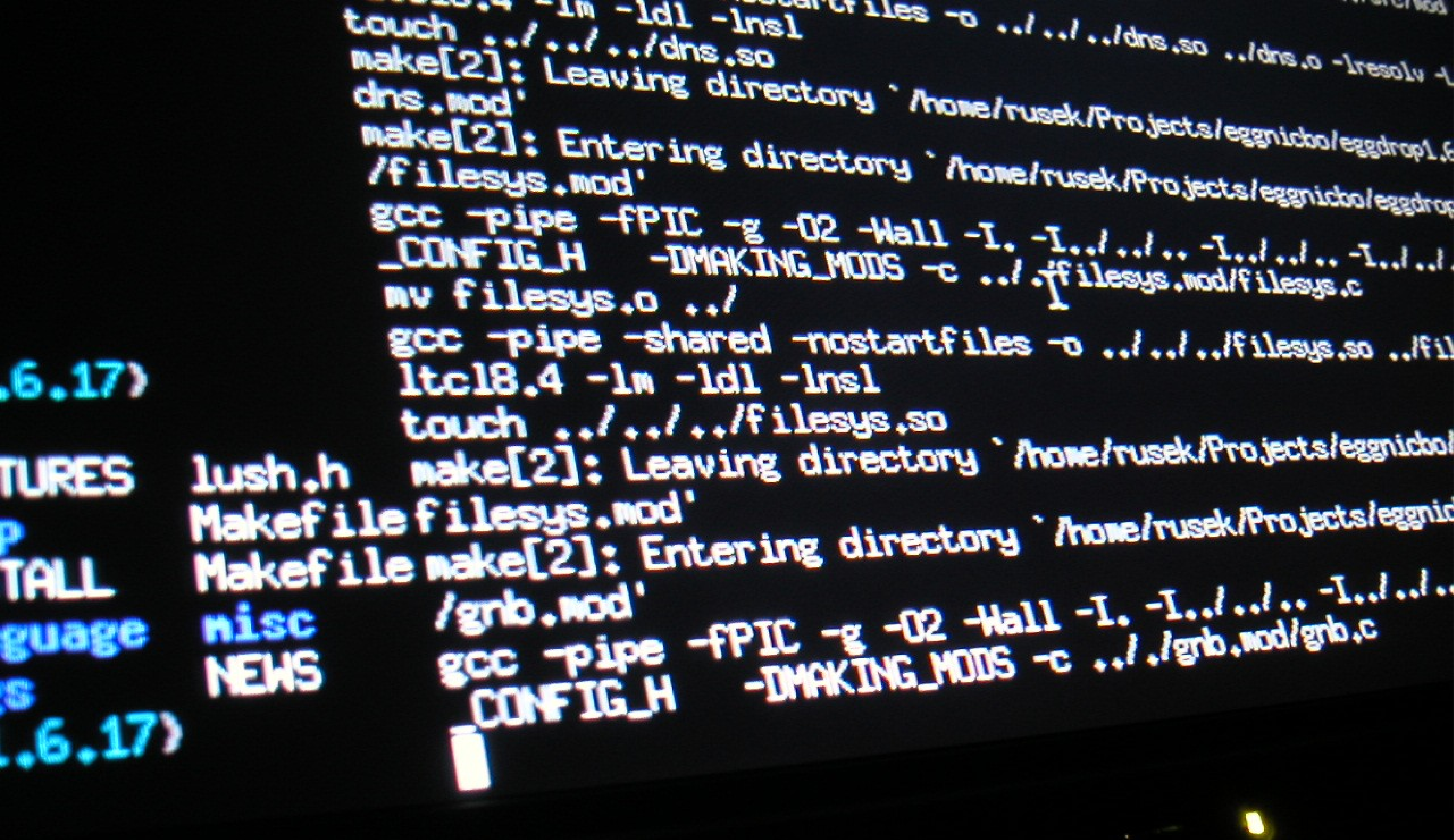


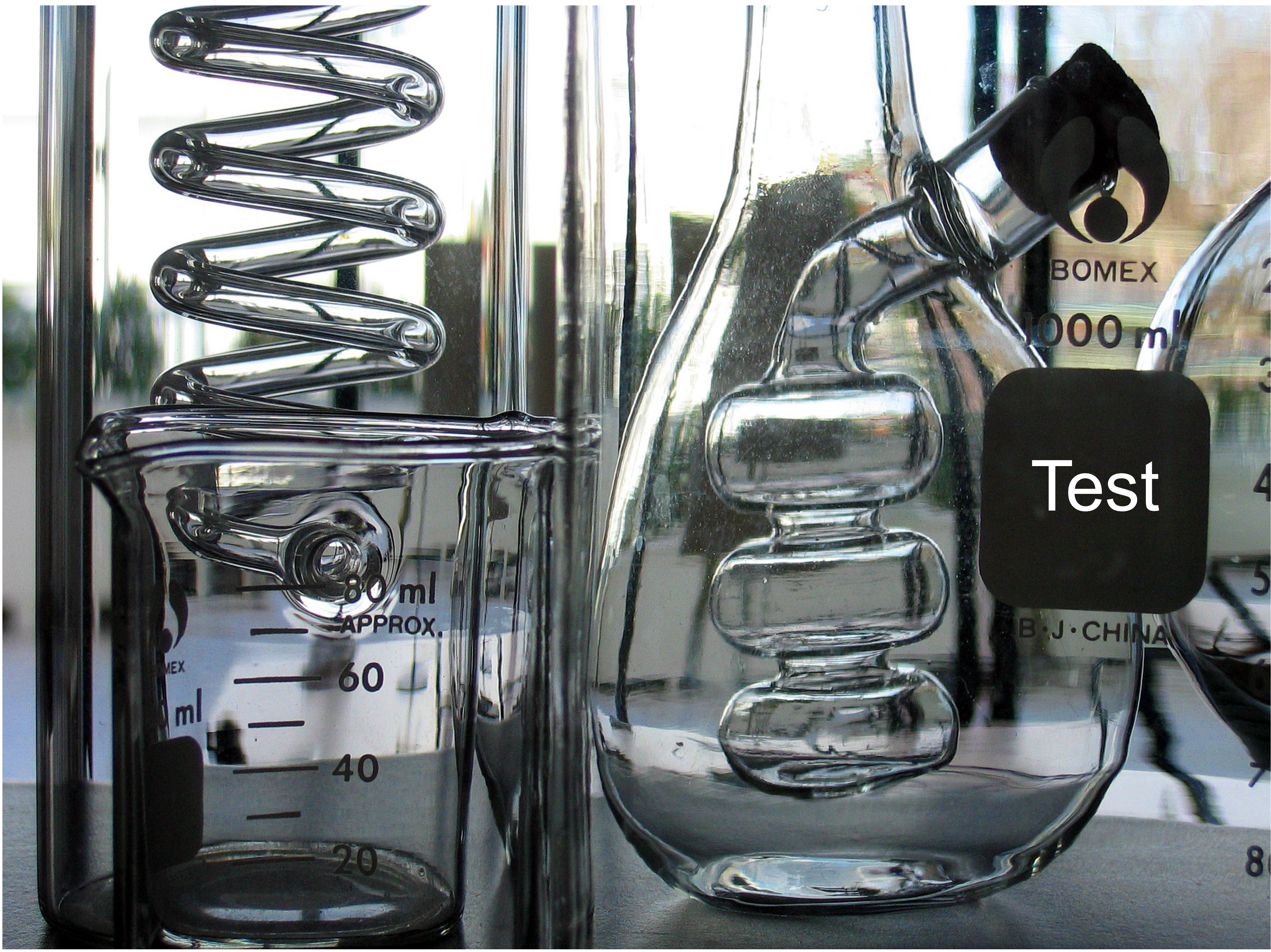
Image
Management

Where to start ?


Go for the low hanging fruit



Development


80 ml
APPROX.
60
40
20
BOMEX
1000 ml
Test
B·J·CHINA

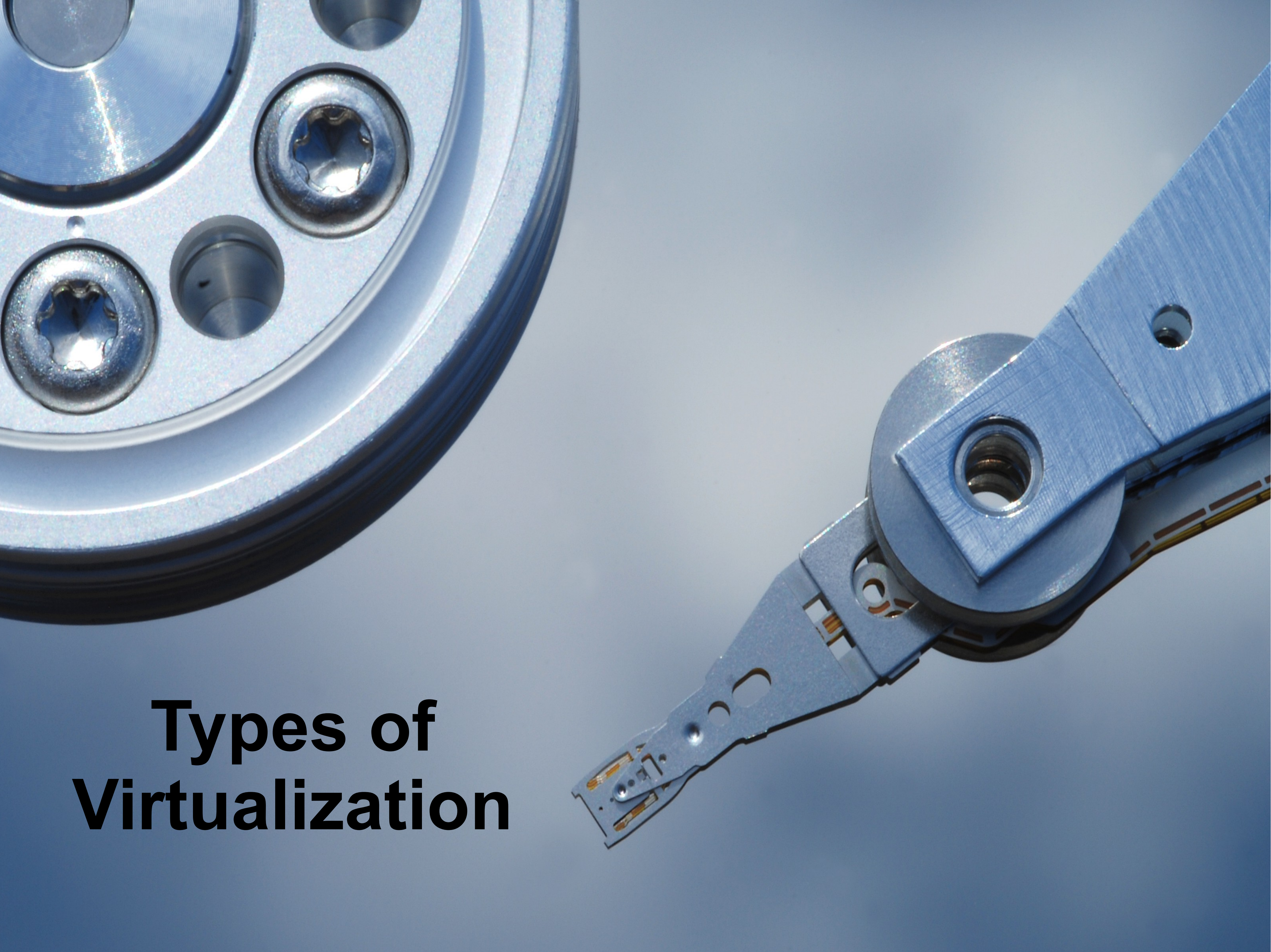
Quality Assurance

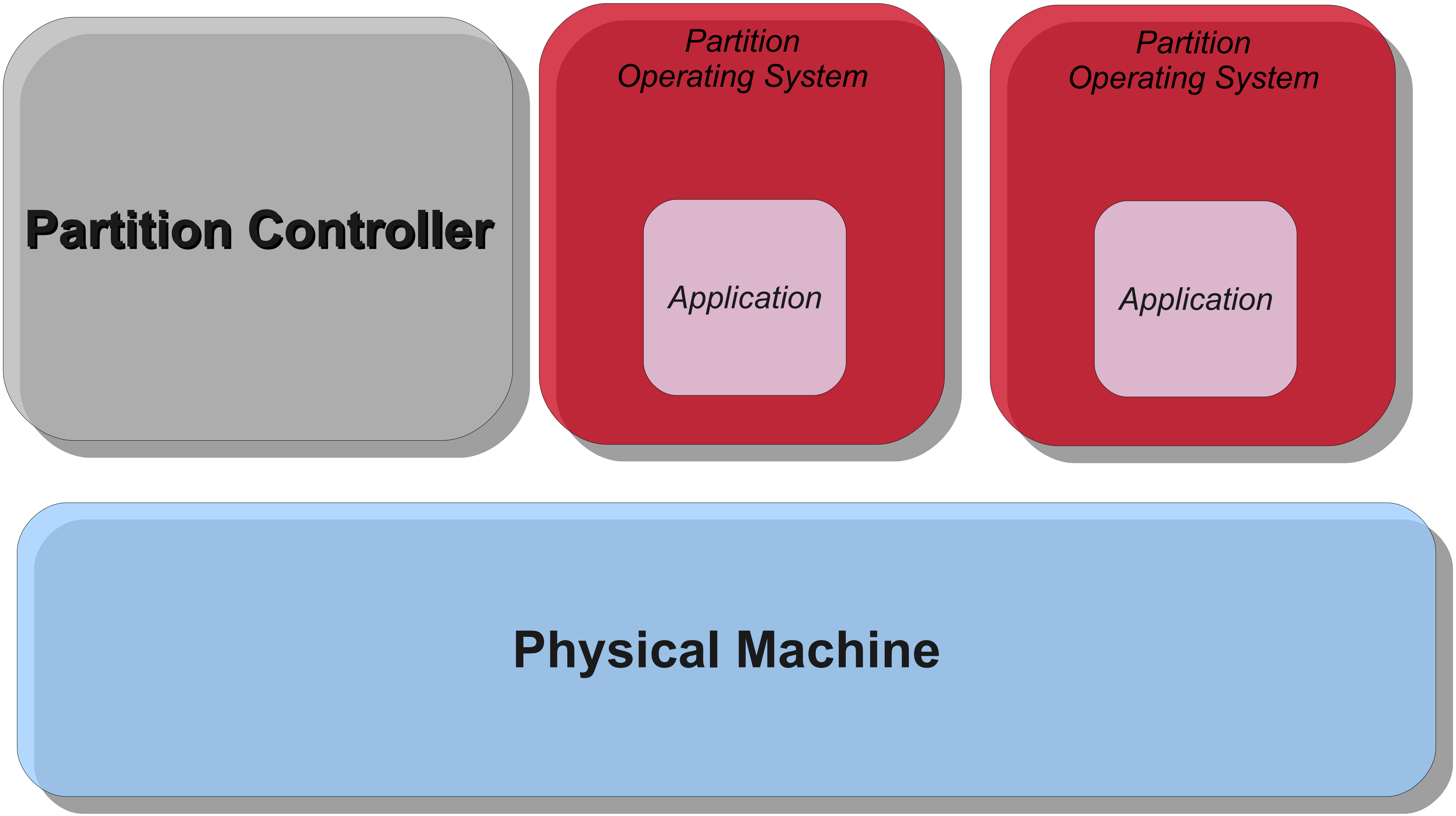

04972G072
1,000,000
100,000
10,000
1,000
PER REV
PER REV
PER REV
PER REV
CUBIC
FEET
Production

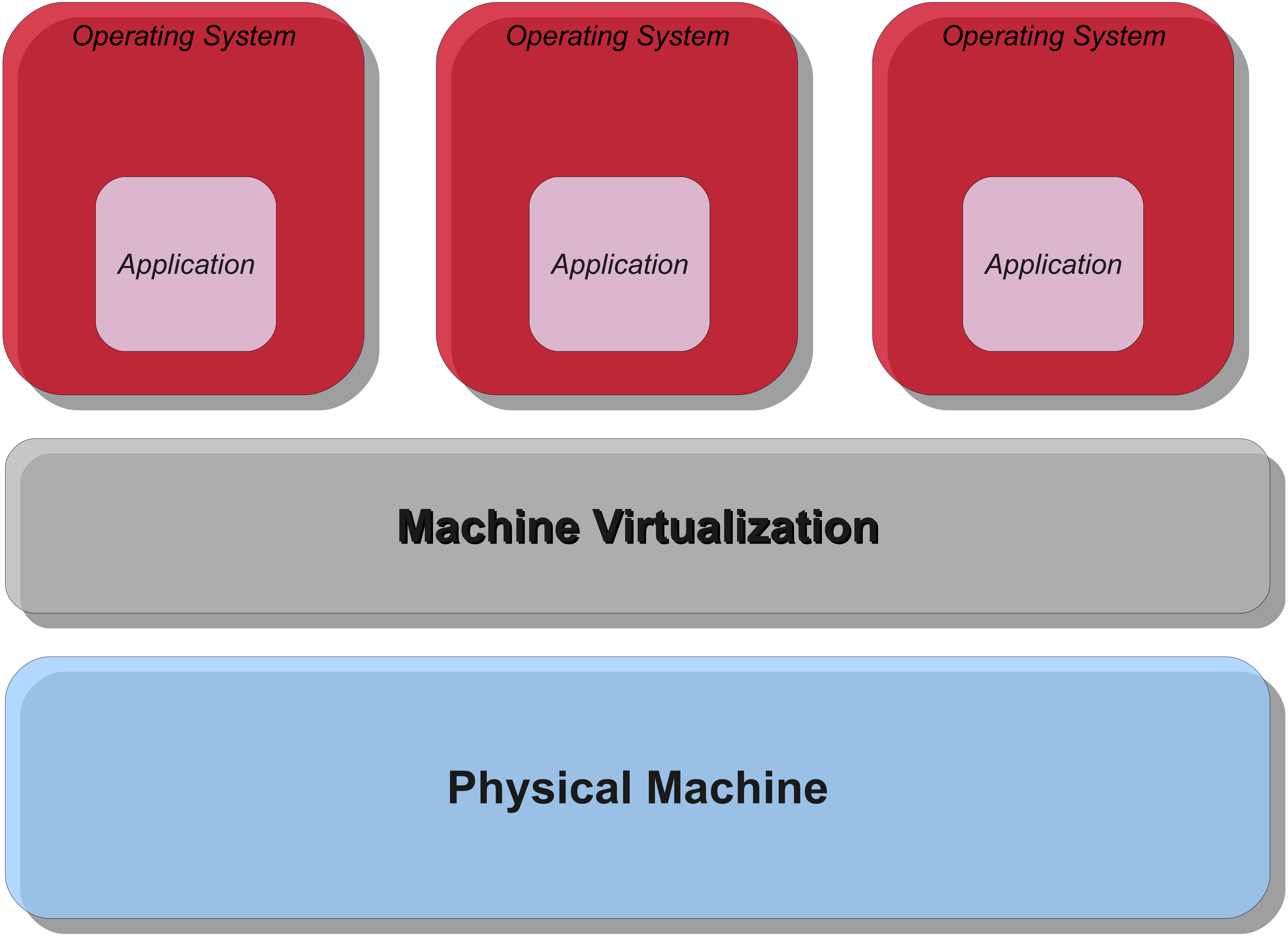

The End

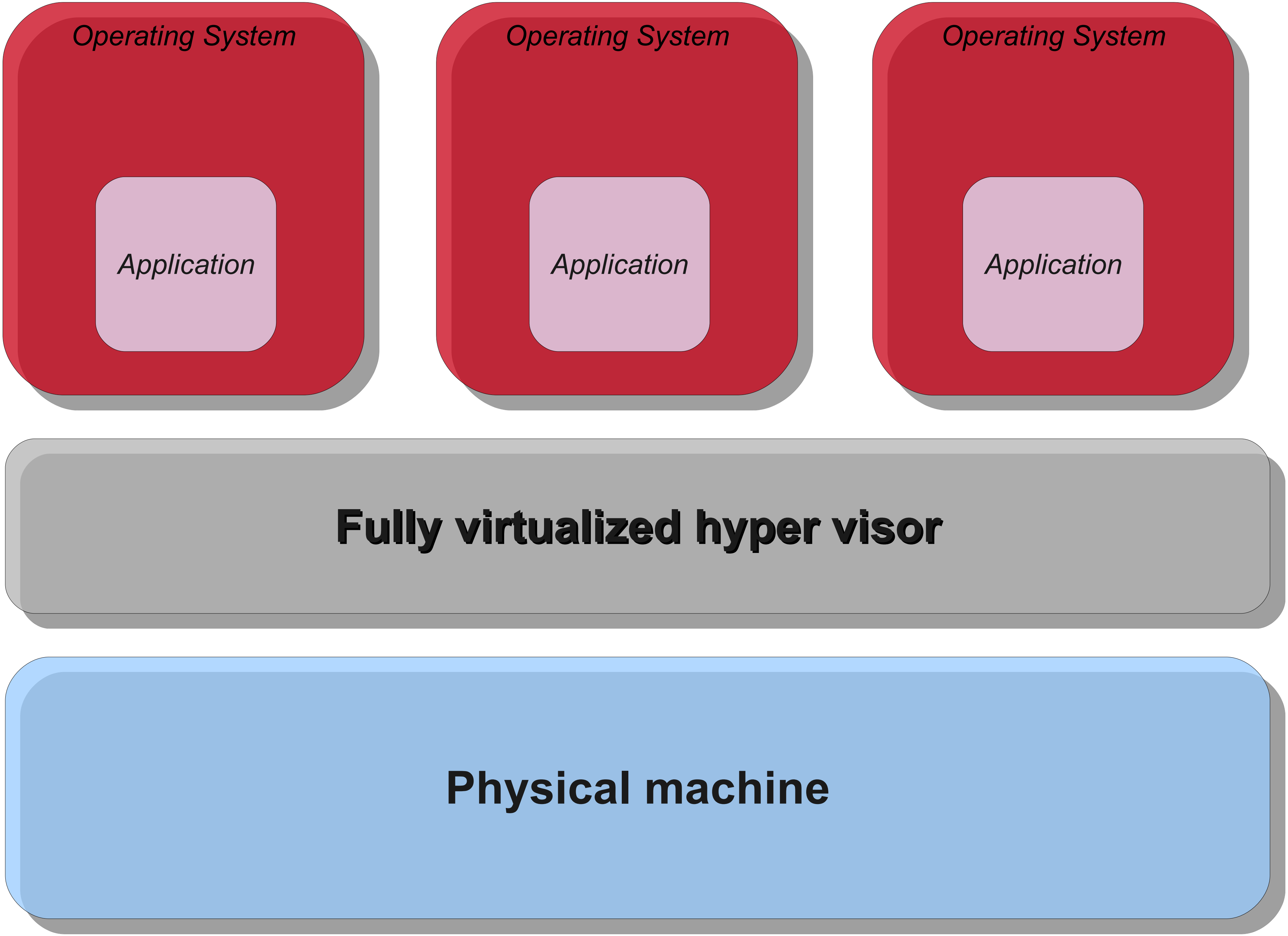
Backup

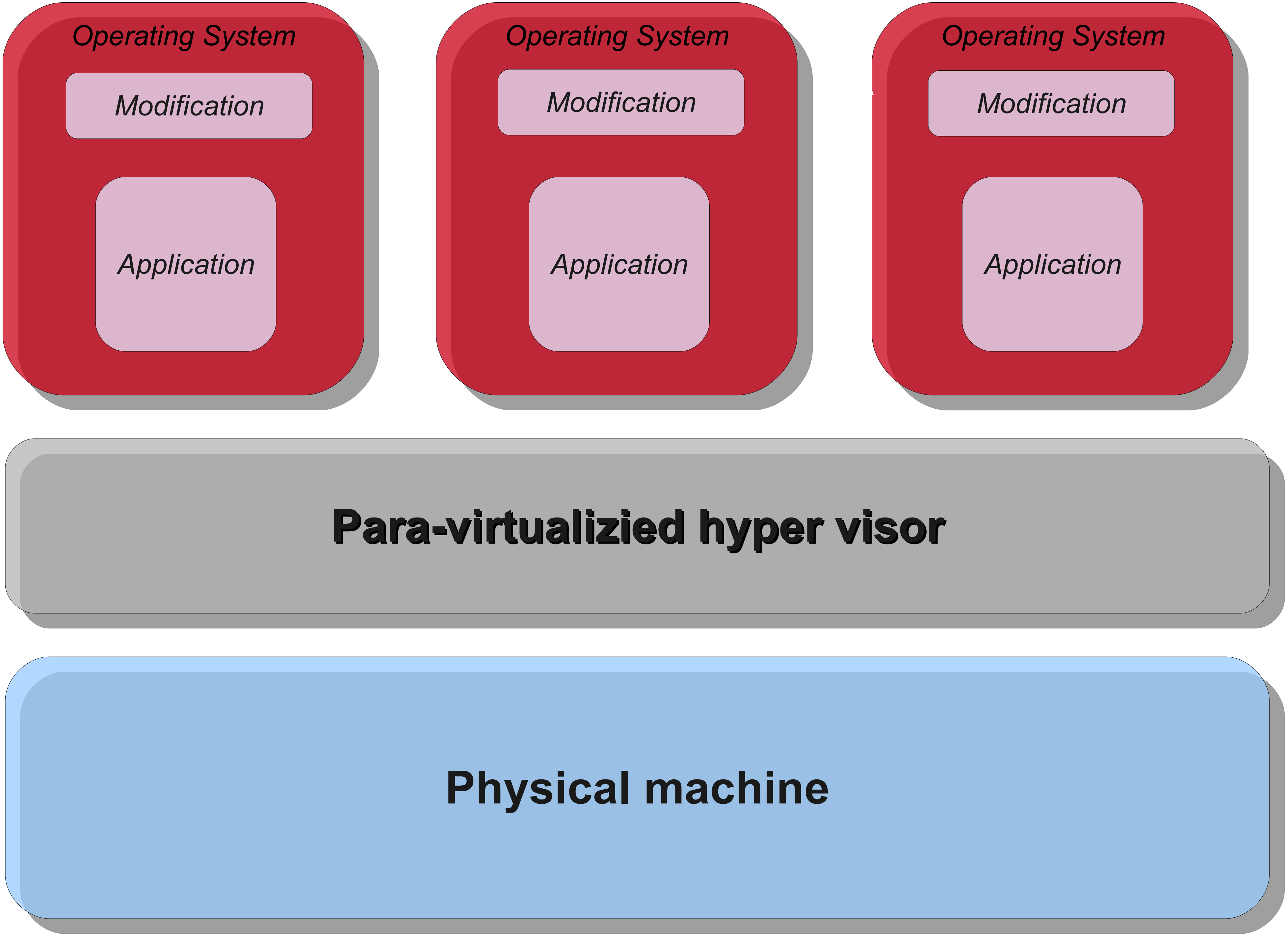
Types of Virtualization

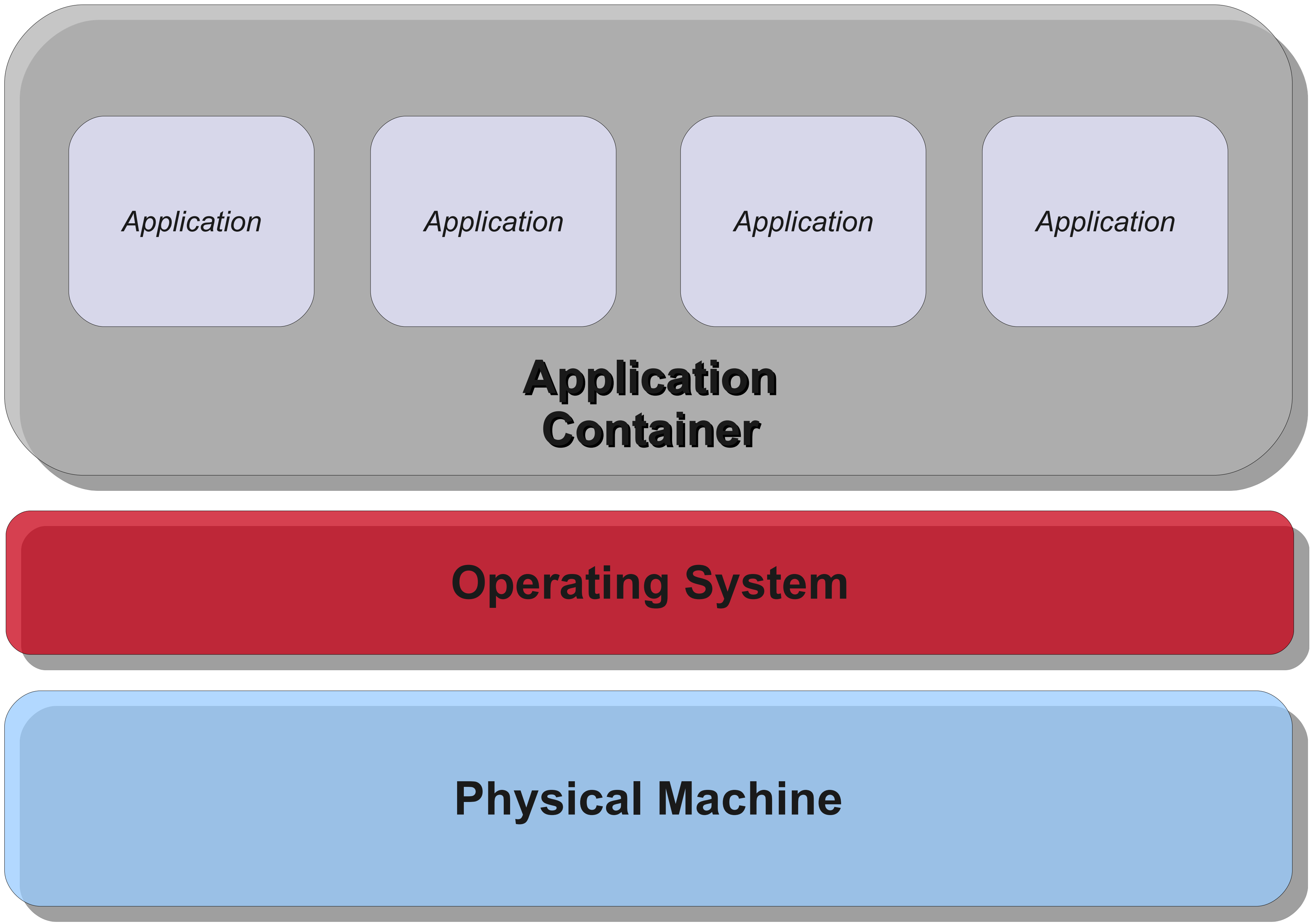


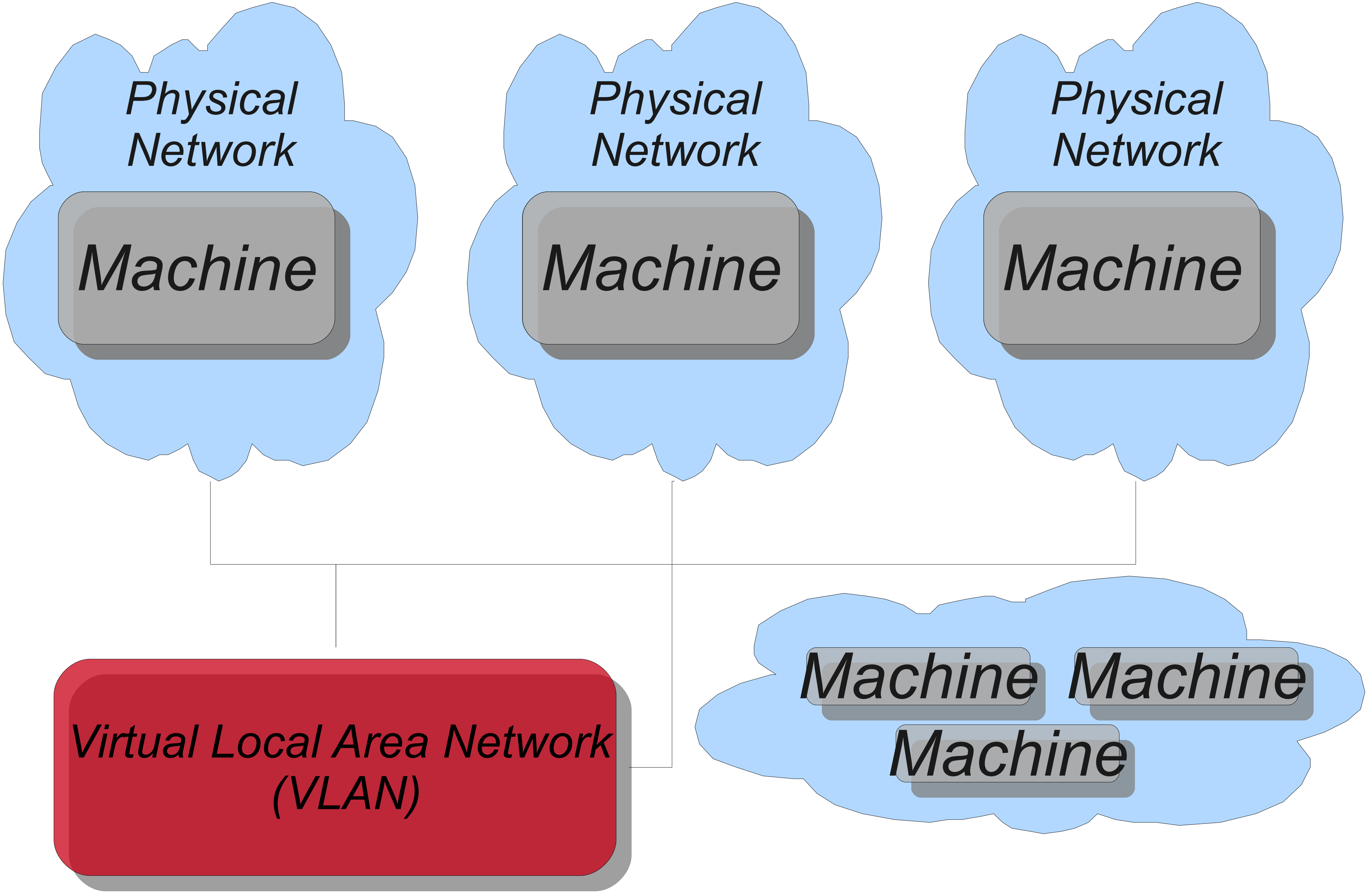

Partition Controller
Partition
Operating System
Application
Partition
Operating System
Application
Physical Machine

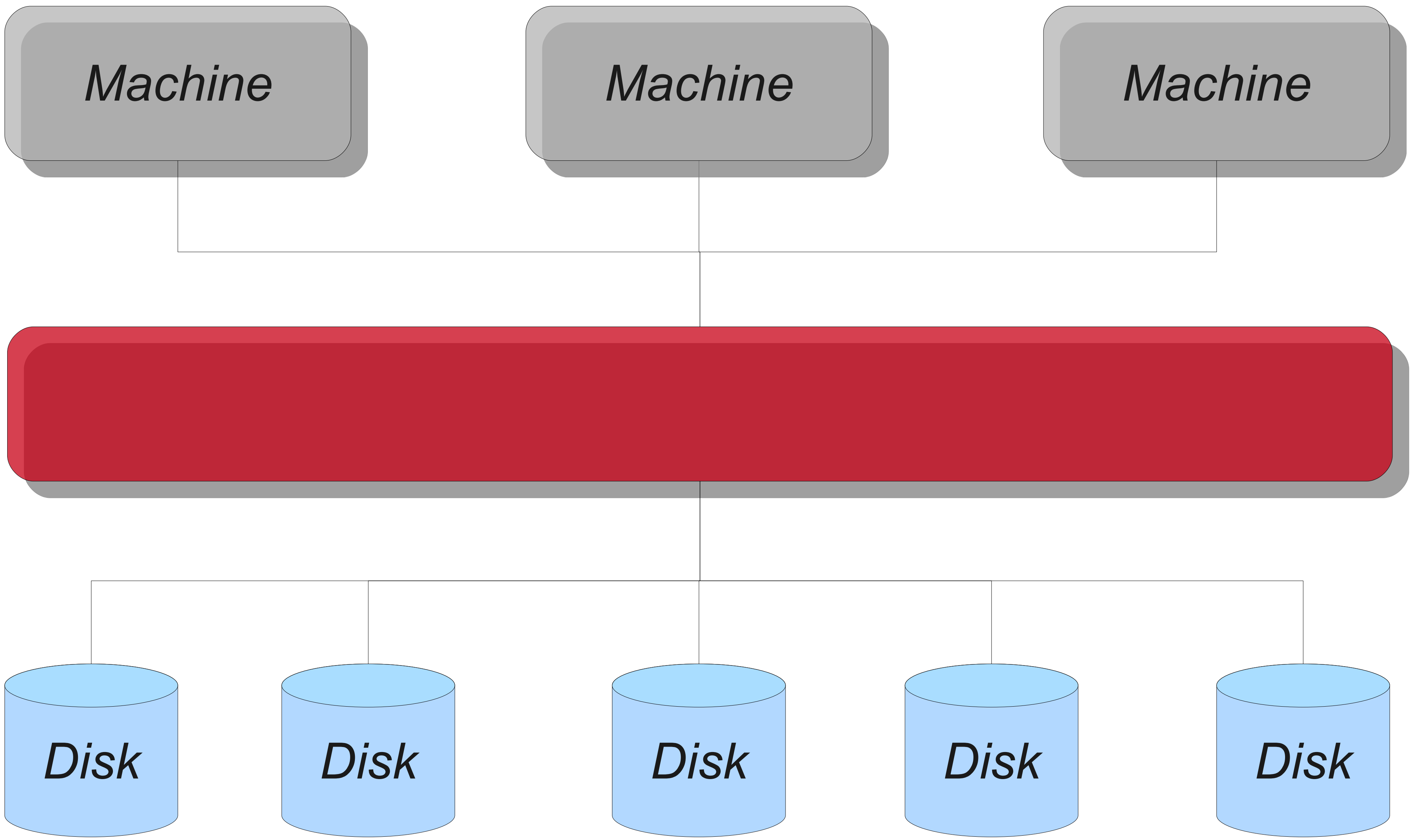

Operating System
Application
Operating System
Application
Operating System
Application
Machine Virtualization
Physical Machine


Operating System
Application
Operating System
Application
Operating System
Application
Fully virtualized hyper visor
Physical machine


Operating System
Modification
Application
Operating System
Modification
Application
Operating System
Modification
Application
Para-virtualizied hyper visor
Physical machine


Application
Application
Application
Application
Application Container
Operating System
Physical Machine


Physical
Network
Machine
Physical
Network
Machine
Physical
Network
Machine
Virtual Local Area Network
(VLAN)
Machine
Machine
Machine


Machine
Machine
Machine
Disk
Disk
Disk
Disk
Disk

Basic CPU Virtualization

- VMM runs in most privileged mode
  - VMM can maintain complete control
- Guest OS runs in an unprivileged mode
  - Privileged instructions will trap
  - VMM then emulates the required instruction in a safe manner

Basic Memory Virtualization

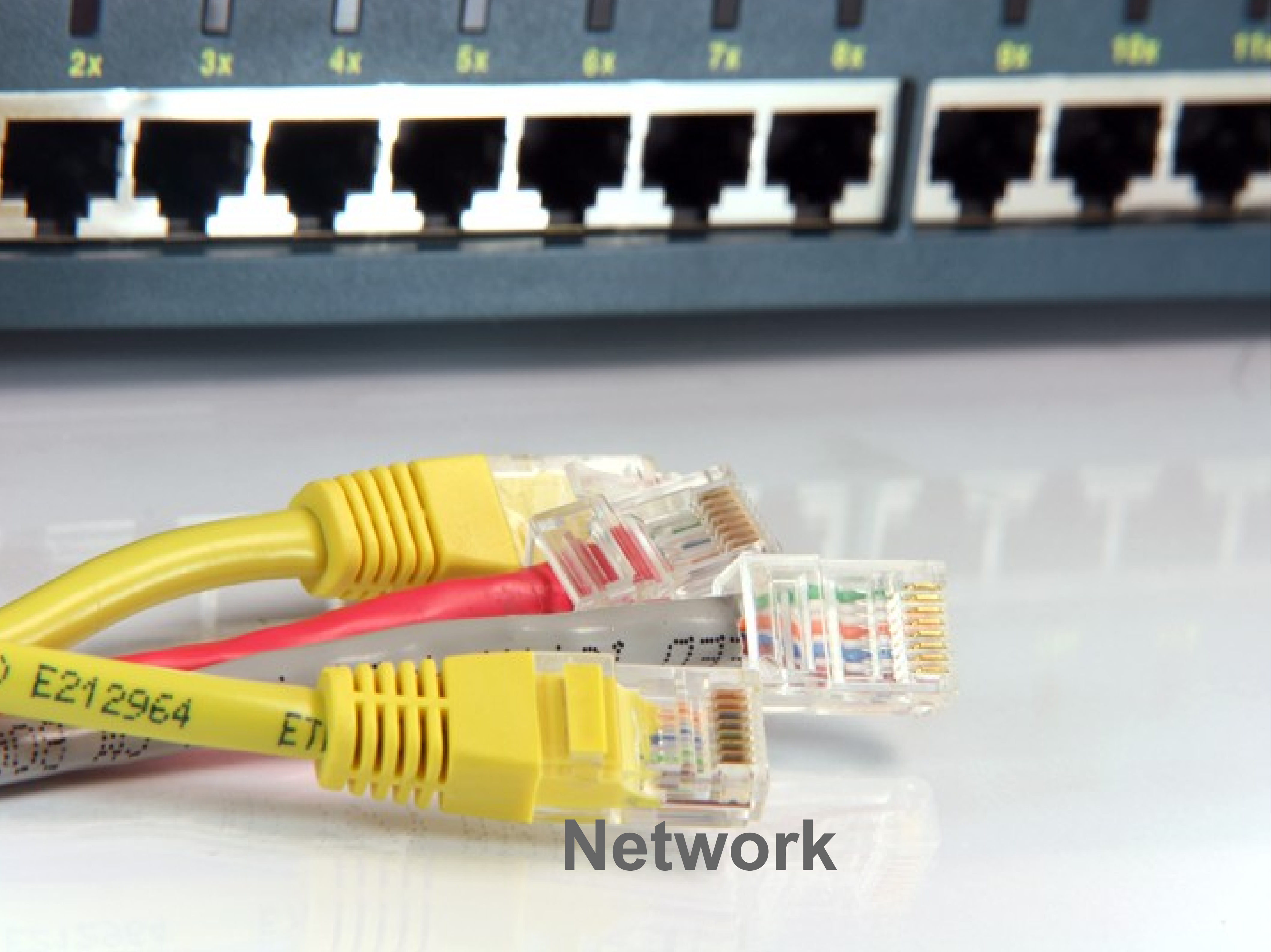
- VMM maintains a “shadow” page table
  - Guest OS establishes a mapping
  - VMM detects changes, updates shadow
  - Hardware uses shadow page table
- VMM can over commit memory
  - Just like normal virtual memory

Why “Basic” Doesn’t Work

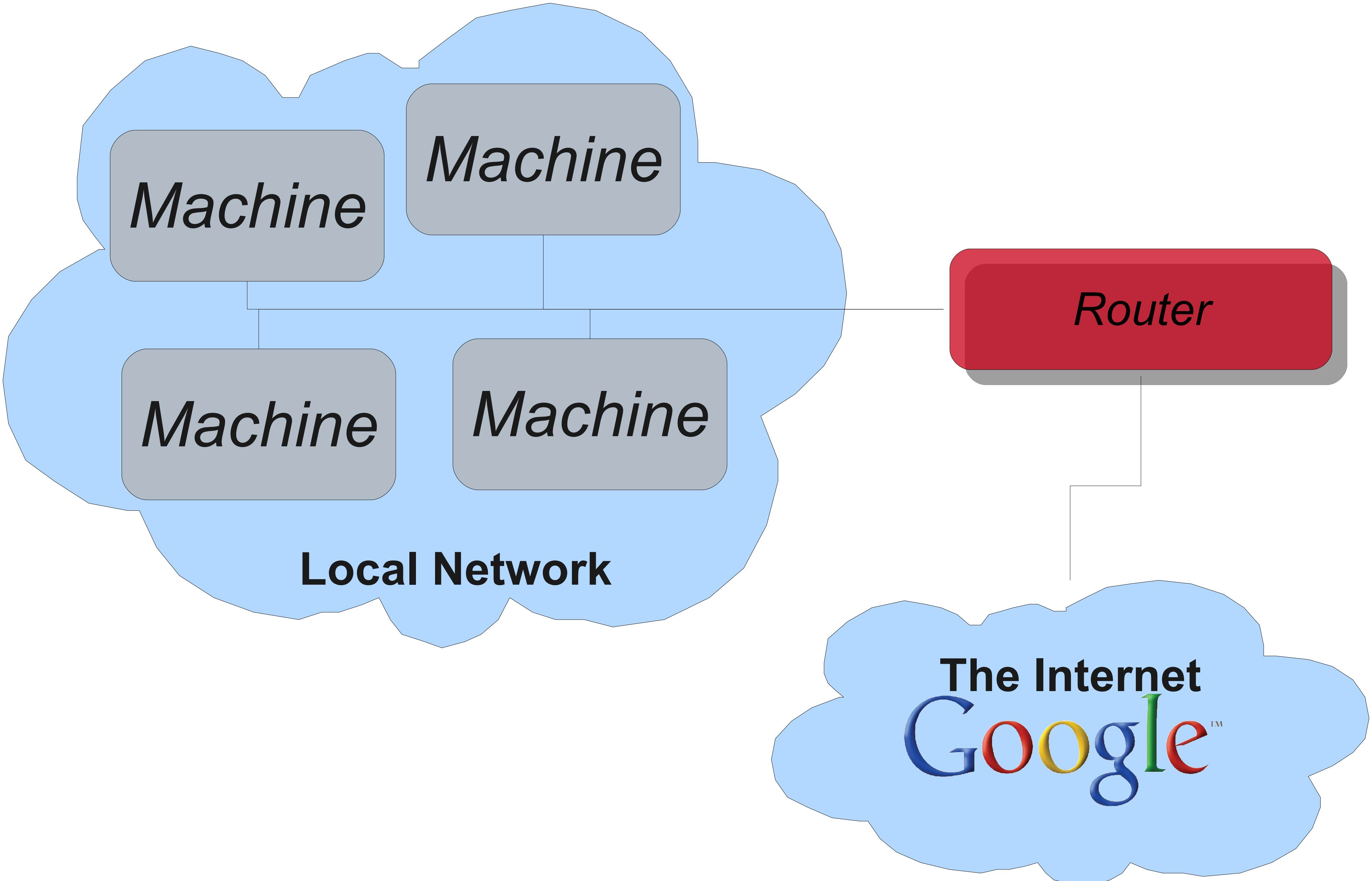
- Architectures not designed for virtualization
  - Unprivileged privileged instructions
- Performance implications
  - Traps are slow
- Wasted resources from redundant code
- Lack of information leads to ineffectiveness

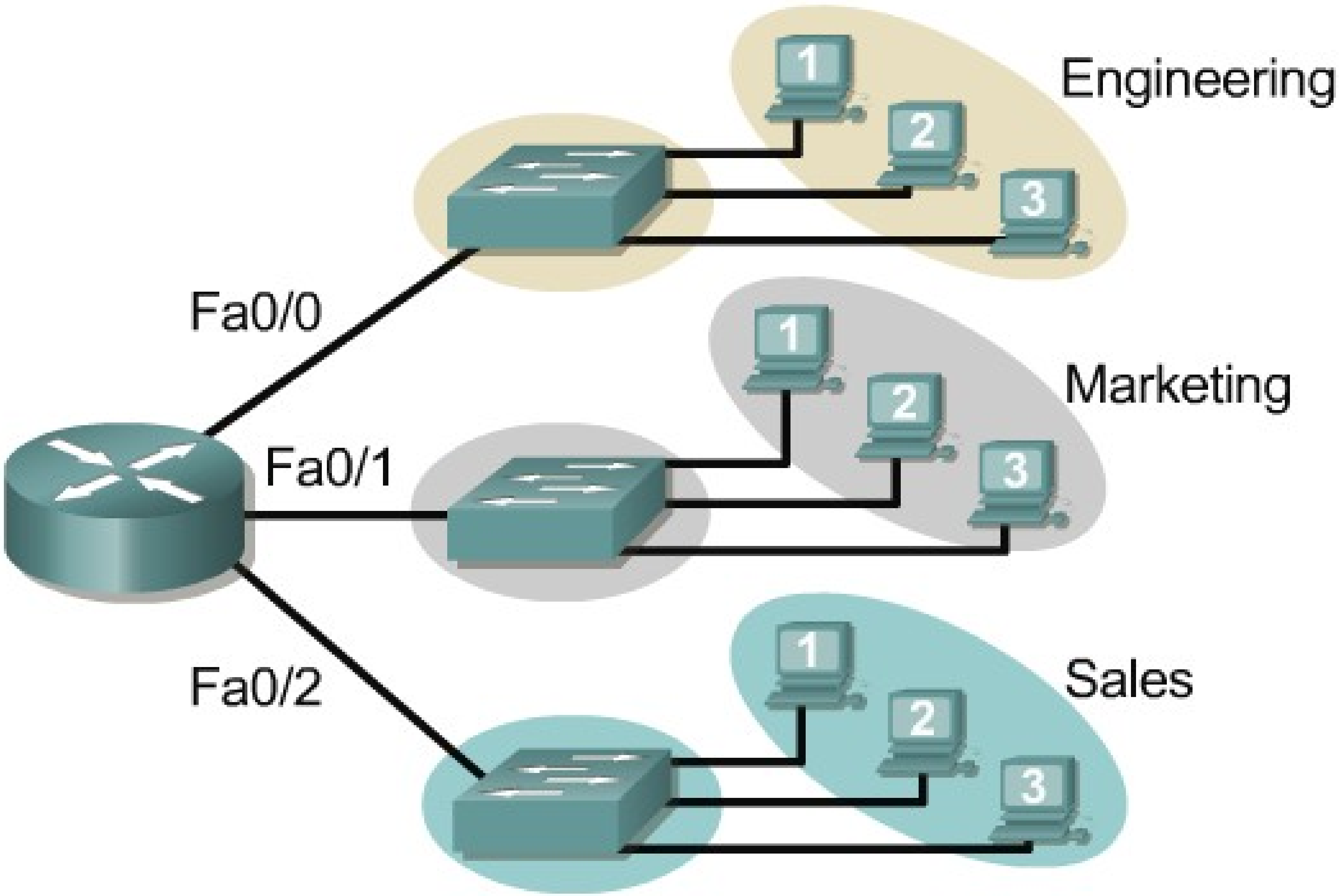
Benefits and Drawbacks

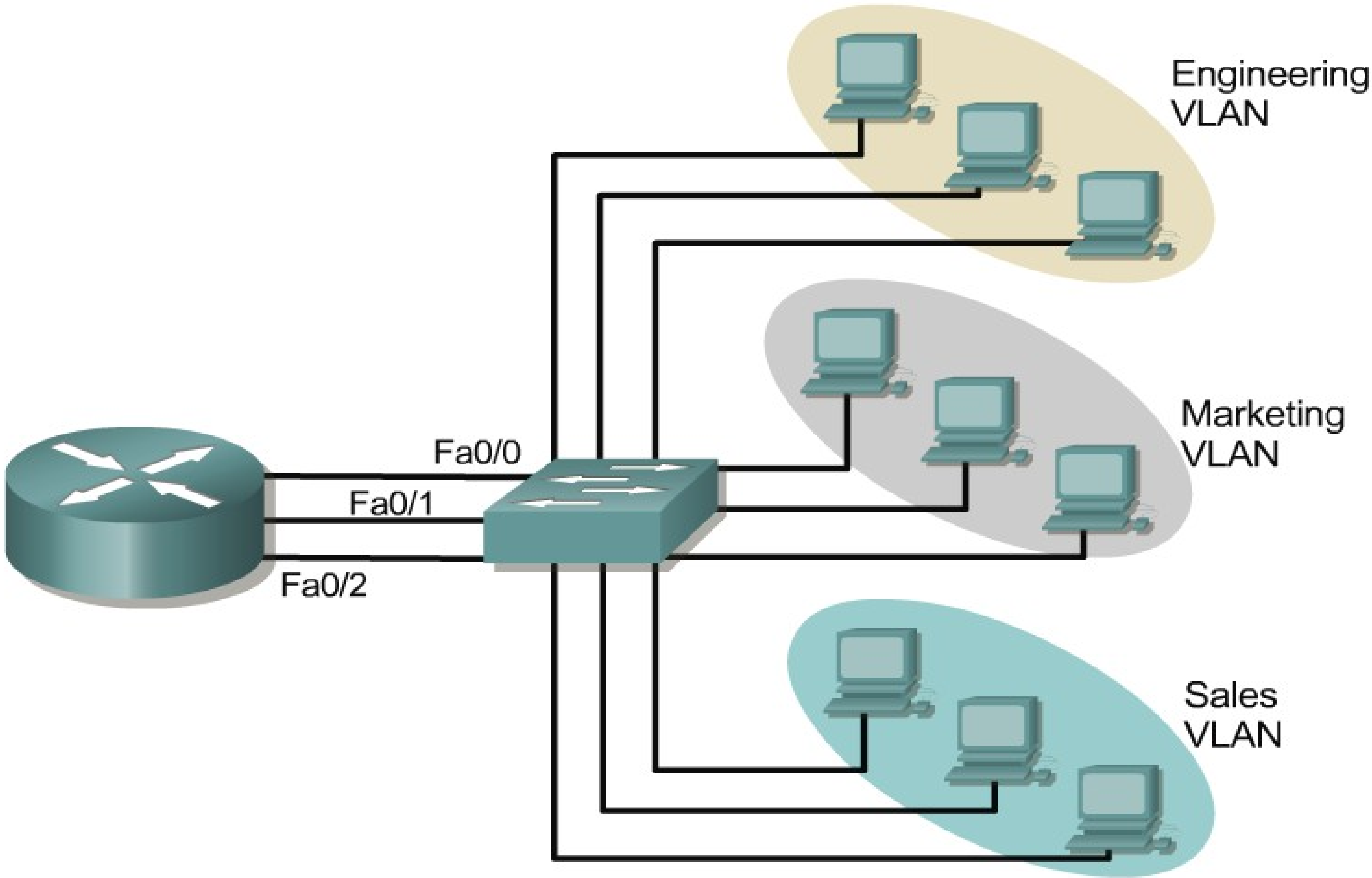
- Unmodified applications and operating systems can run on the VMM
- Performance can suffer because of the need to emulate protected operations
    - Especially bad on the IA-32
    - Virtual Memory Especially Difficult
    - Special tricks can be employed


2x
3x
4x
5x
E212964
Network

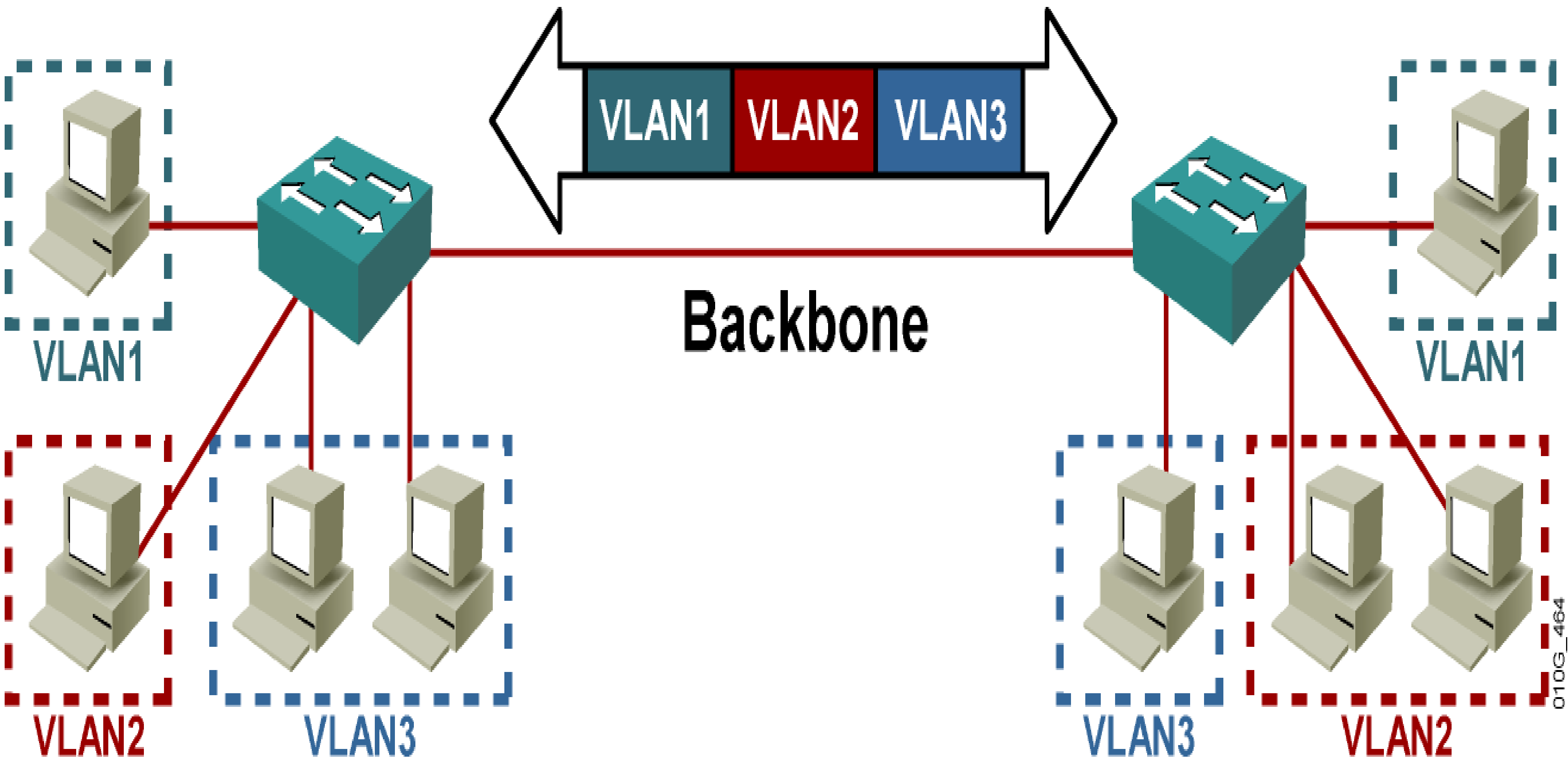
Network Address Translation



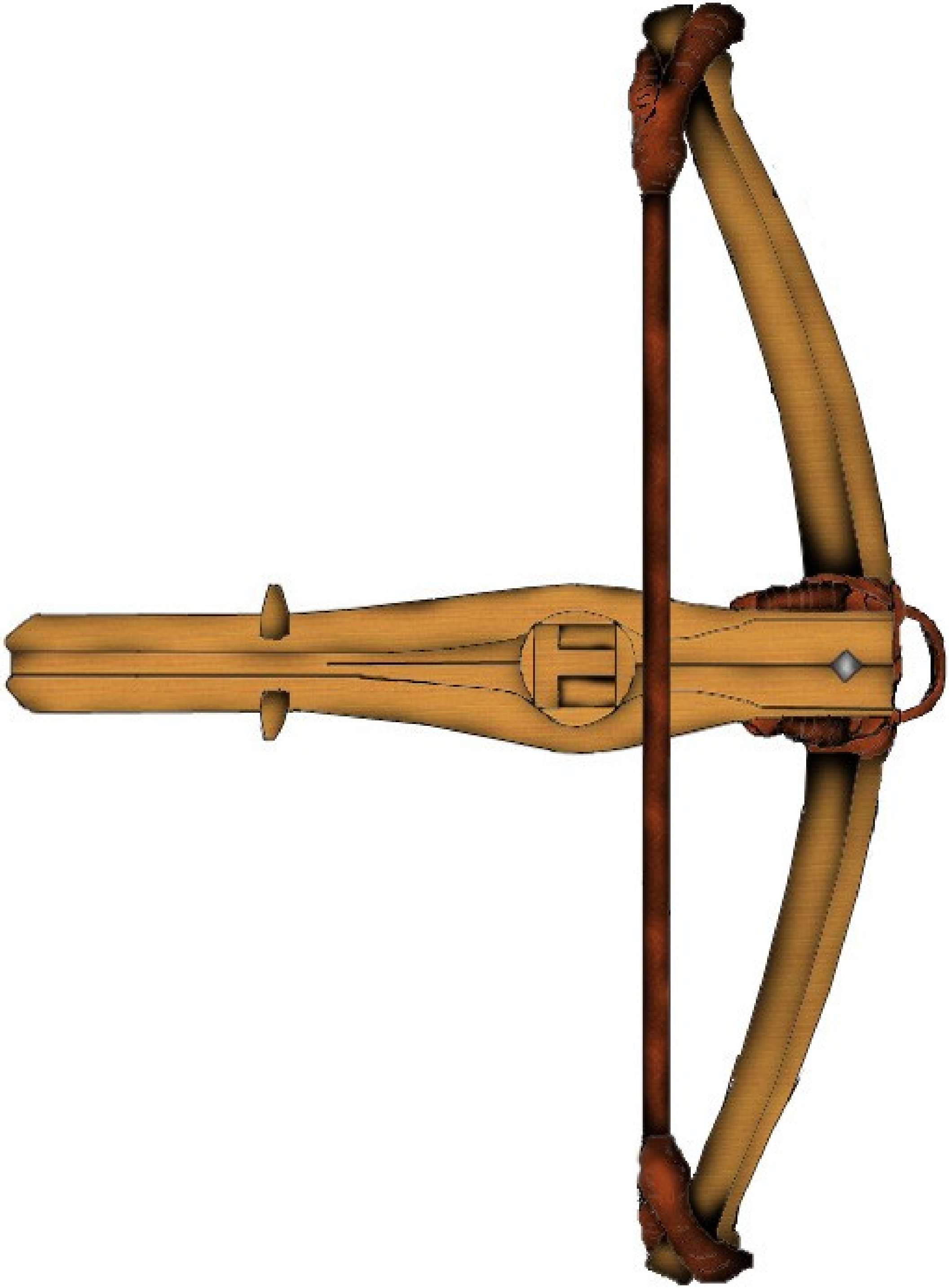

Engineering
Fa0/0
Marketing
Fa0/1
Sales
Fa0/2
1
2
3
1
2
3
1
2
3


Engineering
VLAN
Marketing
VLAN
Sales
VLAN
Fa0/0
Fa0/1
Fa0/2

VLAN Trunking/Tagging

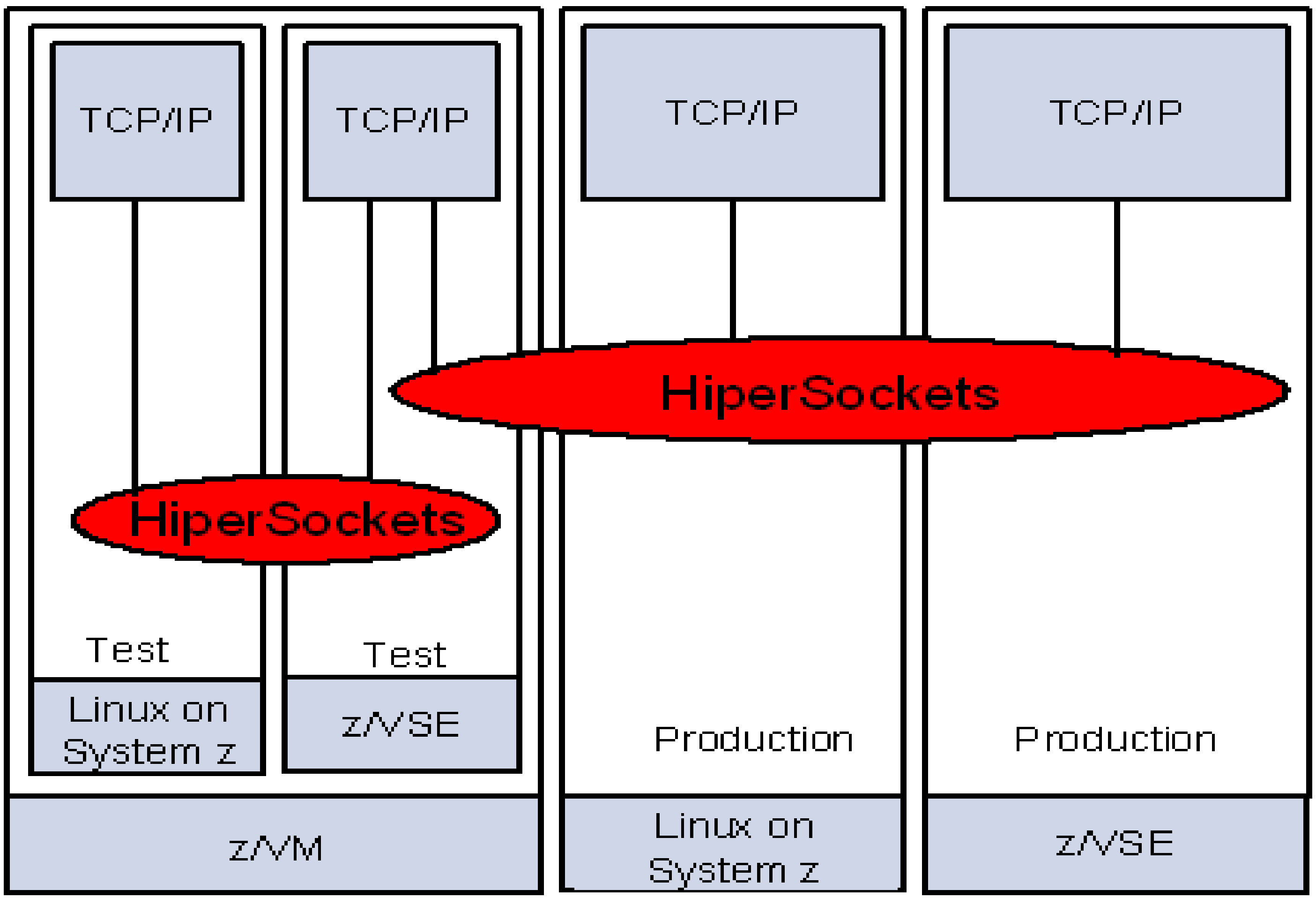


Virtual Private Network



- Sun Crossbow

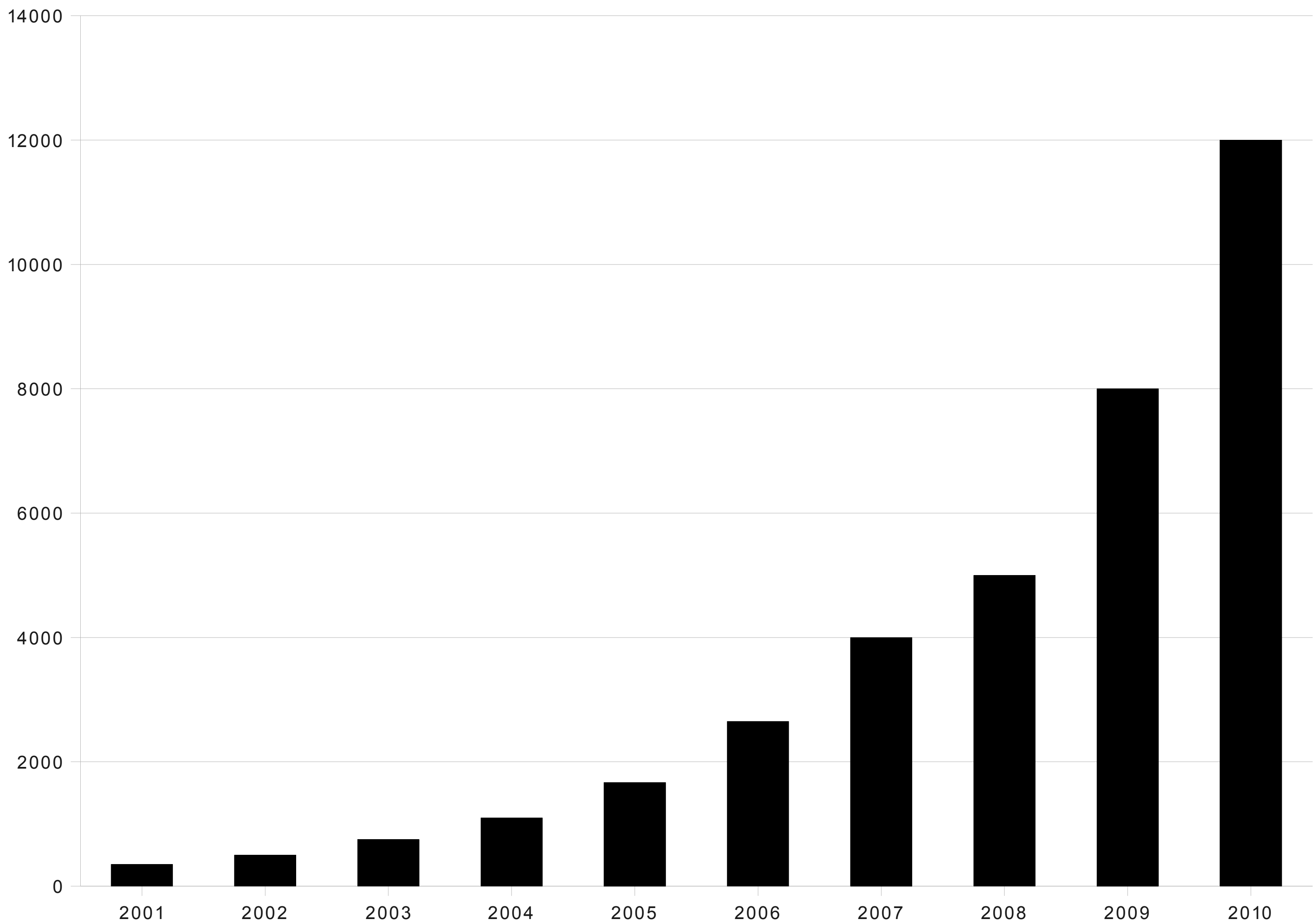
HiperSockets



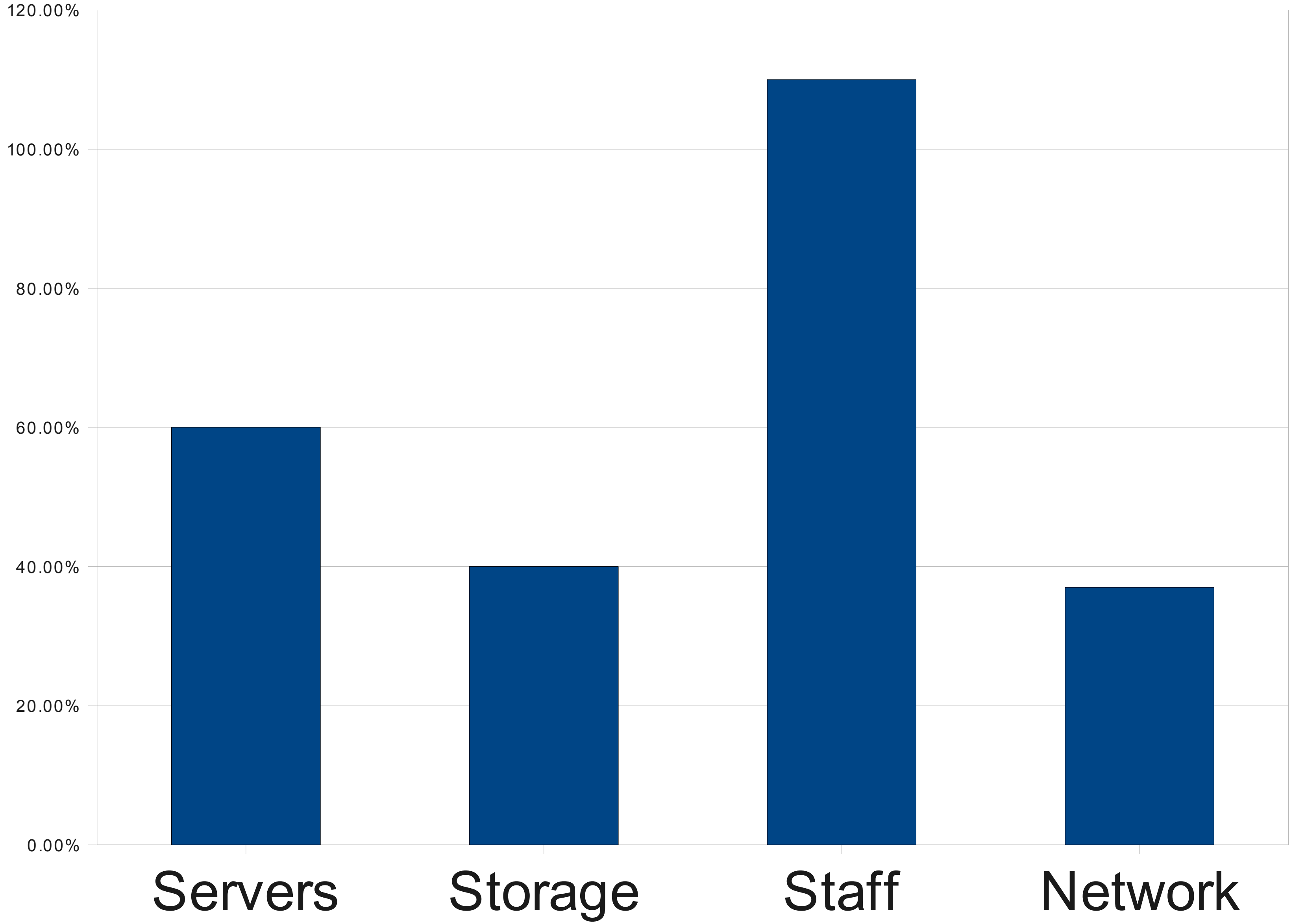
Channel Bonding


Storage

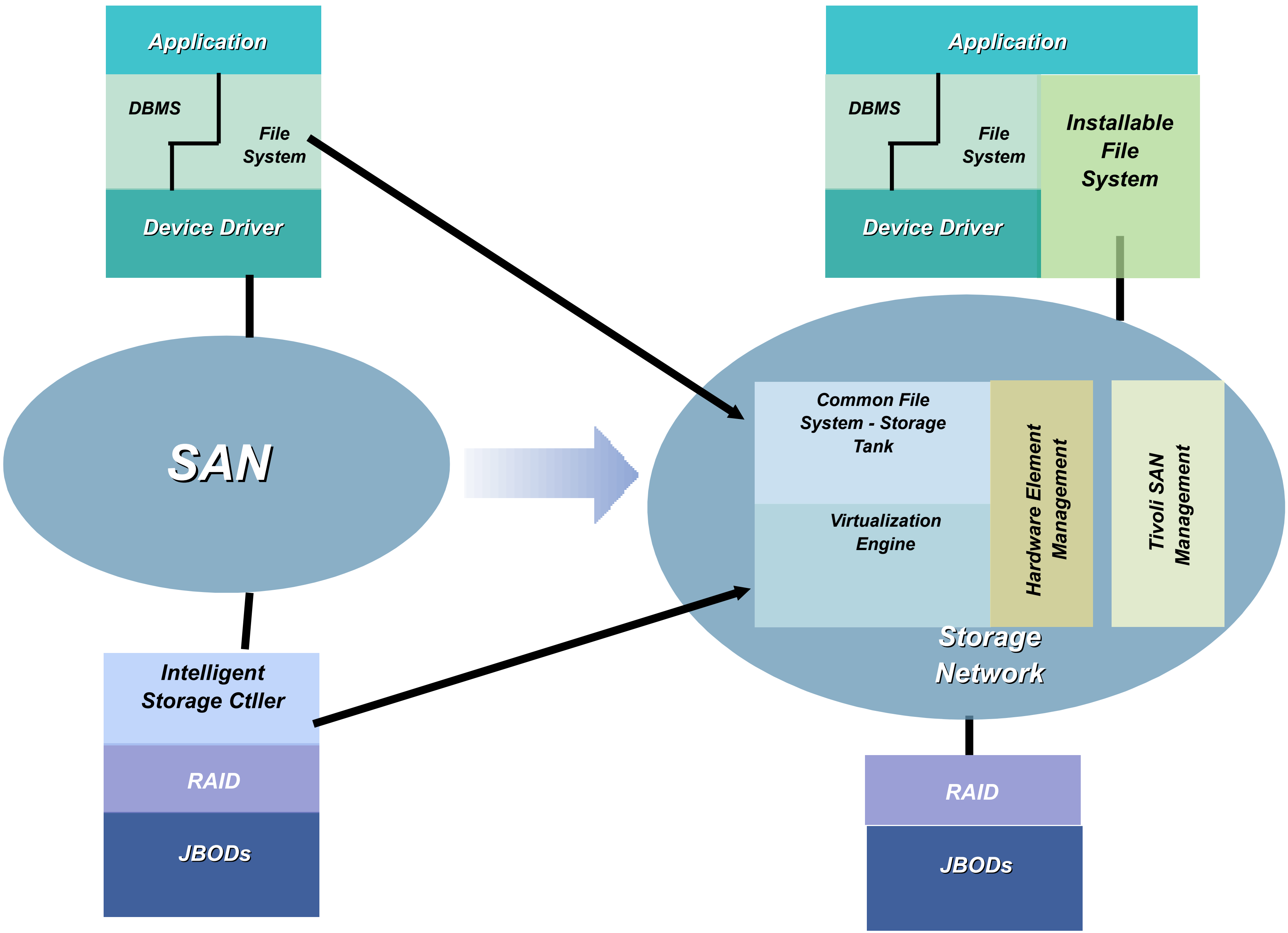
External Terabytes Worldwide



Utilization

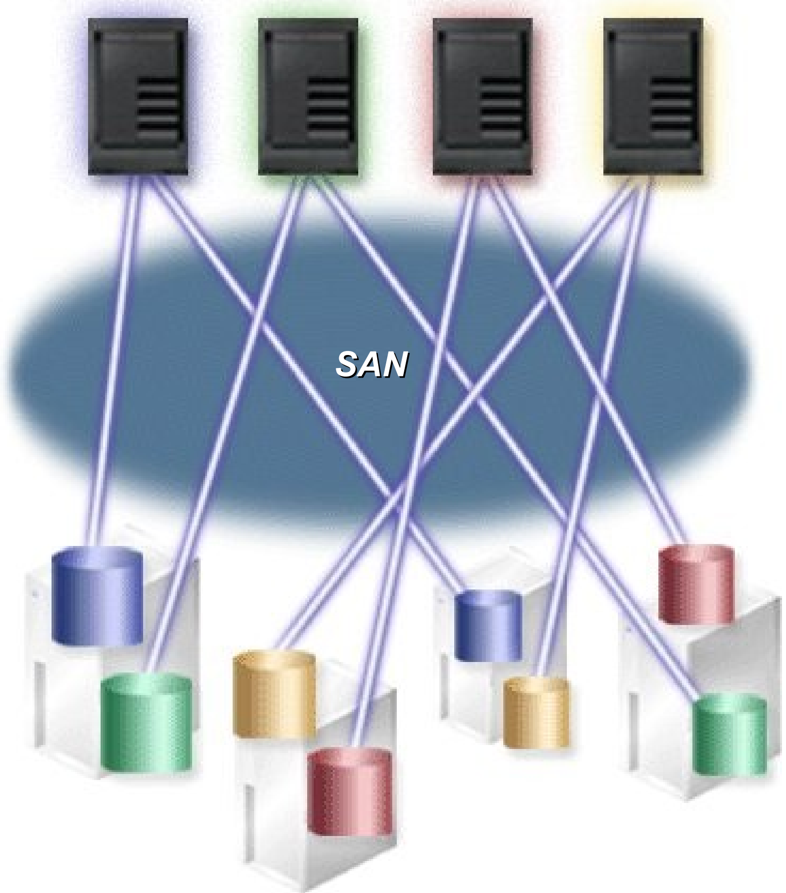


IBM Storage


Application
DBMS
File System
Device Driver
SAN
Intelligent Storage Ctller
RAID
JBODs
Application
DBMS
File System
Installable File System
Device Driver
Common File System - Storage Tank
Virtualization Engine
Hardware Element Management
Tivoli SAN Management
Storage Network
RAID
JBODs

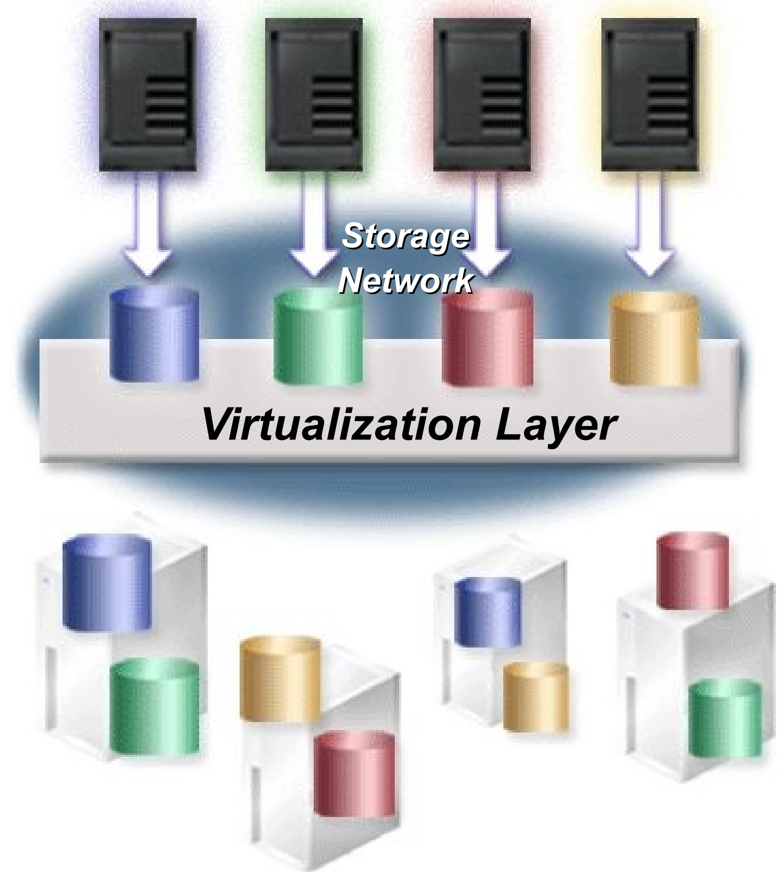
Block Level Virtualization

SANs Today



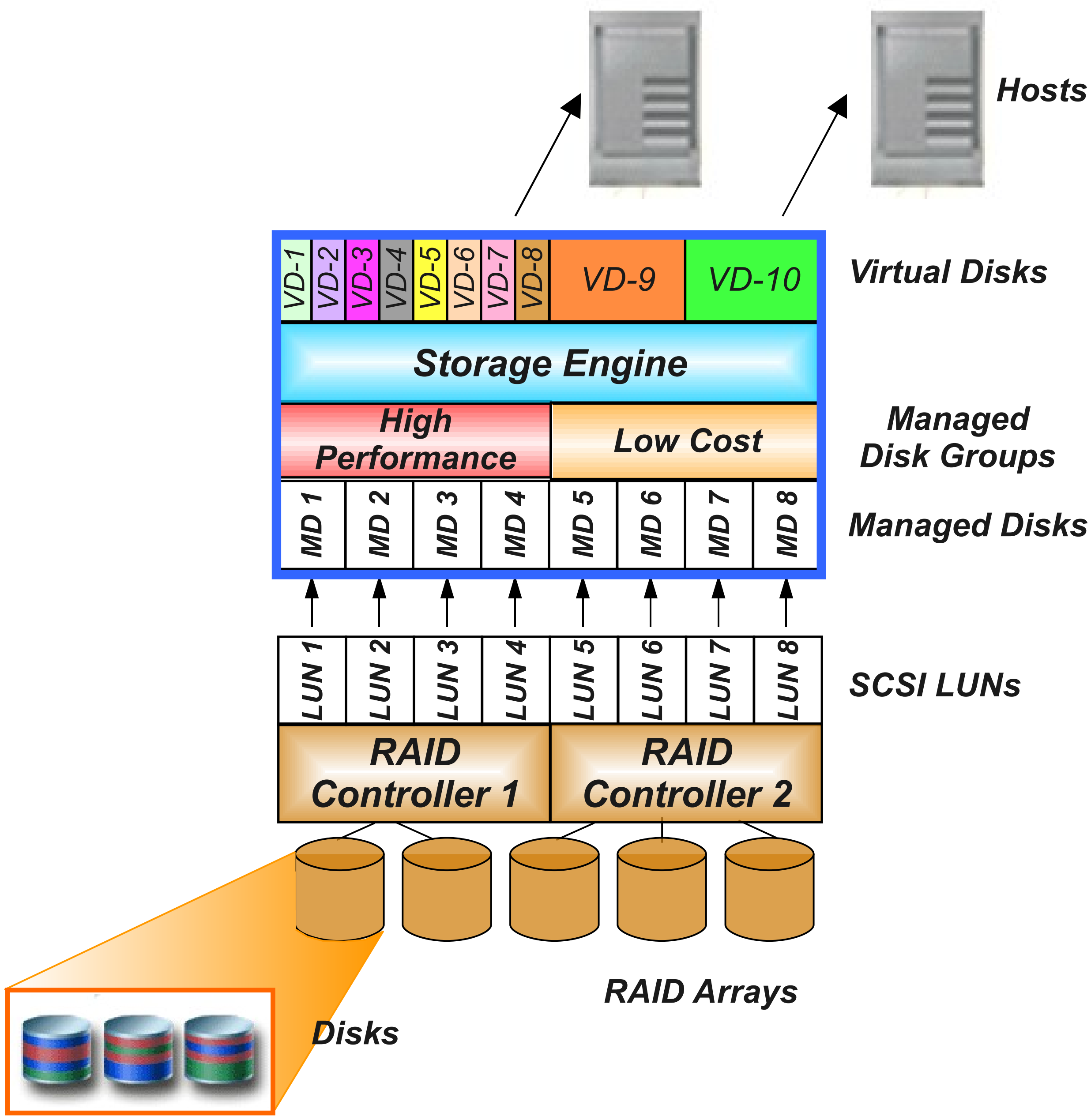
Servers are mapped to specific physical disks i.e., "physical mapping"

Block Virtualization

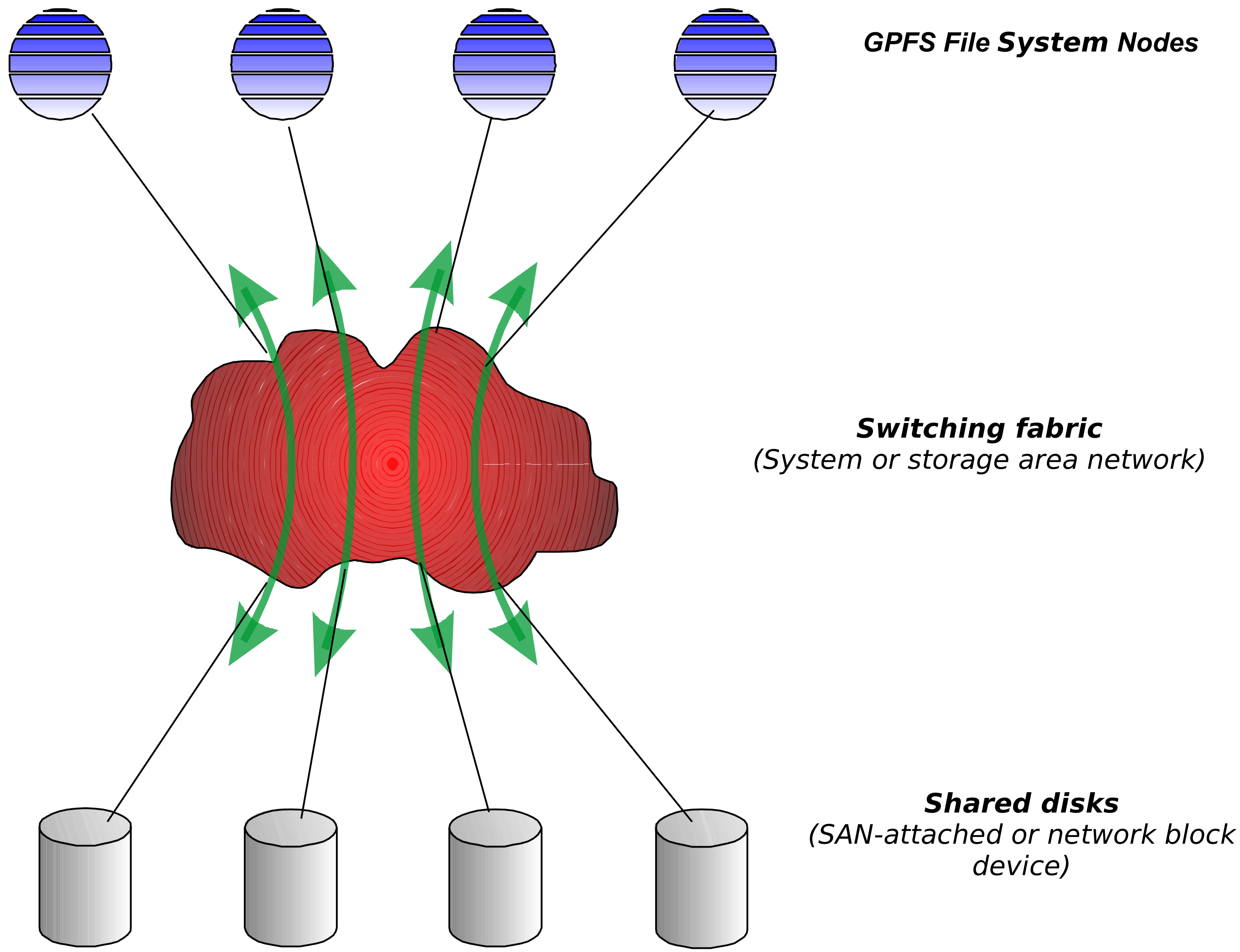


Servers are mapped to a virtual disk i.e., "logical mapping"

SAN Volume Controller

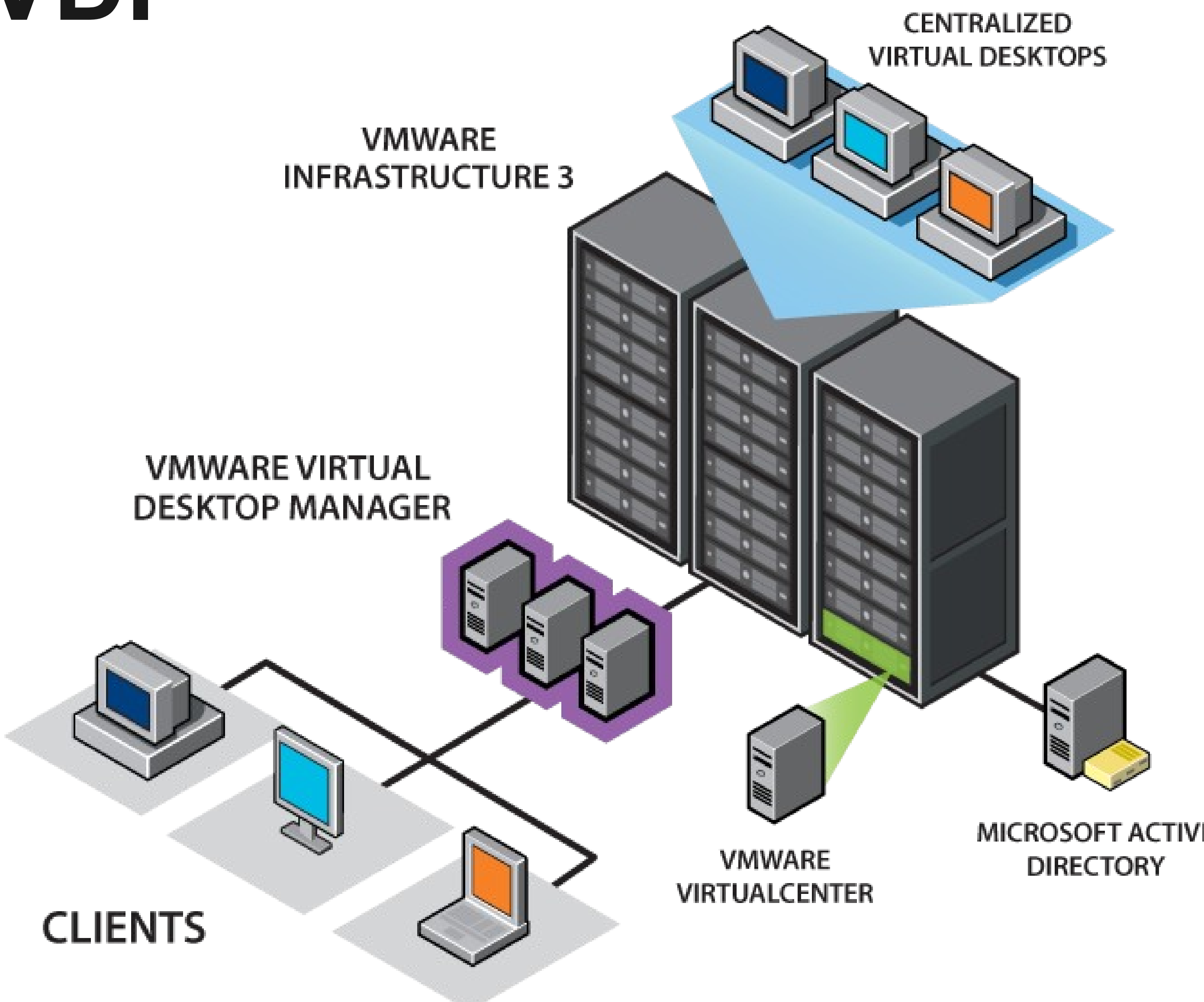


GPFS

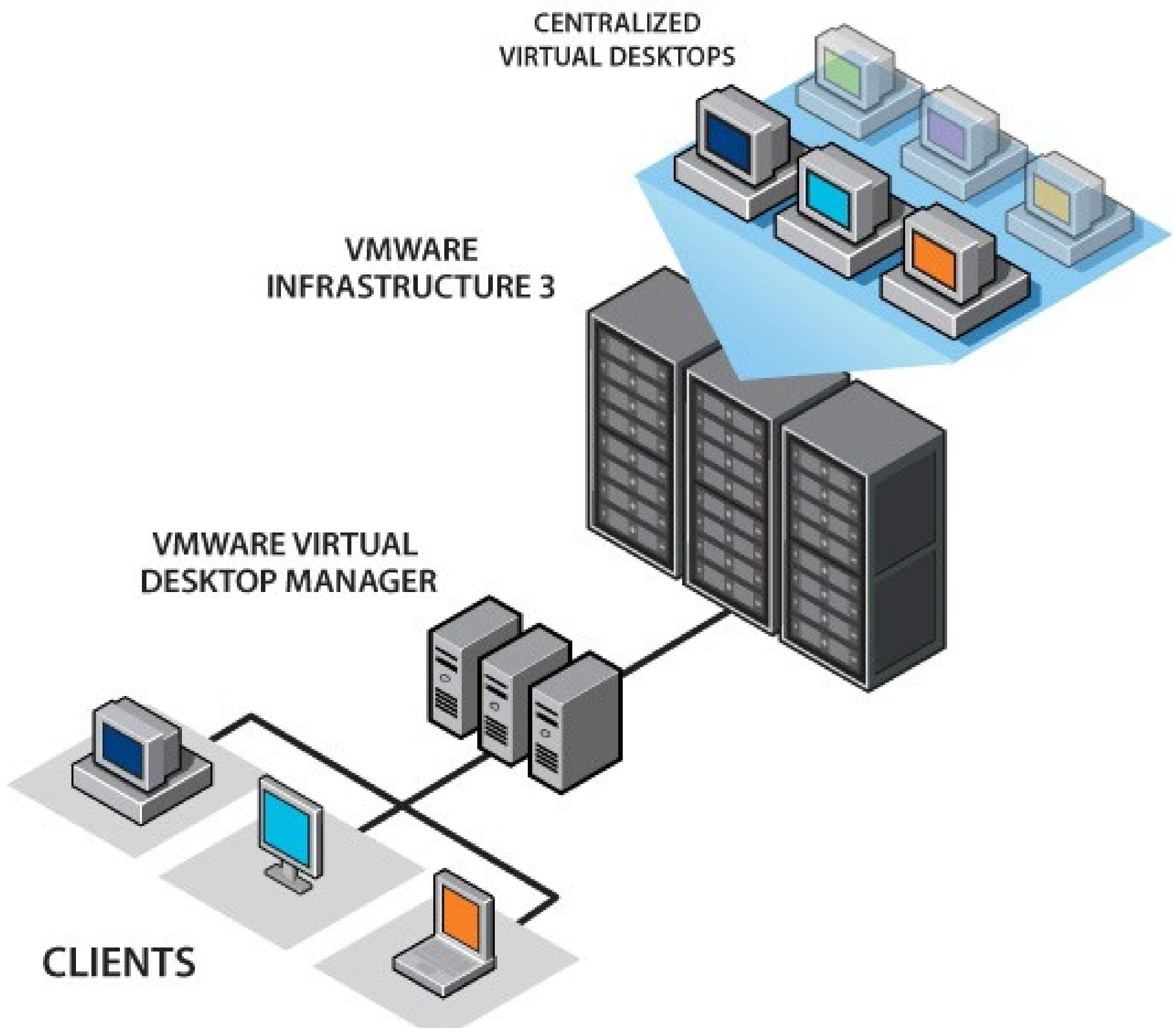

GPFS File System Nodes
Switching fabric
(System or storage area network)
Shared disks
(SAN-attached or network block device)

Desktop

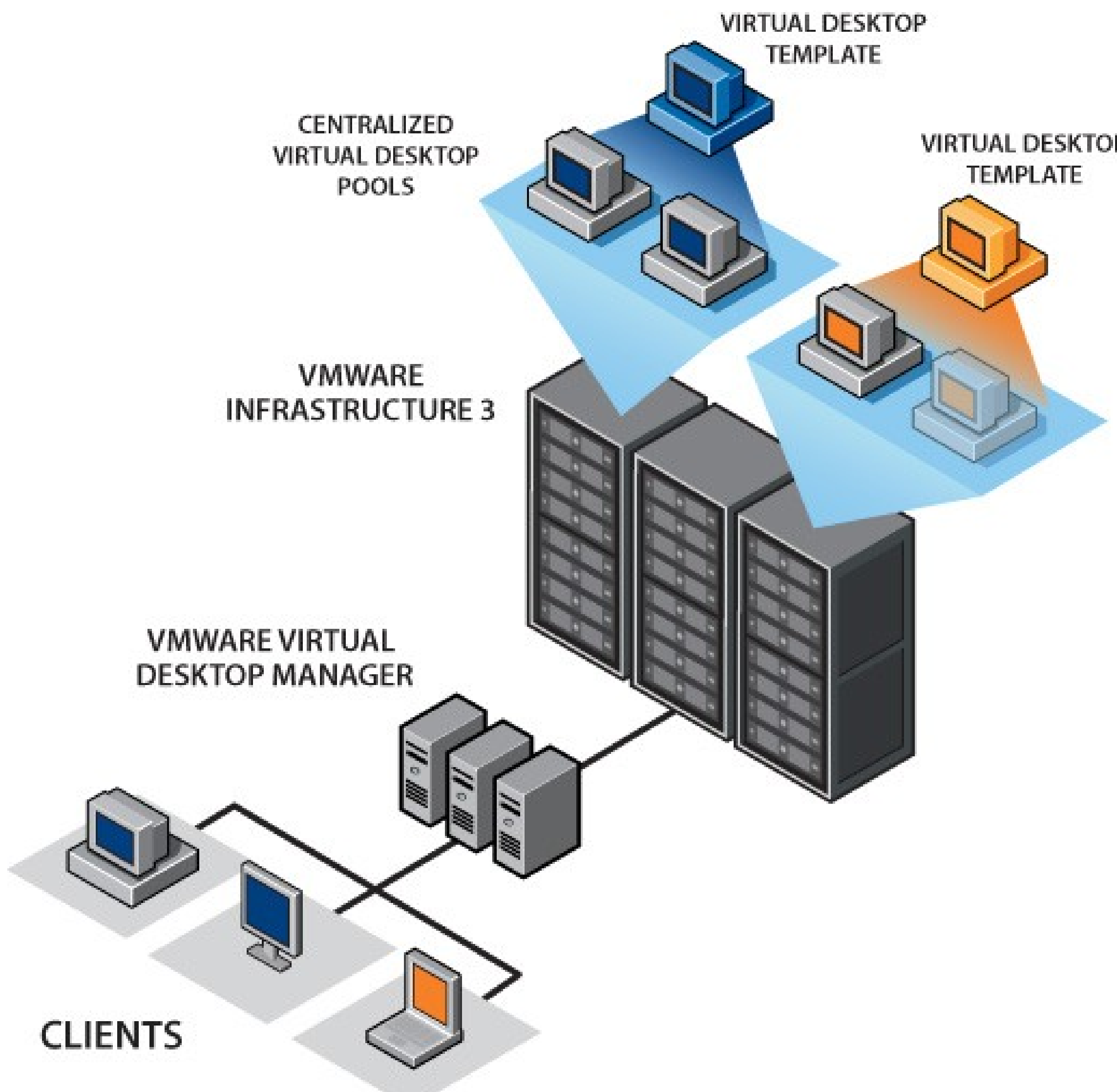
VDI


CENTRALIZED
VIRTUAL DESKTOPS
VMWARE
INFRASTRUCTURE 3
VMWARE VIRTUAL
DESKTOP MANAGER
CLIENTS
VMWARE
VIRTUALCENTER
MICROSOFT ACTIVE
DIRECTORY

Individual Desktops



Connection Broker: Deploying Desktops



How Customers Use VDI

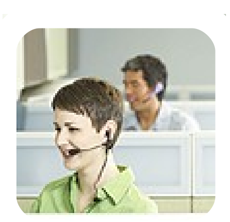
Centralize or Desktop Replacement

Replace traditional PCs with centralized virtual desktops for better control and efficient management. End users have flexibility

---

Disaster Recovery & Business Continuity

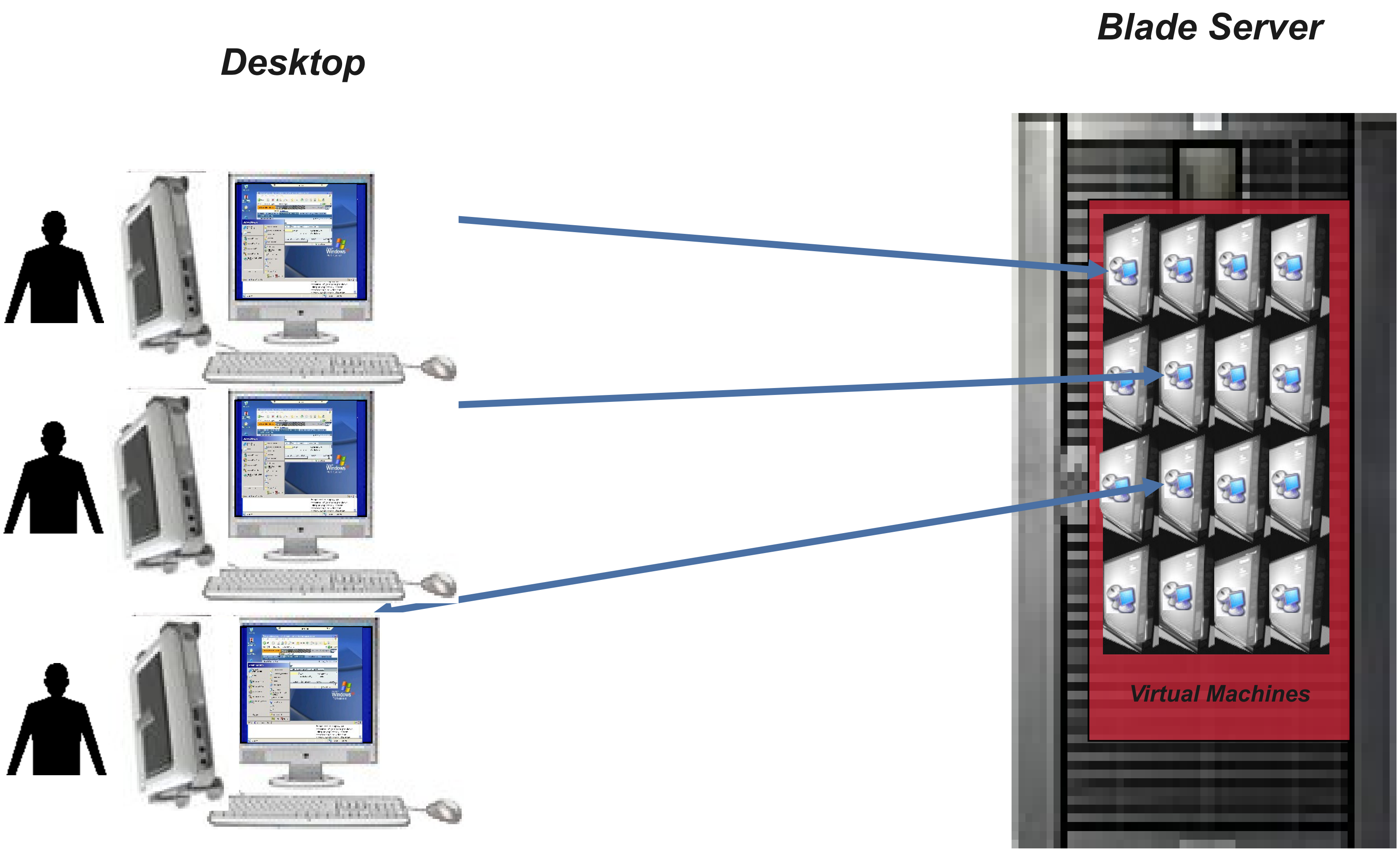
Provide continuous availability of desktops to end users by making high availability and disaster recovery solutions more cost-effective, simpler, and more reliable.



Transactional Office Workers or Developers

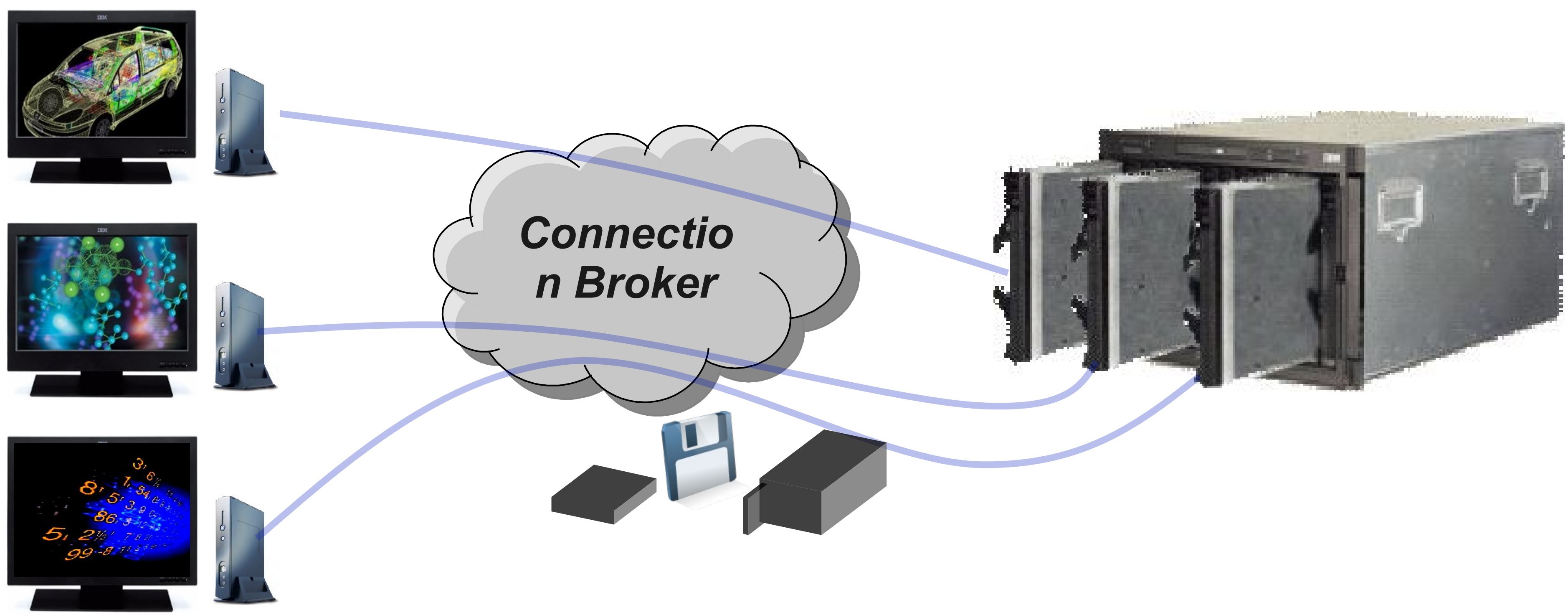
Eliminate the need for moves, adds or changes for call centers. Allow in house developers access to workspace while keeping IP safe in data center.

Thin clients



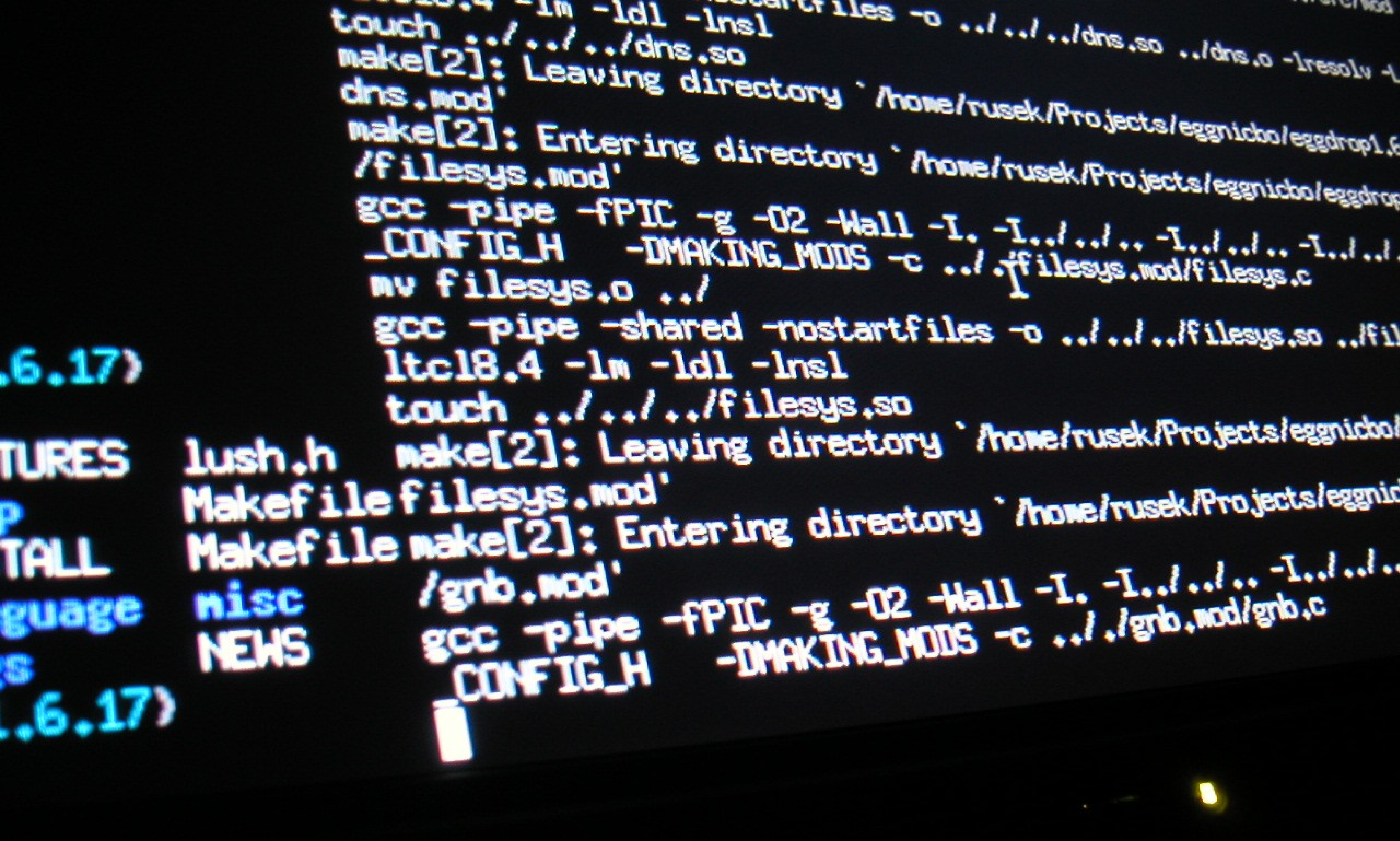
IBM HC10 Blades

Workstation Blade Architecture



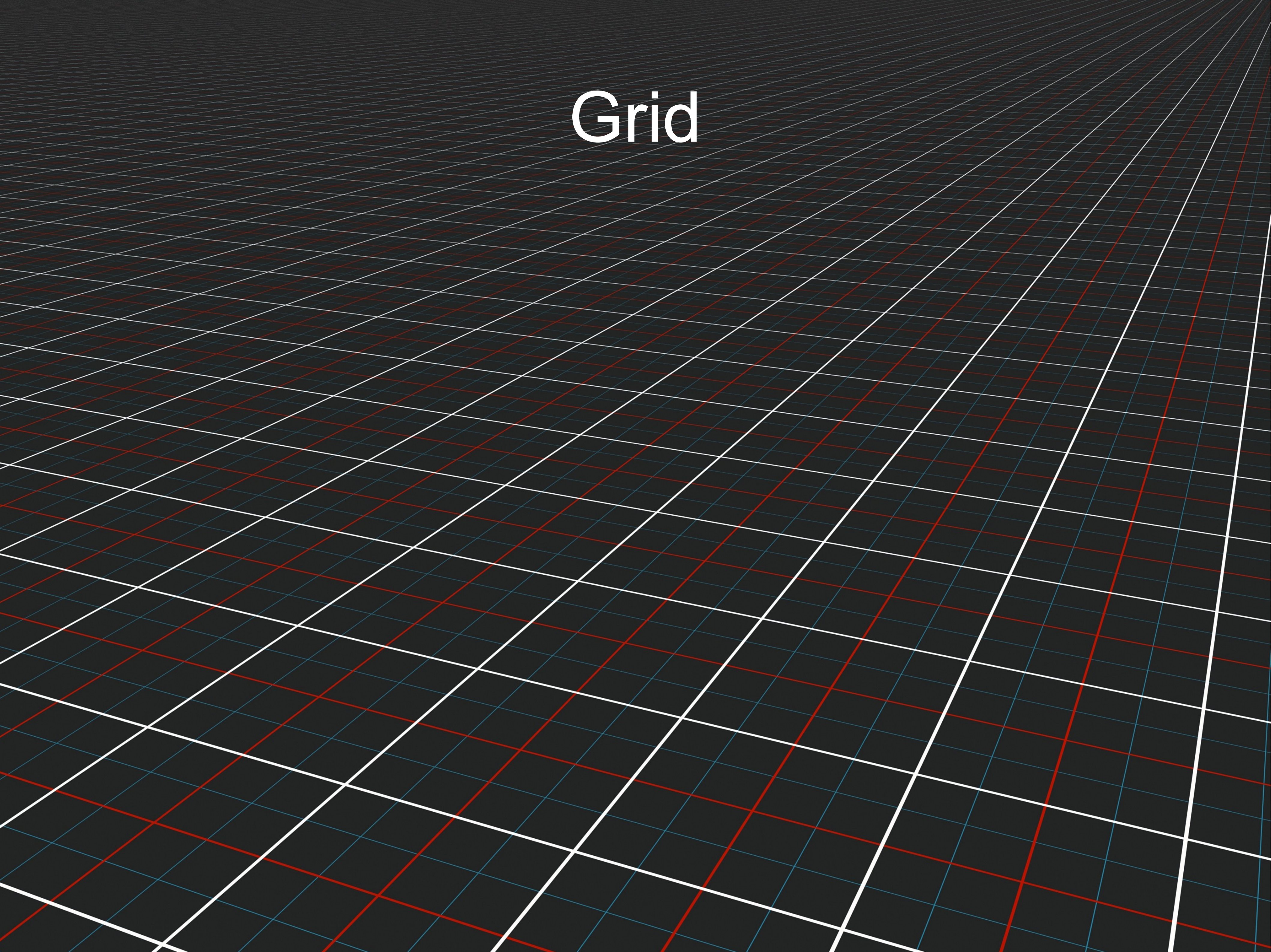
CP20

Dimension with stand	*232mm (H) x 94mm (W) x 174mm (D)*
Weight with stand	*1,100grams*
Ethernet connection to HC10	*10/100/1000BASE-T IEEE 802.3 compliant*
Display	*Two independent DVI-I*
USB	*USB 1.1 compliant, (2) in rear and (2) in front*
Audio	*HD Audio speaker jack (rear), Head phone jack (front), microphone in jack (front)*
LED	*CP20 Power Status LED, Connected HC10 Power Status (sleep mode indicator) LED, Session Status LED*
Control button	*Remote HC10 power button, Session disconnection button*
Acoustics	*0db (fanless)*
Power consumption	*25W (nominal)*



Applicatoin

Grid



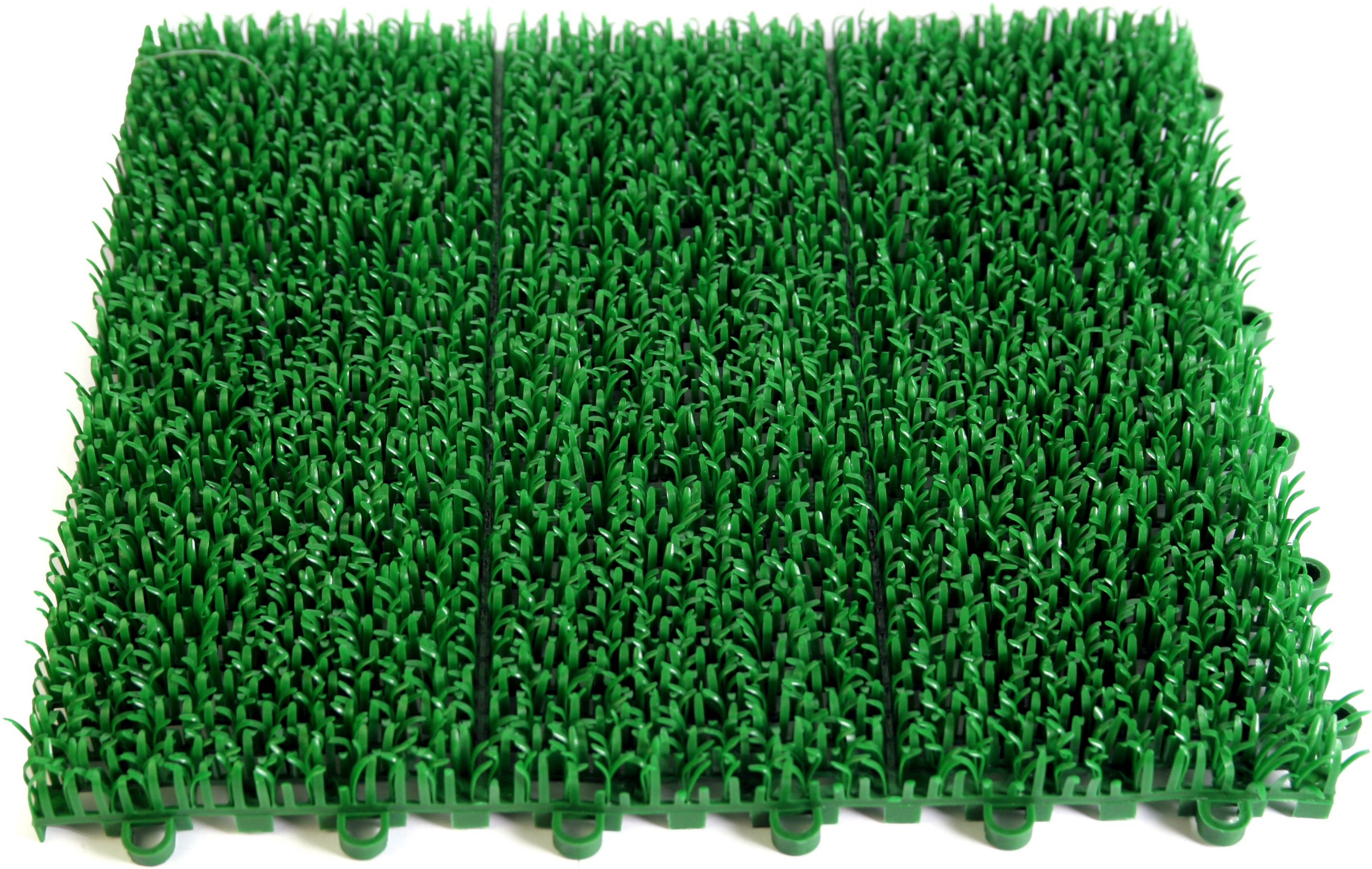
Chroot (jail)

ThinApp

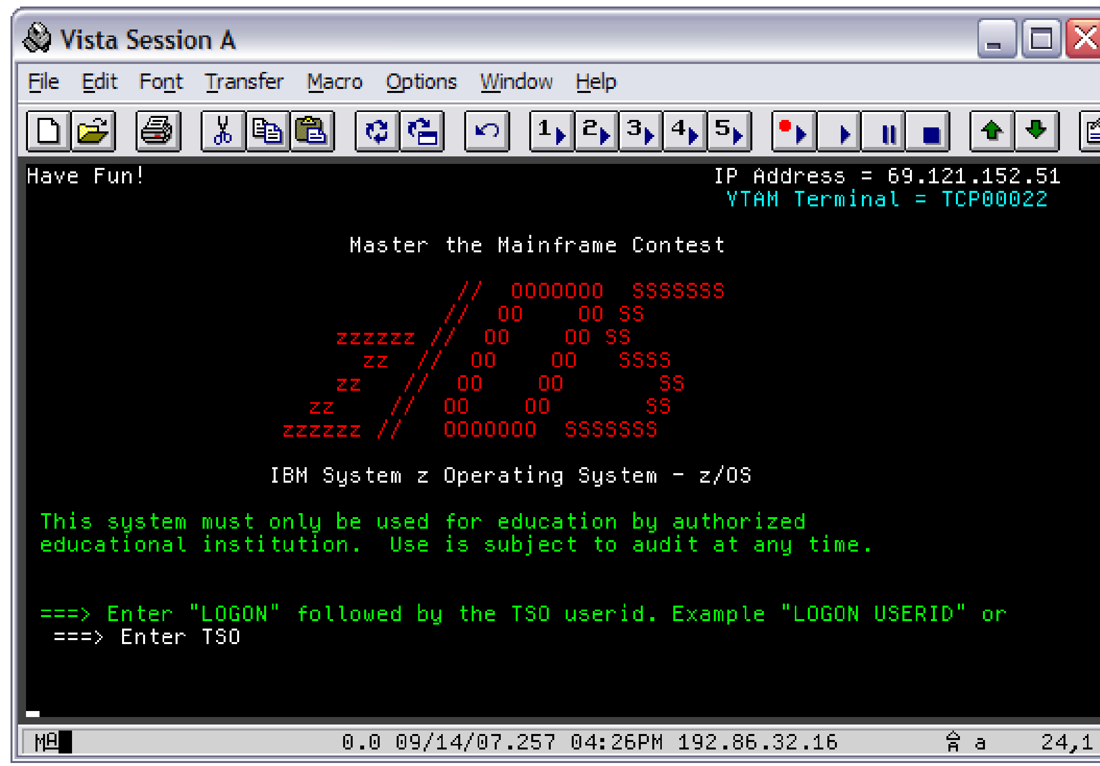
XenApp

Application Virtual Machines

Emulation



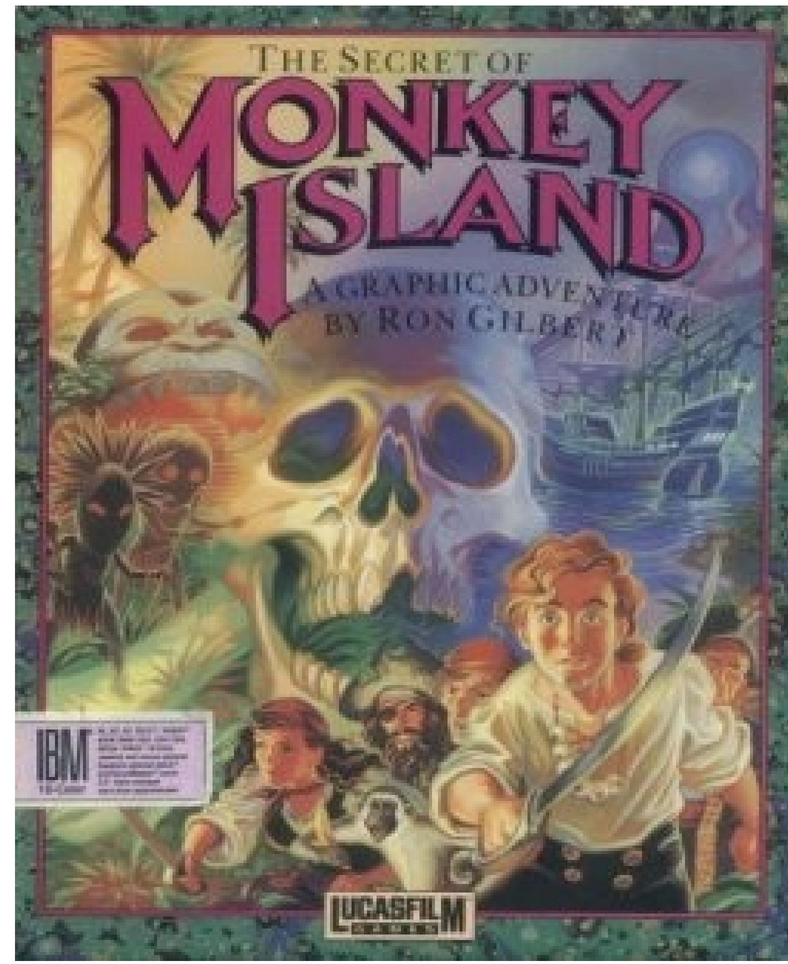
3270


Vista Session A
File Edit Font Transfer Macro Options Window Help
Have Fun!
IP Address = 69.121.152.51
VTAM Terminal = TCP00022
Master the Mainframe Contest
IBM System z Operating System - z/OS
This system must only be used for education by authorized
educational institution. Use is subject to audit at any time.
===> Enter "LOGON" followed by the TSO userid. Example "LOGON USERID" or
===> Enter TSO
0.0 09/14/07.257 04:26PM 192.86.32.16
24,1

Wine

ScummVM

Script Creation Utility for Maniac Mansion

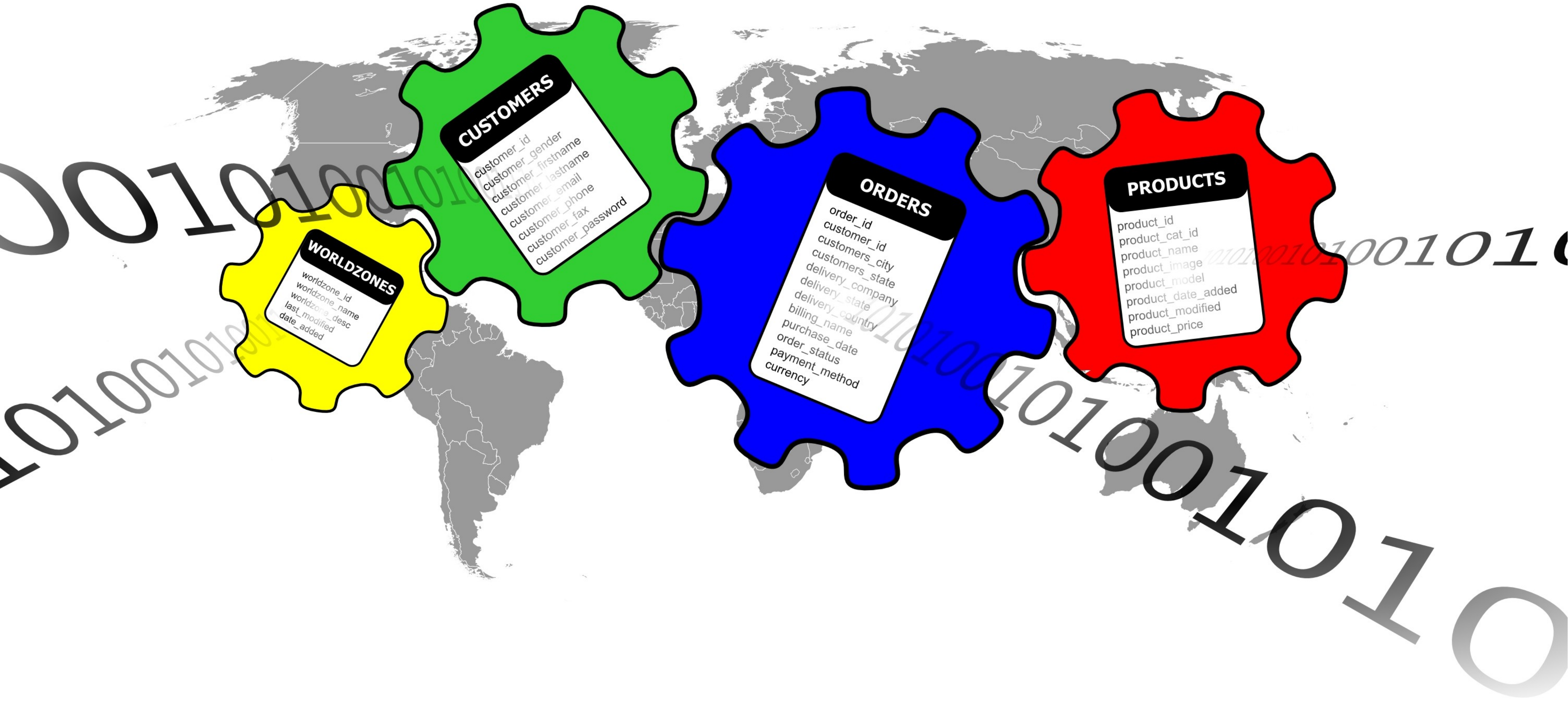


Wii

Architecture Translation

The Future

SaaS


CUSTOMERS
customer_id
customer_gender
customer_firstname
customer_lastname
customer_email
customer_phone
customer_fax
customer_password
WORLDZONES
worldzone_id
worldzone_name
worldzone_desc
last_modified
date_added
ORDERS
order_id
customer_id
customers_city
customers_state
delivery_company
delivery_state
delivery_country
billing_name
purchase_date
order_status
payment_method
currency
PRODUCTS
product_id
product_cat_id
product_name
product_image
product_model
product_date_added
product_modified
product_price

App Engine

SimpleDB

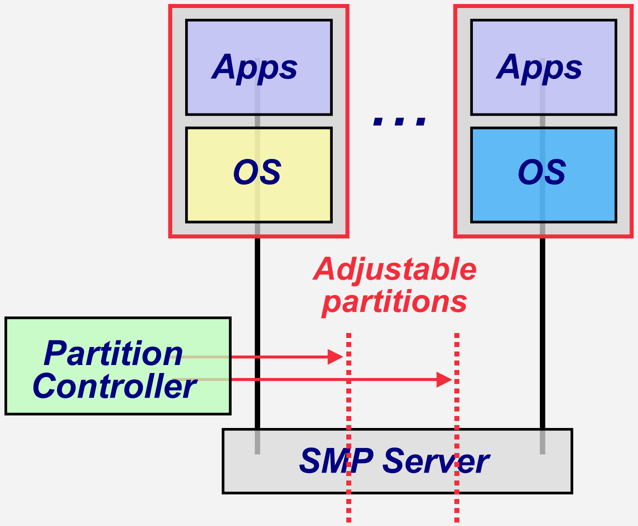

BigTable

If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.

—John McCarthy, MIT Centennial in 1961

Cloud / Utility

Hardware Partitioning



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

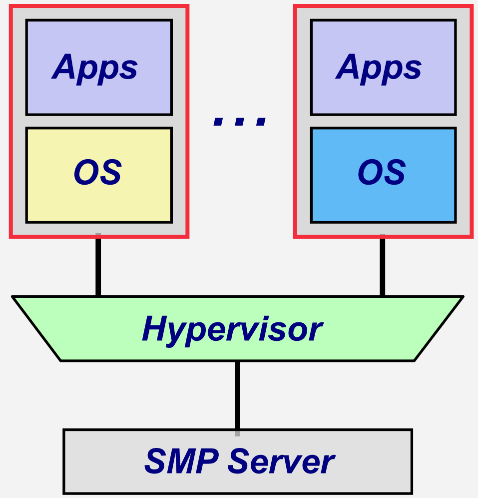
Physical partitioning

S/370™ SI-to-PP and PP-to-SI, Sun Domains, HP nPartitions

Logical partitioning

IBM eServer™ pSeries® LPAR, HP vPartitions

Bare-metal Hypervisor

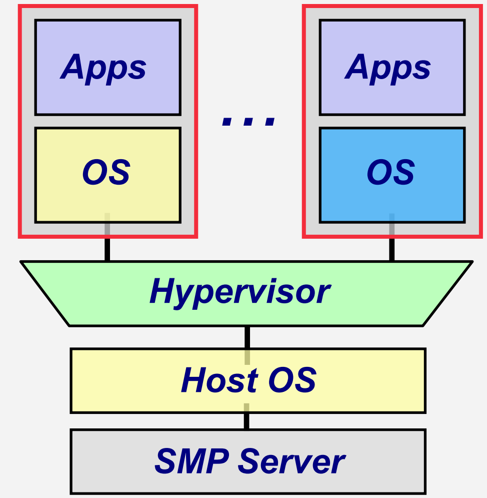


Hypervisor provides fine-grained timesharing of all resources

Hypervisor software/firmware runs directly on server

System z LPAR and z/VM®
POWER™ Hypervisor
VMware ESX Server
Xen Hypervisor

Hosted Hypervisor



Hypervisor uses OS services to do timesharing of all resources

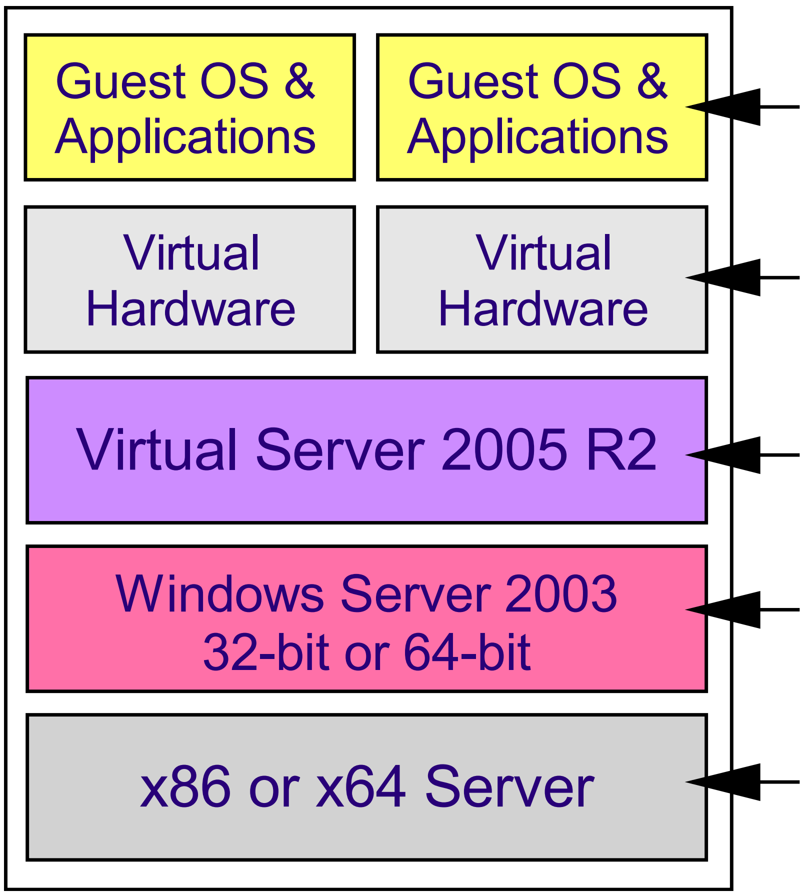
Hypervisor software runs on a host operating system

VMware Server
Microsoft® Virtual Server
HP Integrity VM
User Mode Linux®

Characteristics:

- *Bare-metal hypervisors offer high efficiency and availability*
- *Hosted hypervisors are useful for clients where host OS integration is important*
- *Hardware partitioning is less flexible than hypervisor-based solutions*

Microsoft Virtual Server 2005



Windows NT 4.0 Server, Windows 2000 Server, Windows 2003 Server supported (32-bit)

1 CPU per virtual machine
Up to 3.6 GB of memory per virtual machine

Multithreaded virtual machine monitor (VMM) provides isolation

Windows Server 2003
Broad device compatibility

Optimized for 2 to 8 way servers
Scales up to 32 CPUs and up to 64 GB

Source: "Microsoft Virtual Server 2005 R2 Technical Overview", December 2005

Experiments

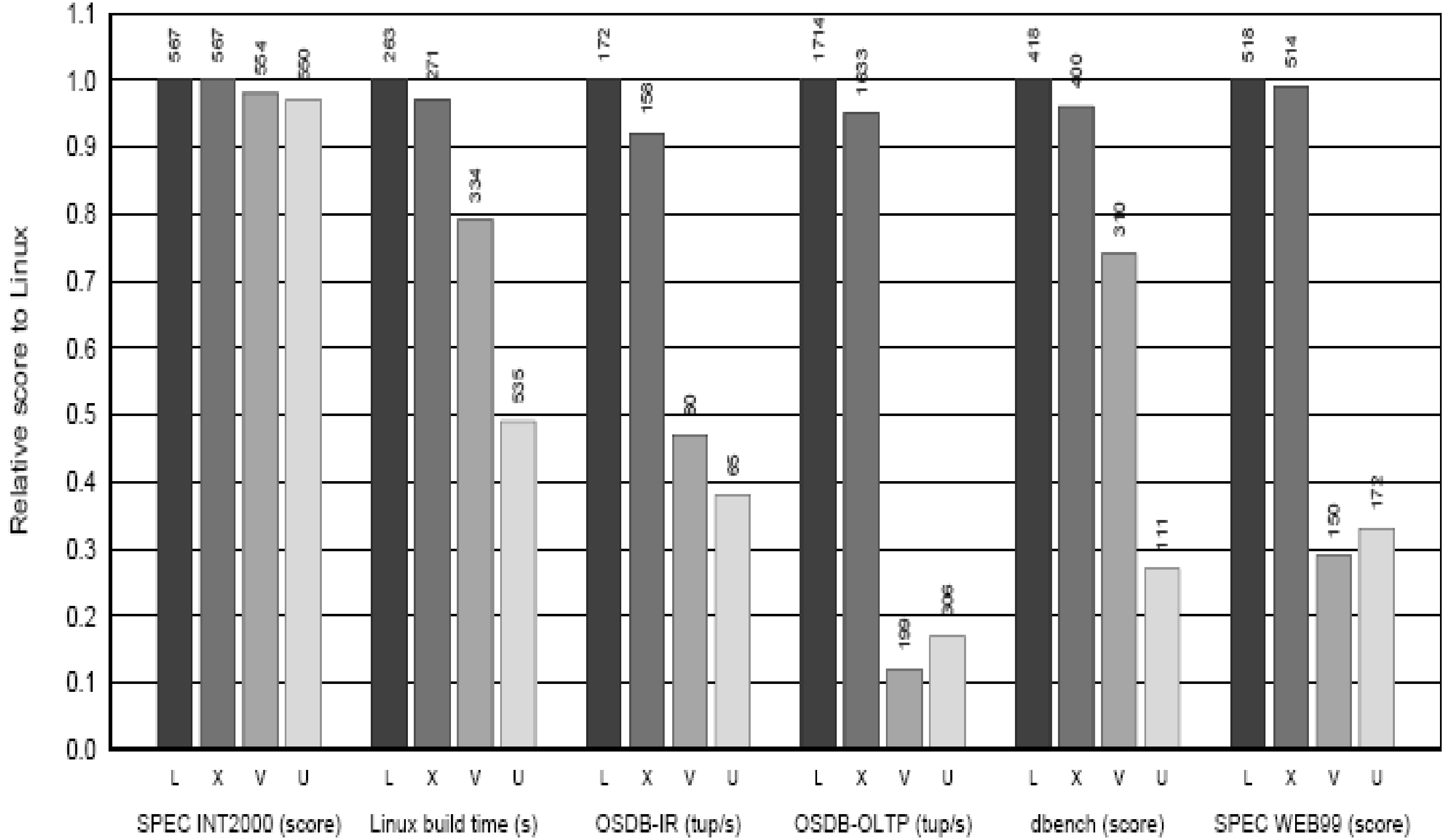
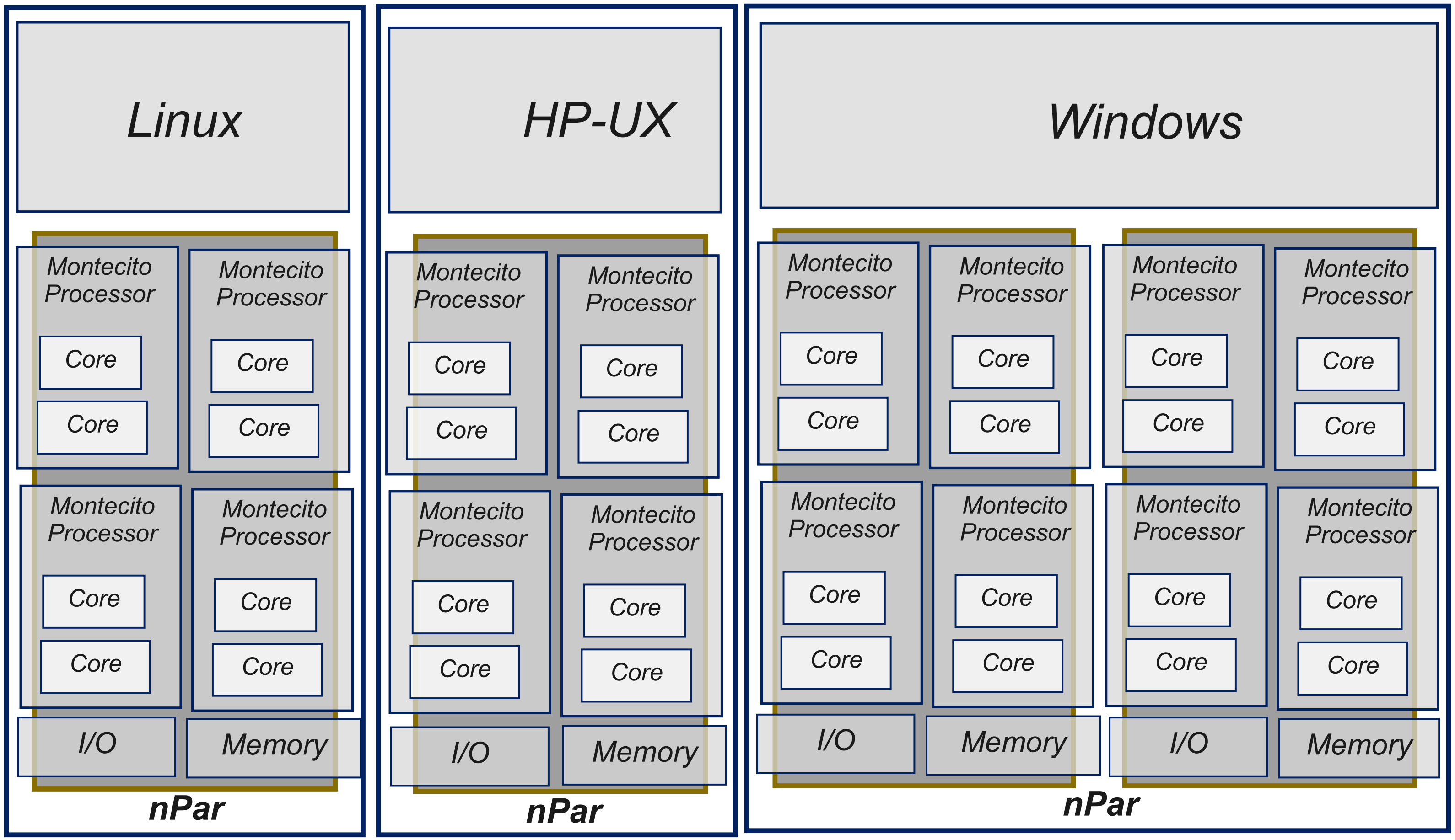


Figure 3: Relative performance of native Linux (L), XenoLinux (X), VMware workstation 3.2 (V) and User-Mode Linux (U).

HP Integrity Servers - nPars



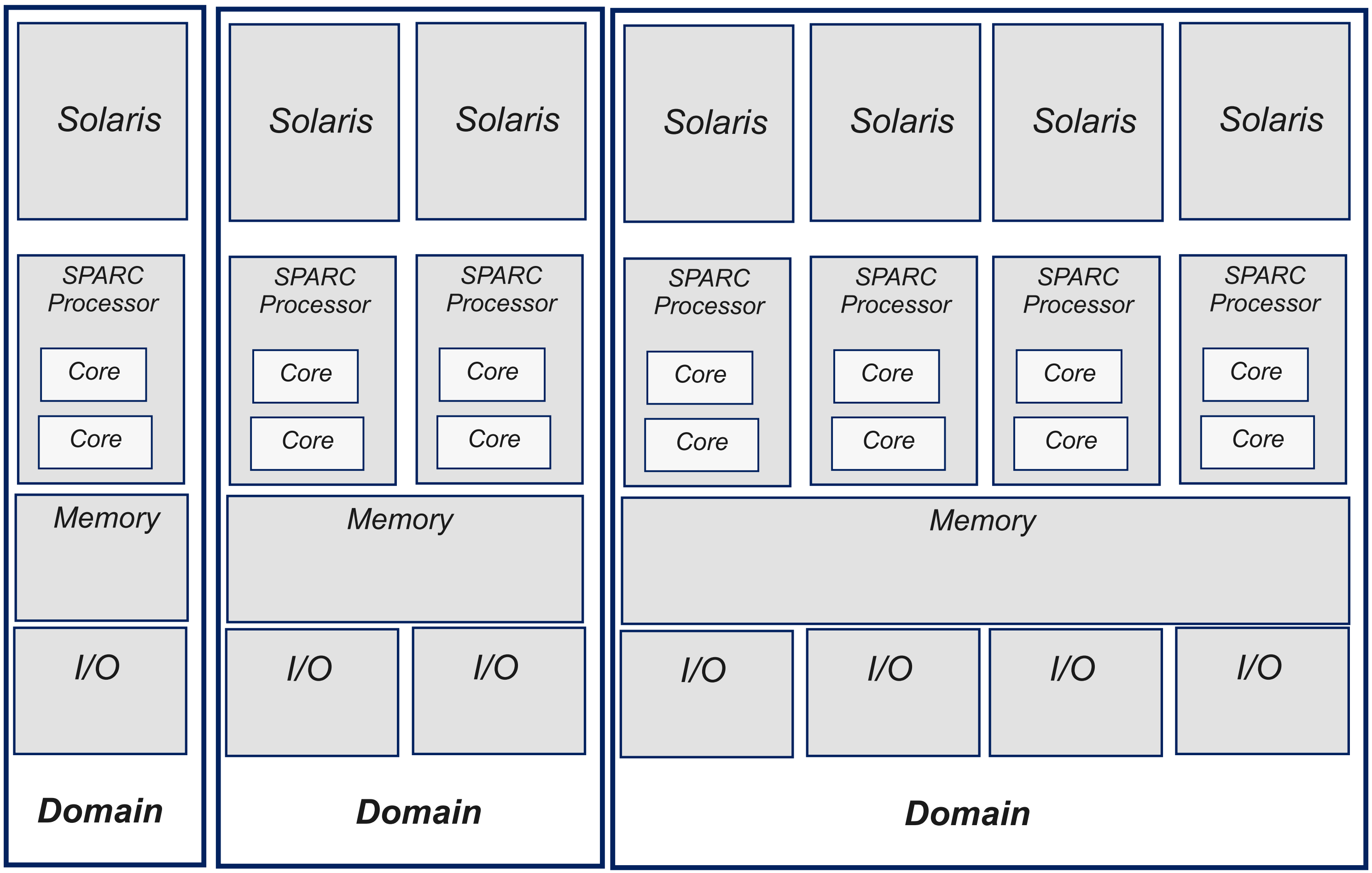
HP Integrity Servers Soft Partitions / VMs

- Virtual Partitions (vPars)
  - Soft Partitions for HP-UX workloads
  - Each vPar has subset of CPUs, memory & I/O
  - Granularity of one or more cores per vPar
  - Minimum performance overhead
  - Dynamically reconfigurable
- Integrity Virtual Machines
  - VM platform hosted on HP-UX
  - Can host HP-UX, Linux or Windows operating systems
  - Shared I/O
  - Granularity up to 20 VMs per core
  - Greater performance overhead than nPars & vPars

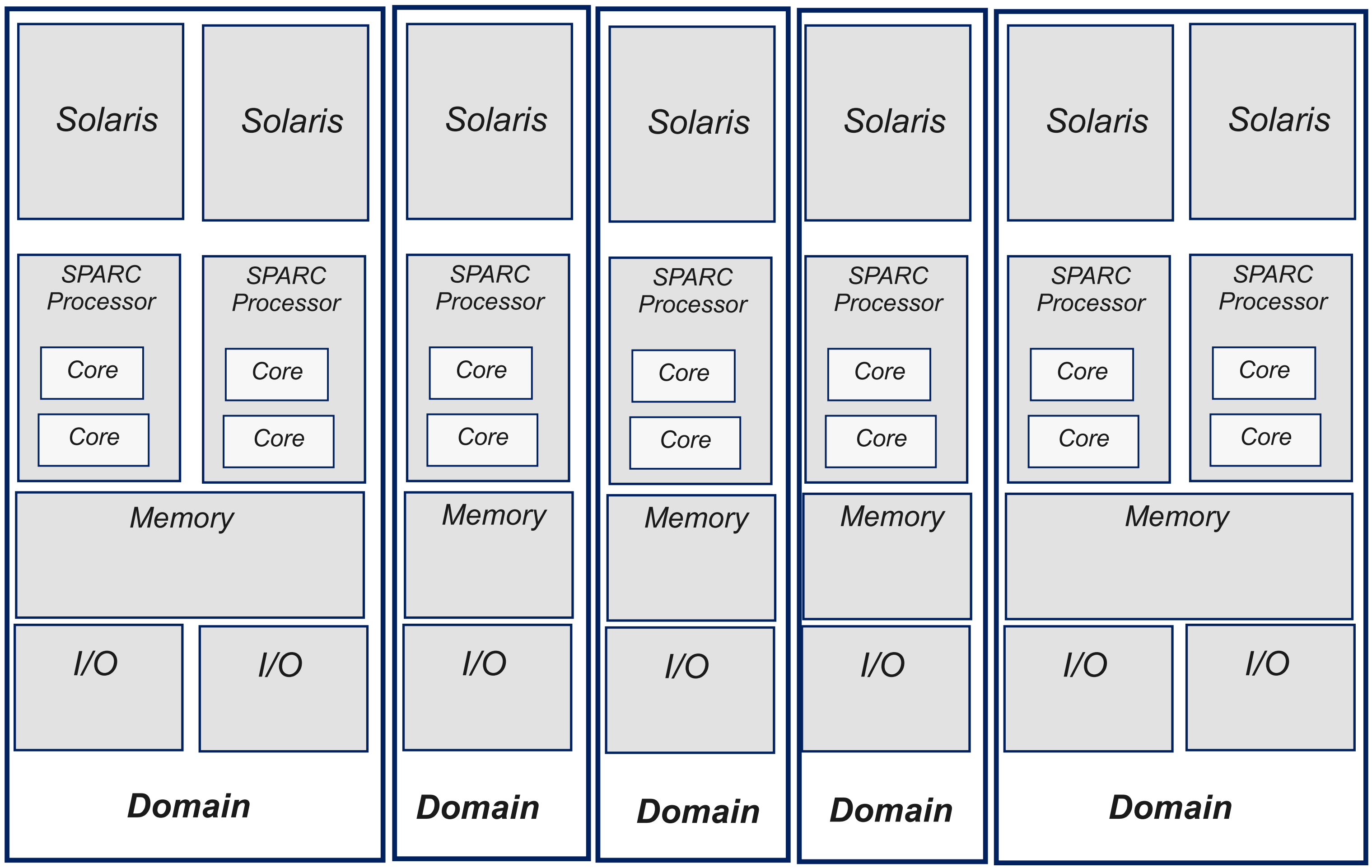
HP Virtual Server Environment (VSE)

- Tight integration of key system components
  - Virtualization functions (nPars, vPars, Integrity Virtual Machines)
  - Workload management tools (gWLM)
  - Serviceguard HA/DR tools
  - Utility pricing
- Helps maintain service levels and increase business agility
  - Admins control which apps are the most important
  - Designate how much of available resources apps get
  - Can automatically / dynamically readjust resource allocations in response to changes in workload demand or failure conditions
  - Goal-based WLM policies

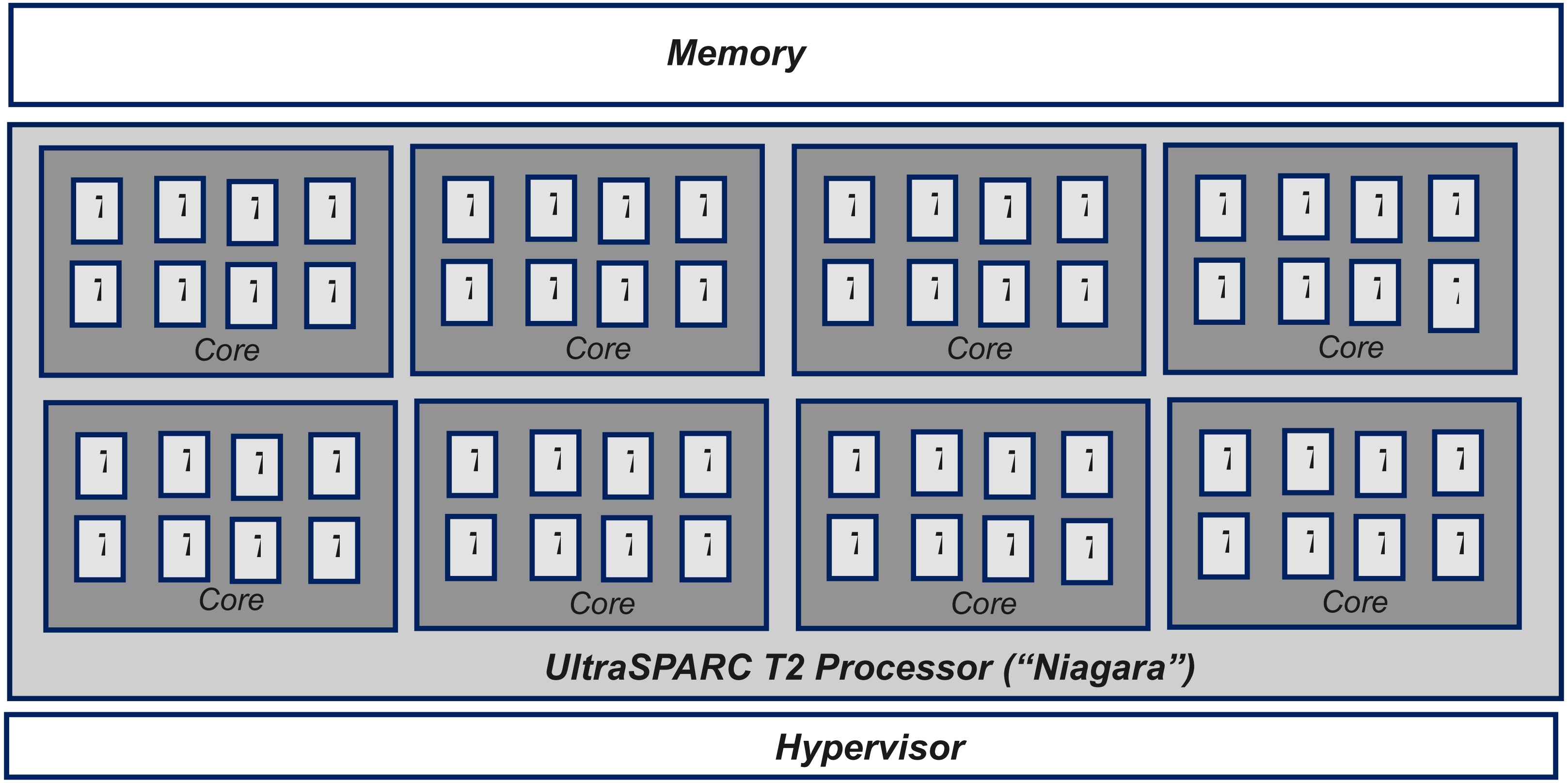
Sun Dynamic Domains



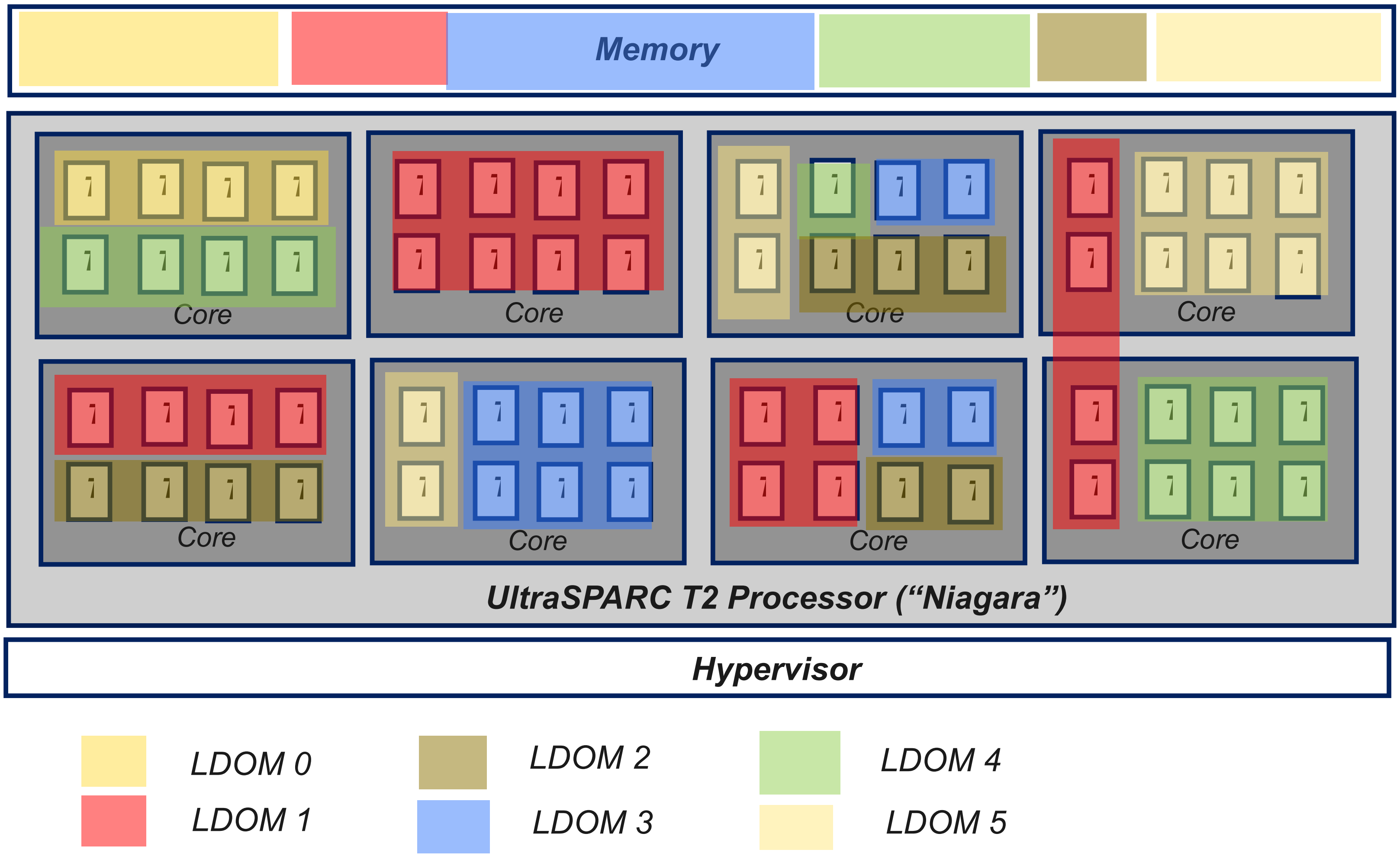
Sun Dynamic Domains



Sun SPARC Servers
UltraSPARC T2 CoolThreads



Sun CoolThreads & Logical Domains (LDOMs)



x86 Server Virtualization Options

- Virtual Machine Platforms
  - VMware
  - Xen implementations
    - Citrix/XenSource
    - Virtual Iron
    - Red Hat Enterprise Linux
    - SUSE Linux Enterprise Server
    - Sun xVM
    - Oracle
  - Microsoft Hyper-V (July)
  - Linux Kernel Virtual Machine (KVM)

» *Virtual Server Solutions*

- Parallels (Swsoft Virtuozzo)
- Solaris Containers

» *Hardware-assisted Partitions*

- Unisys ES7000
- IBM X Architecture
- Hitachi Virtage

The Players -- Server Virtualization

- VMware
  - ESX Server, Virtual Infrastructure 3
  - VMware Server
- Microsoft
  - Windows Server Virtualization Service (coming in Windows Server 2008)
  - Virtual Server 2005
- Xen
  - XenSource, Virtual Iron, Novell, Red Hat

The Players – OS Virtualization

- SWsoft
  - Virtuozzo
  - Linux, Windows
- Sun
  - Solaris Containers
- HP
  - Virtual Server Environment (VSE)
- IBM
  - Virtual Partition Manager

Non-x86 Virtualization Platforms

- IBM Mainframes (System z)
  - z/OS, Linux
- IBM System p
  - UNIX (AIX), Linux
- HP Integrity
  - UNIX (HP-UX), Linux, Windows
- Sun SPARC
  - UNIX (Solaris)

Non-x86 Virtualization Platforms

- IBM Mainframes (System z)
  - z/OS, Linux
- IBM System p
  - UNIX (AIX), Linux
- HP Integrity
  - UNIX (HP-UX), Linux, Windows
- Sun SPARC
  - UNIX (Solaris)

Network Virtualization

- Emerging Technology
- Cisco (Network-Based)
  - The efficient utilization of network resources through logical segmentation of a single physical network.
  - Generic Routing Encapsulation (GRE)
  - Virtual Routing and Forwarding (VRF)
  - Multi-protocol Label Switching (MPLS) VPNs
  - Single Physical Network -- multiple closed logical groups

Network Virtualization

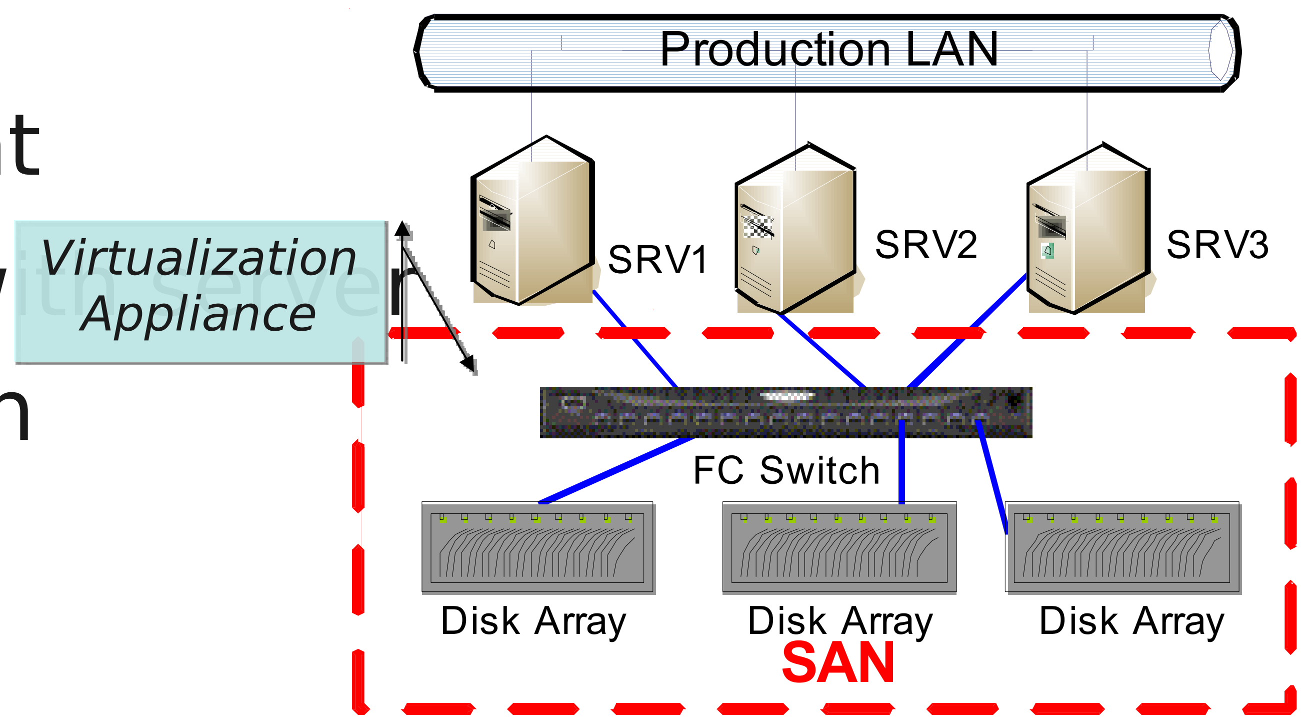
- Sun Crossbow (Host-Based)
  - Physical NIC divided into multiple virtual NICs
  - Bandwidth resource control and virtualization included in the network stack
  - Bandwidth and priority dynamically assigned to services (FTP, SMB, etc) or Virtual Machines

Storage Virtualization

- Host-Based
  - RAID
- Network-Based
  - In-band
    - Virtualization appliance resides directly in the data path
    - Commonly associated with SANs
      - Cisco, IBM, EMC, Pillar
  - Out-of-band
    - Virtualization appliance resides outside of the data path
      - DFS, Global Namespace

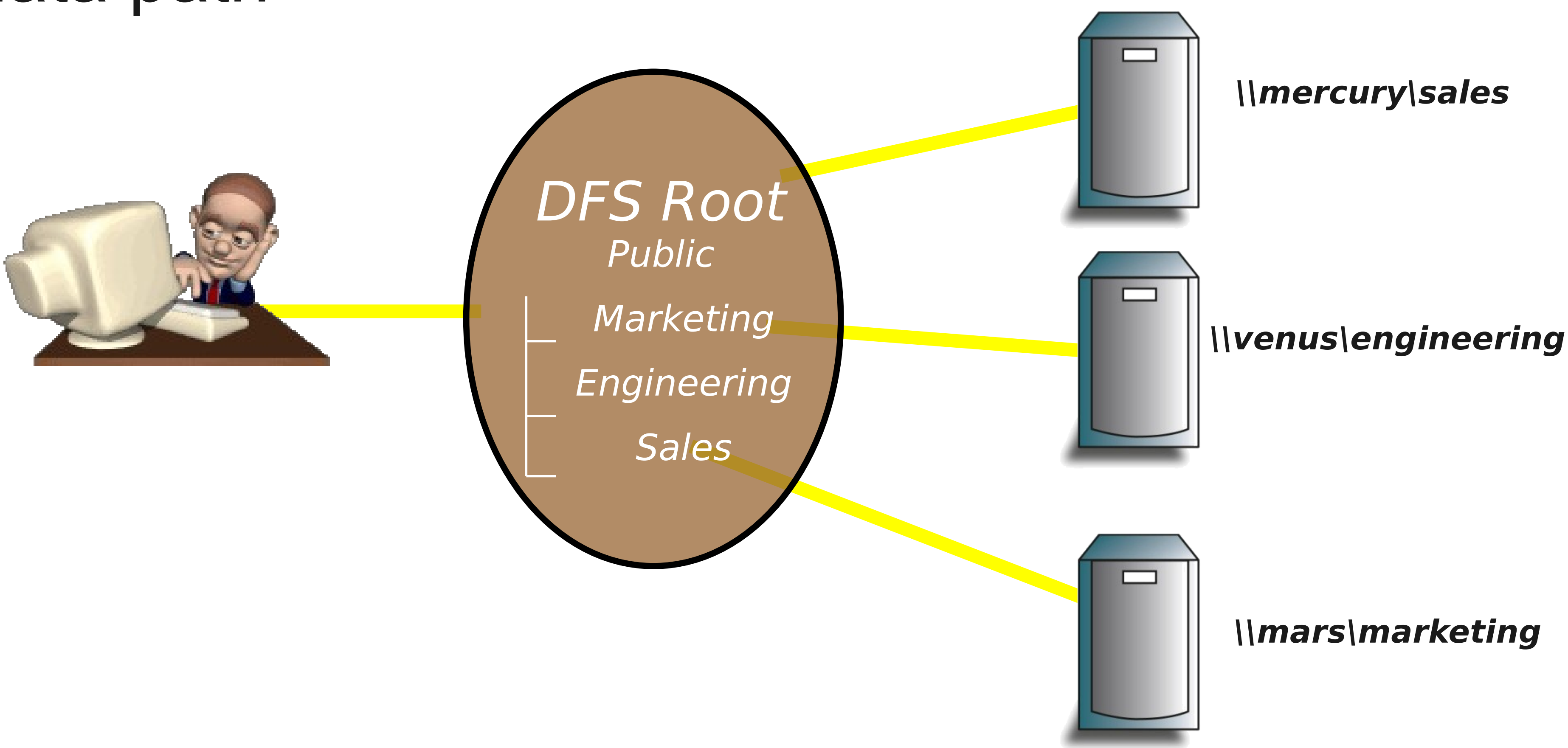
In-band Virtualization

- Physical storage resources translated into virtual resources
- Replication
- Data Management
- Combined with server virtualization



Distributed File Systems

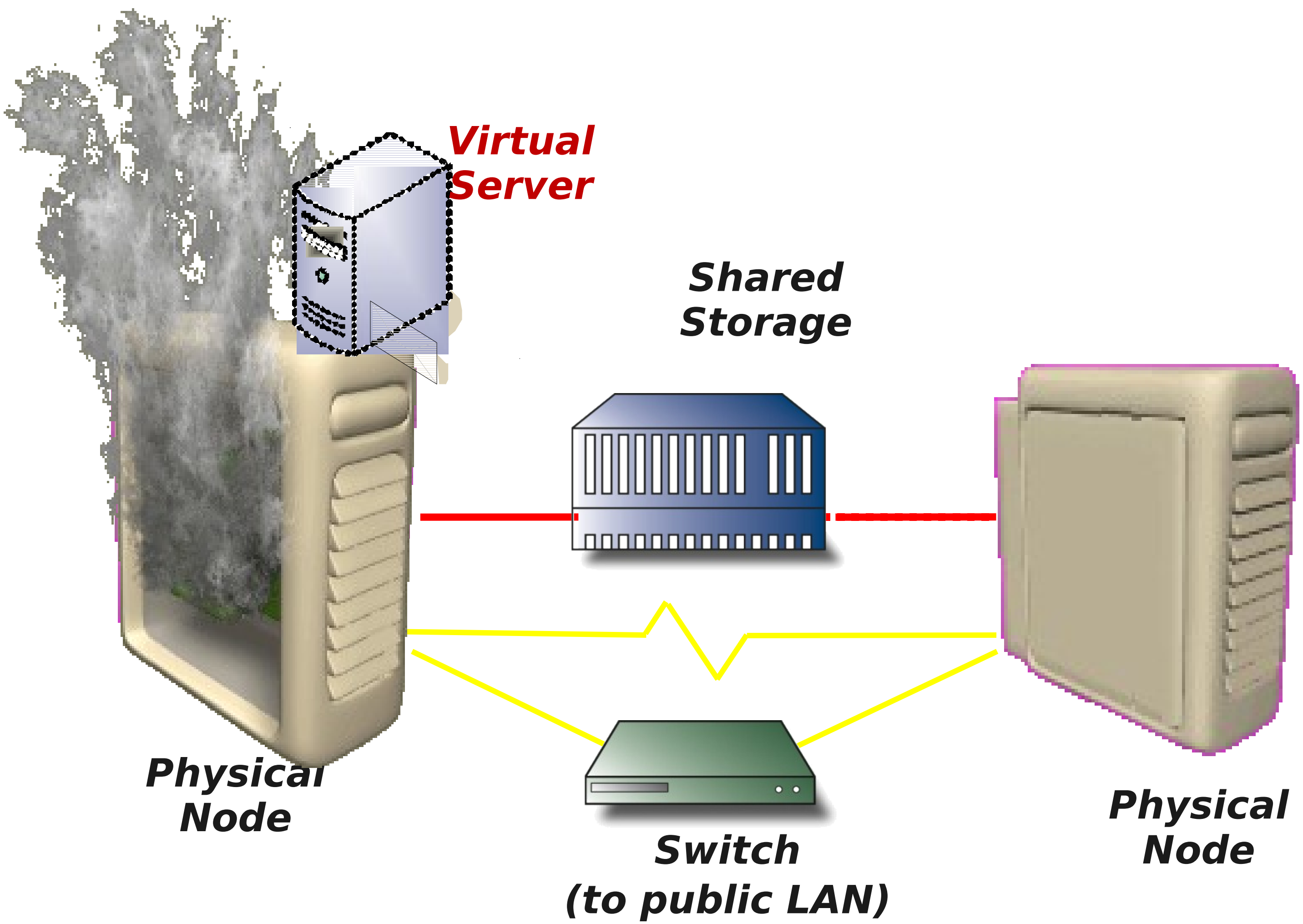
- Abstract the physical storage location from the data path



Clustering

- Applications run within virtual servers that reside on one or more physical host systems.
- Common Implementations:
  - Failover Clusters
  - Load Balanced Clusters
  - Shared Data Clusters

Failover Clustering



Virtualizing Storage

- More efficient utilization of storage resources
- Simplifies administration
- Provides additional methods for data protection
- Better Server virtualization flexibility
- Considerations:
  - Integration with existing hardware platforms and operating systems
  - Integration with existing data protection/backup products

Server Virtualization Methods

Technique	Advantages	Disadvantages
Hard Partitions	•Maximum isolation between virtualized workloads •Can protect workloads against some HW failures •Can enable online HW upgrade	•Limited granularity (1+ CPUs/ partitions) •May have limited flexibility (w/ out OS support) •Reconfiguration can be time-consuming
Logical ("Soft") Partitions (LPARs)	•Finer granularity than hard partitions	•Lacks protection from extreme hardware failures
Virtual Machines	•Finer granularity than LPARs	•Incurs higher performance overhead
Virtual Servers (OS Virtualization)	•Maximum granularity & responsiveness	•All workloads must run on same OS instance (i.e. kernel revision & patch level)

Positioning Virtualization Options

Platform	Key Benefits	Notes
IBM System z Mainframes	•*Industry-leading maturity, performance & functionality*	•*Suitable for z/OS & Linux only*
IBM System p	•*Strong virtual infrastructure support w/live migration capabilities*	•*Suitable for UNIX, i5/OS & Linux only*
HP Integrity	•*Heterogeneous platform: suitable for UNIX, Linux & Windows workloads* •*Leading HA/DR & WLM integration w/HP Virtual Server Environment*	•*I/O overhead in Integrity Virtual Machines* •*Lacks migration functions for virtualized workload*
Sun SPARC Servers	•*Proven isolation & reconfiguration in Dynamic Domains*	•*Suitable for UNIX only* •*LDOMs limited by lack of true SMP in Niagara* •*Lacks migration functions for virtualized workload*
x86/x64 Hardware	•*Growing choice of platforms*	•*I/O overhead still major concern in VMs*

x86 Comparison

Name	Host CPU	Guest CPU	Host OS	Guest OS
Xen	*x86, AMD64, (PowerPC and IA-64 ports in progress)*	*(Same as host)*	*NetBSD, Linux, Solaris*	*Linux, Solaris, Windows XP & 2003 Server, Plan 9*
Hyper-V	*x64 + (Intel VT or AMD-V)*	*x64,x86*	*Windows 2008 w/ Hyper-V Role*	*Windows 2008, Windows XP, Windows Vista, Linux*
VMware ESX Server 3.0	*x86, AMD64*	*x86, AMD64*	*none (bare metal install)*	*Windows, Red Hat, SuSE, Netware, Solaris*
KVM	*Intel/AMD processor with X86 virtualization*	*x86/AMD64*	*Linux*	*Linux, Windows*

System p

Name	Host CPU	Guest CPU	Host OS	Guest OS
PowerVM	POWER4, PowerPC 970, POWER5, POWER6	POWER4, PowerPC 970, POWER5, POWER6	hardware / firmware, no host OS	Linux-PPC, AIX, i5/OS

System z

Name	Host CPU	Guest CPU	Host OS	Guest OS
z/VM	z/Architecture	z/Architecture	None or itself	Linux on zSeries, z/OS, z/VSE, z/TPF, z/VM, VM/CMS, MUSIC/SP, and predecessors
z LPARs	z/Architecture	z/Architecture	Intrinsic feature of System z mainframes	Linux on zSeries, z/OS, z/VSE, z/TPF, z/VM, VM/CMS, MUSIC/SP, and predecessors

Features

Name	Can boot an OS on another disk partition as guest	USB	GUI	Live memory allocation	3D acceleration	Live migration
KVM	Yes	Yes	Yes		Supported with VMGL	Yes
PowerVM	Yes	Yes	No	Yes	No	Yes (on POWER 6-based systems, requires PowerVM Enterprise Licensing)
VMware ESX Server 3.0	Yes	Yes	Yes	No	No	Yes
Xen	Yes	Yes	Enomalism	Yes	Supported with VMGL	Yes
z/VM	Yes	N/A	with add-ons	Yes	No	with GDPS
z LPARs	Yes	N/A	Yes	Yes	No	with GDPS