Comparing and Contrasting Virtualization Technologies

12359

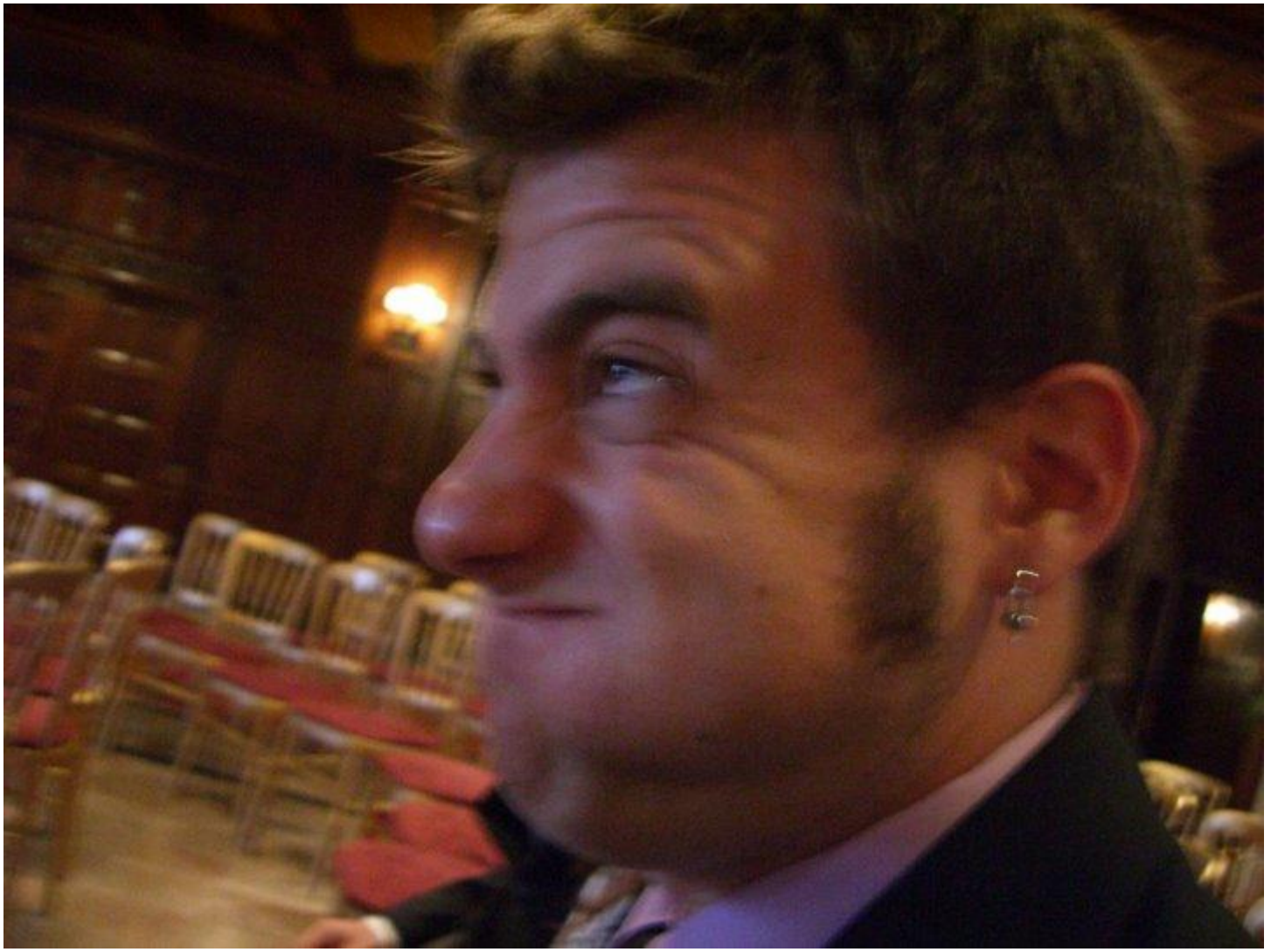
Do you know the differences between
Xen and VMware?

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

Virtualization can be a complicated subject with many different facets.

It is not always easy to choose the strategy that best fits your needs.

I am here to help
buzzetti@us.ibm.com



Worldwide Design Centers

Poughkeepsie, NY

design@us.ibm.com

Montpellier, France

design.center@fr.ibm.com

Makuhari, Japan

design@jp.ibm.com

Boeblingen, Germany

design@de.ibm.com

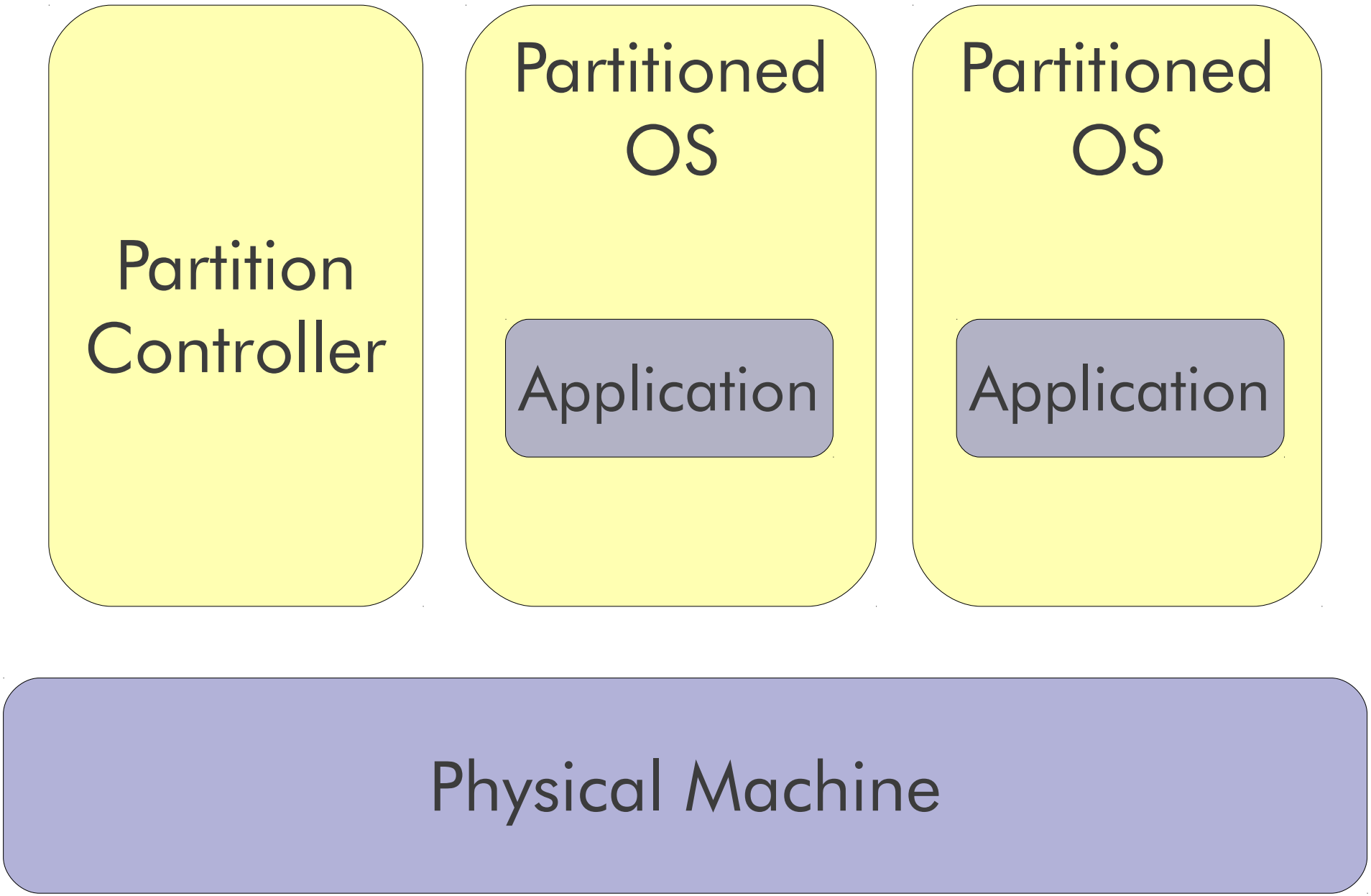
Worldwide Design Centers

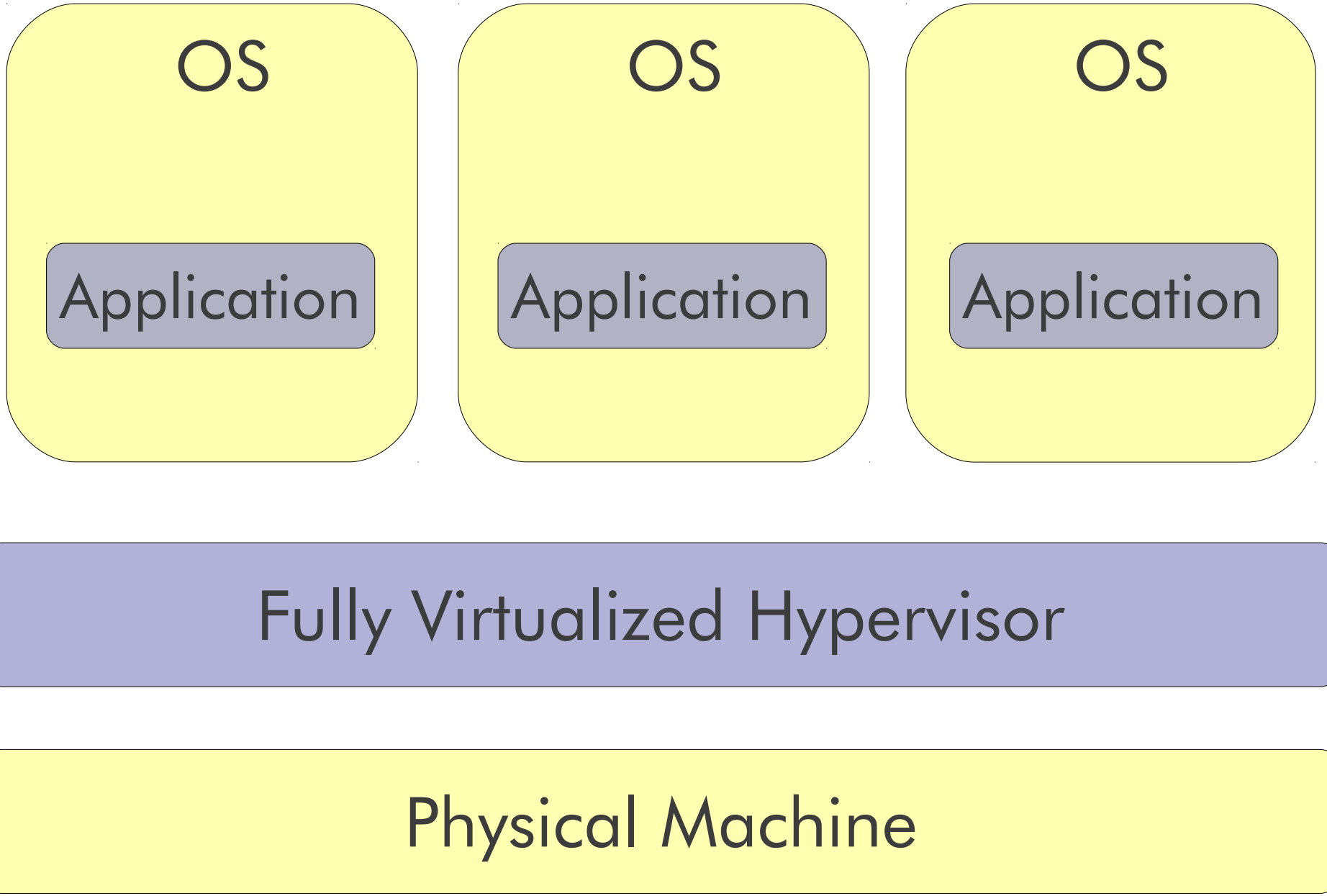
Our mission – To architect & design innovative end to end solutions with selected worldwide clients that leverage leading-edge IBM technologies to accelerate their IT transformation

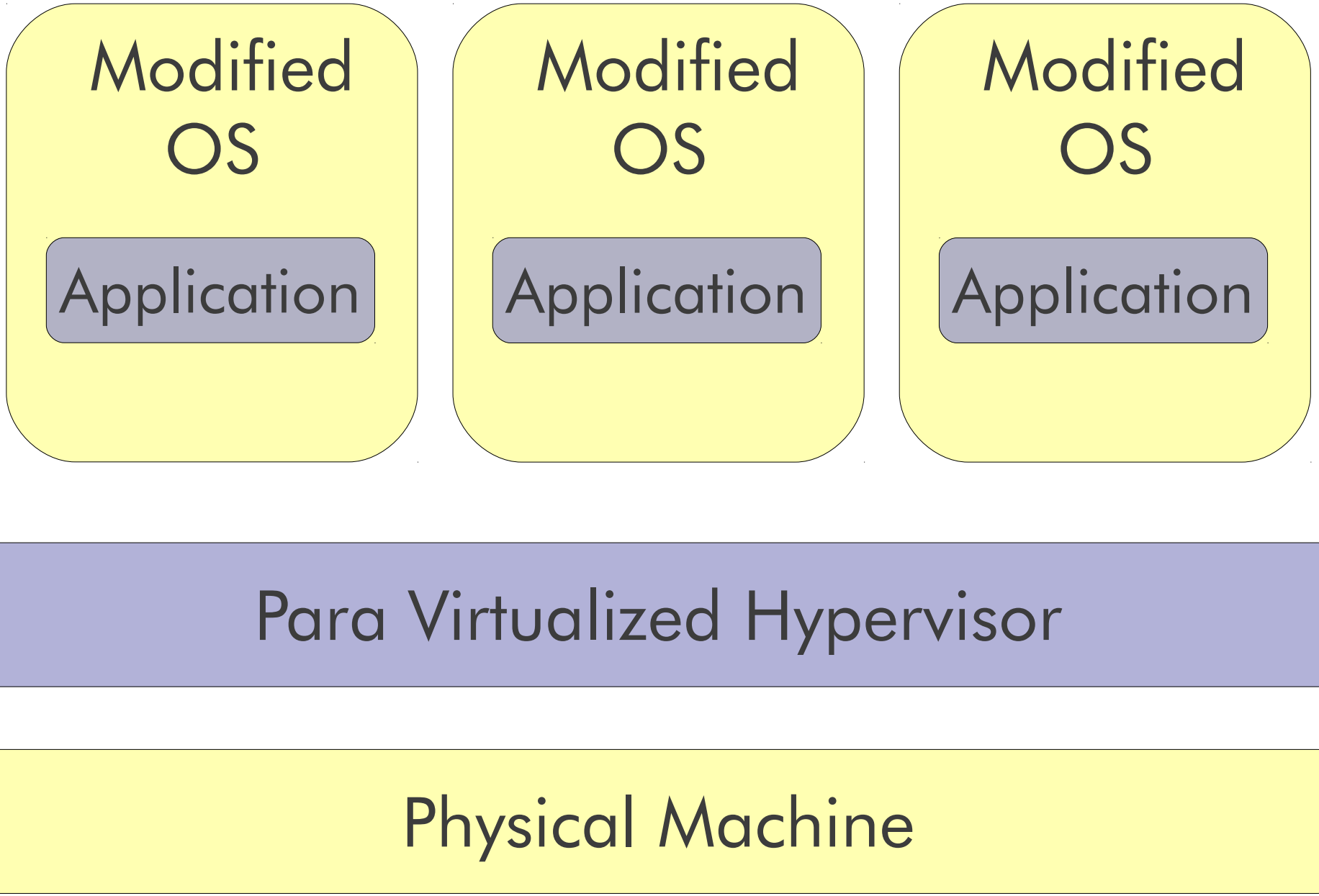
- **Analysis**
  Define and analyze various solution options that meet existing and future requirements
- **Assessment**
  Review and validate a planned solution, strategy, IT transformation, or architecture, and provide recommendations and roadmaps
- **Design**
  Define an architecture and high level design for an IT infrastructure to meet business requirements

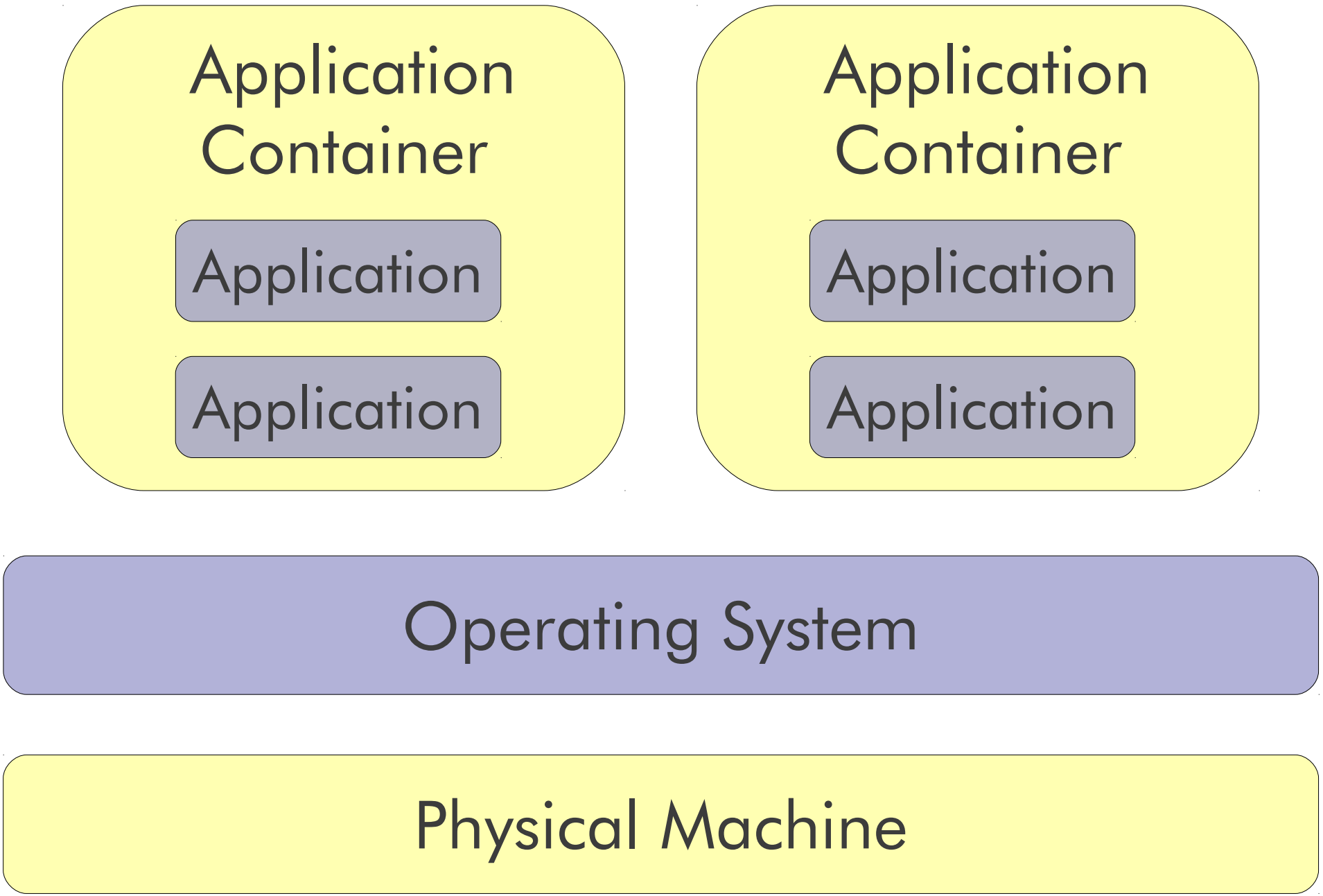
There is no free lunch

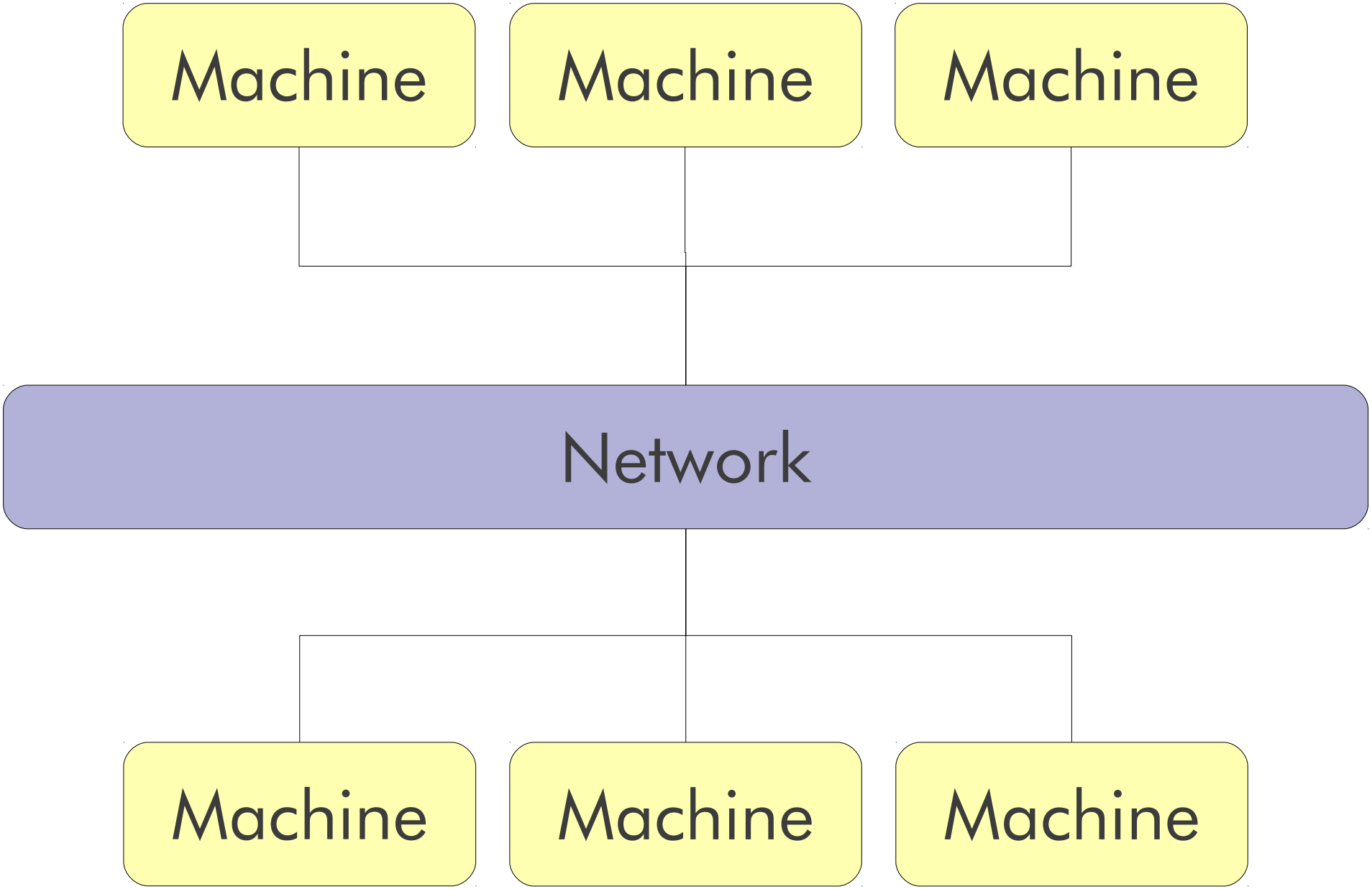
Types of Virtualization

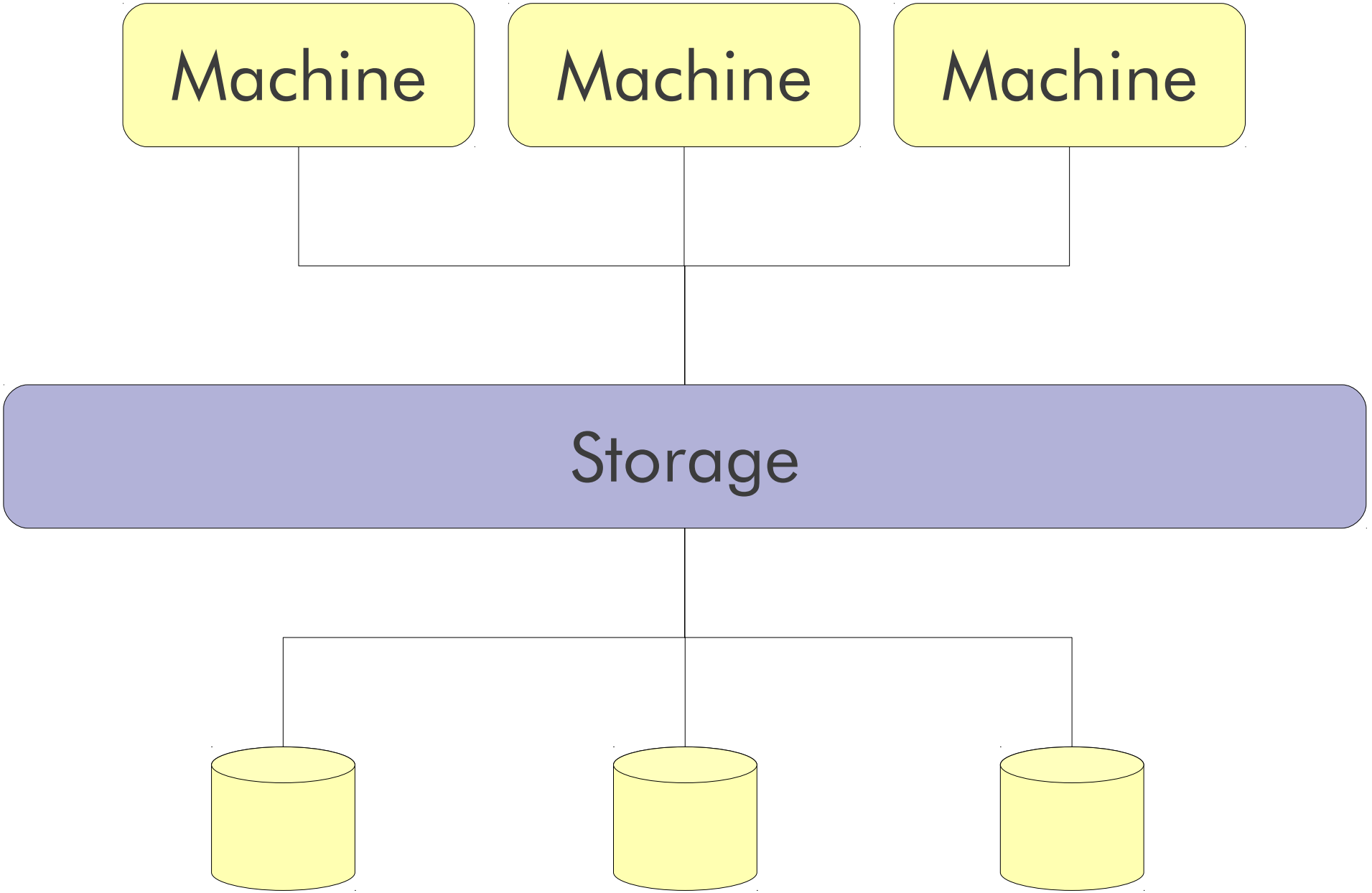

Partition Controller
Partitioned OS
Application
Partitioned OS
Application
Physical Machine


OS
Application
OS
Application
OS
Application
Fully Virtualized Hypervisor
Physical Machine


Modified OS
Application
Modified OS
Application
Modified OS
Application
Para Virtualized Hypervisor
Physical Machine

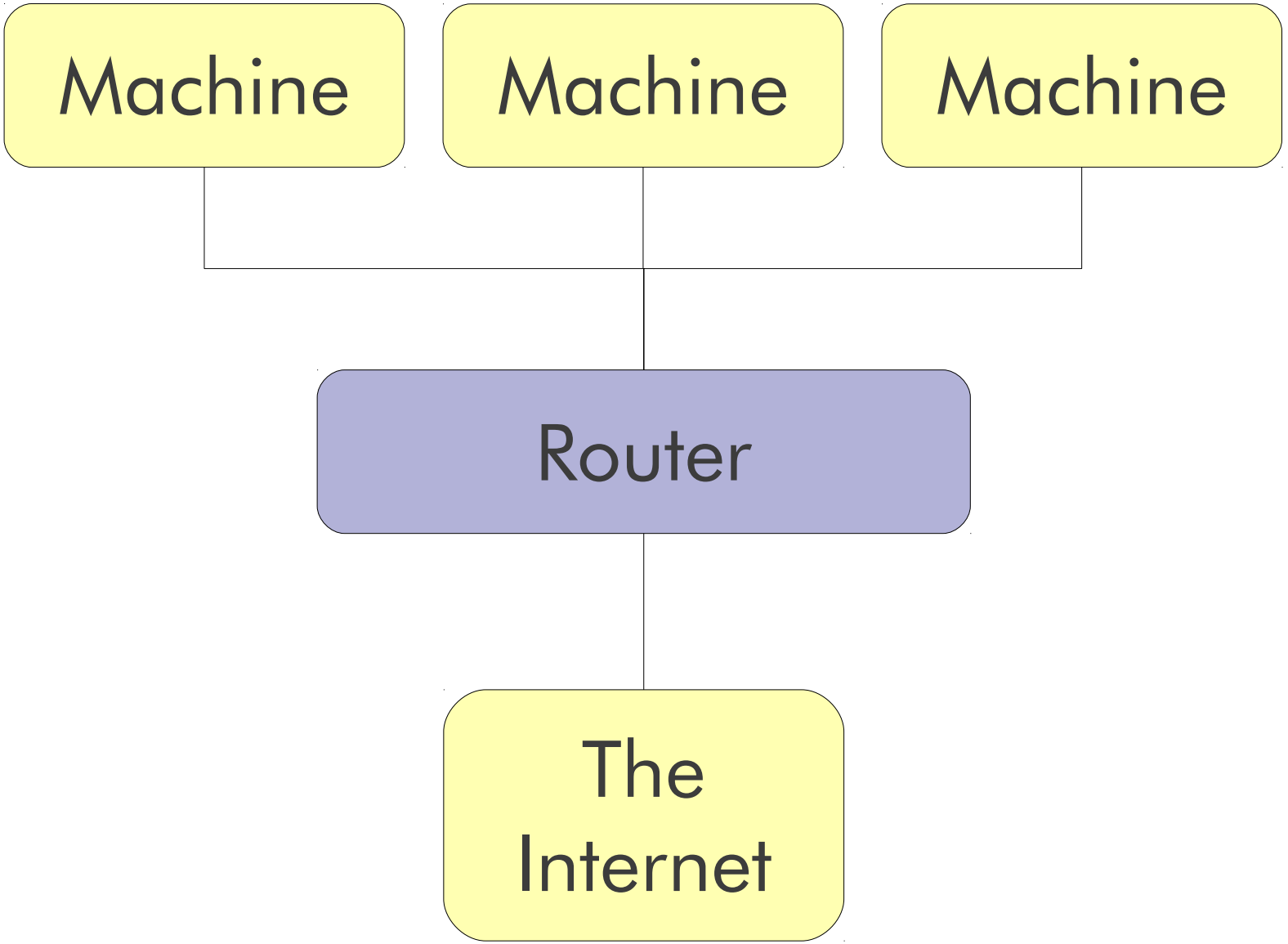

Application Container
Application
Application
Application Container
Application
Application
Operating System
Physical Machine


Machine
Machine
Machine
Network
Machine
Machine
Machine

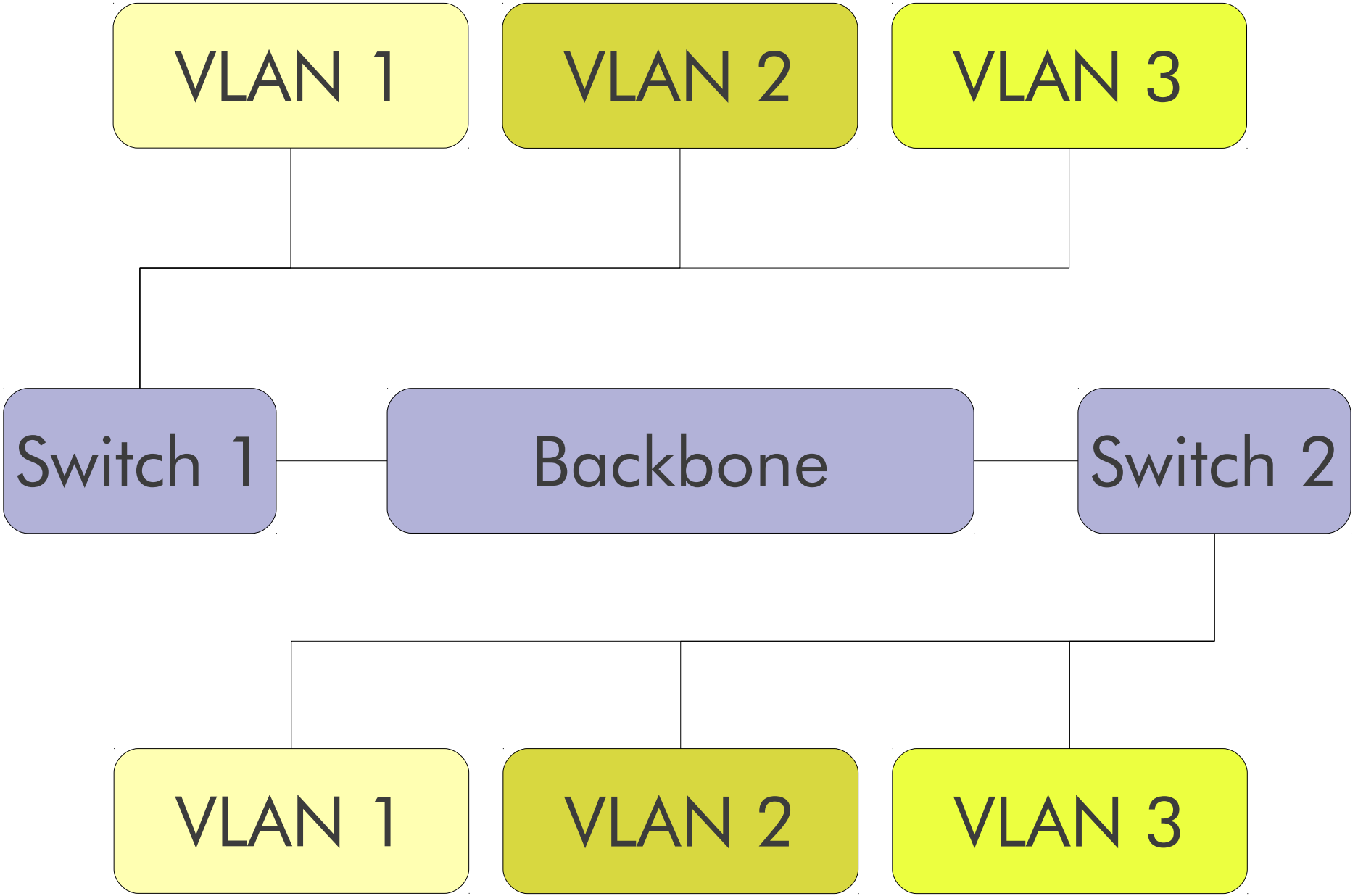

Machine
Machine
Machine
Storage


Network

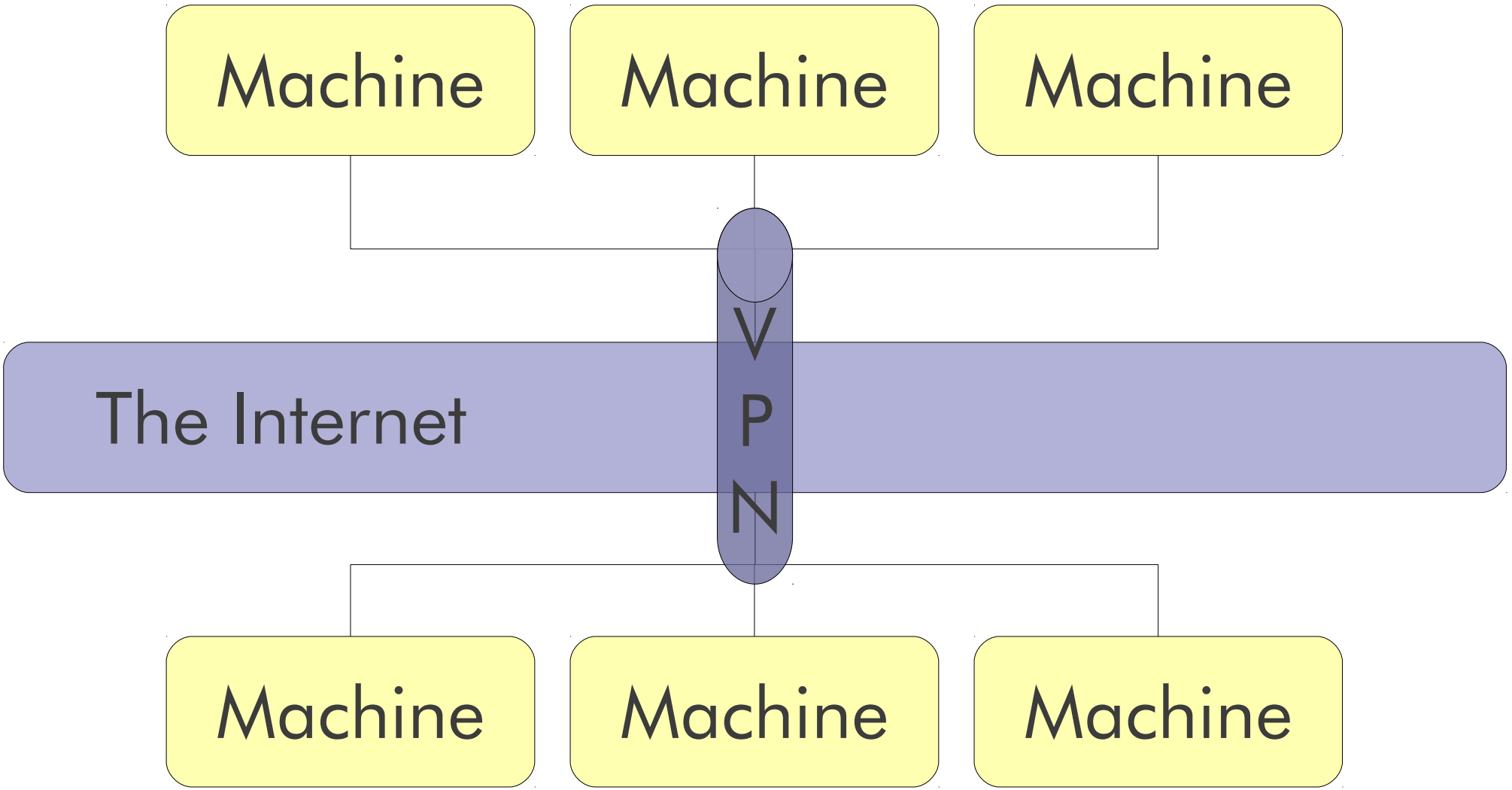
Network Address Translation



VLAN Trunking/Tagging

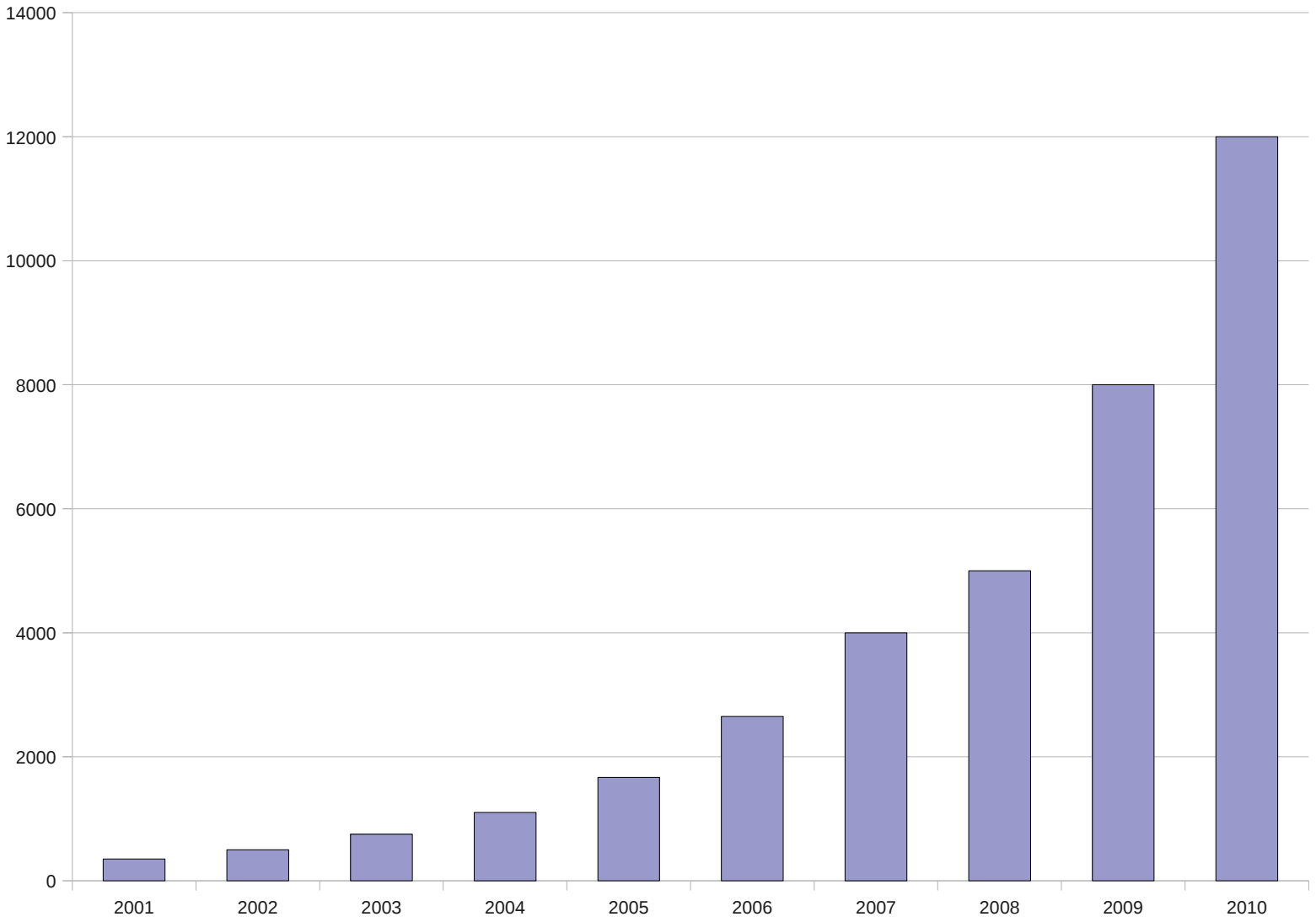


Virtual Private Network

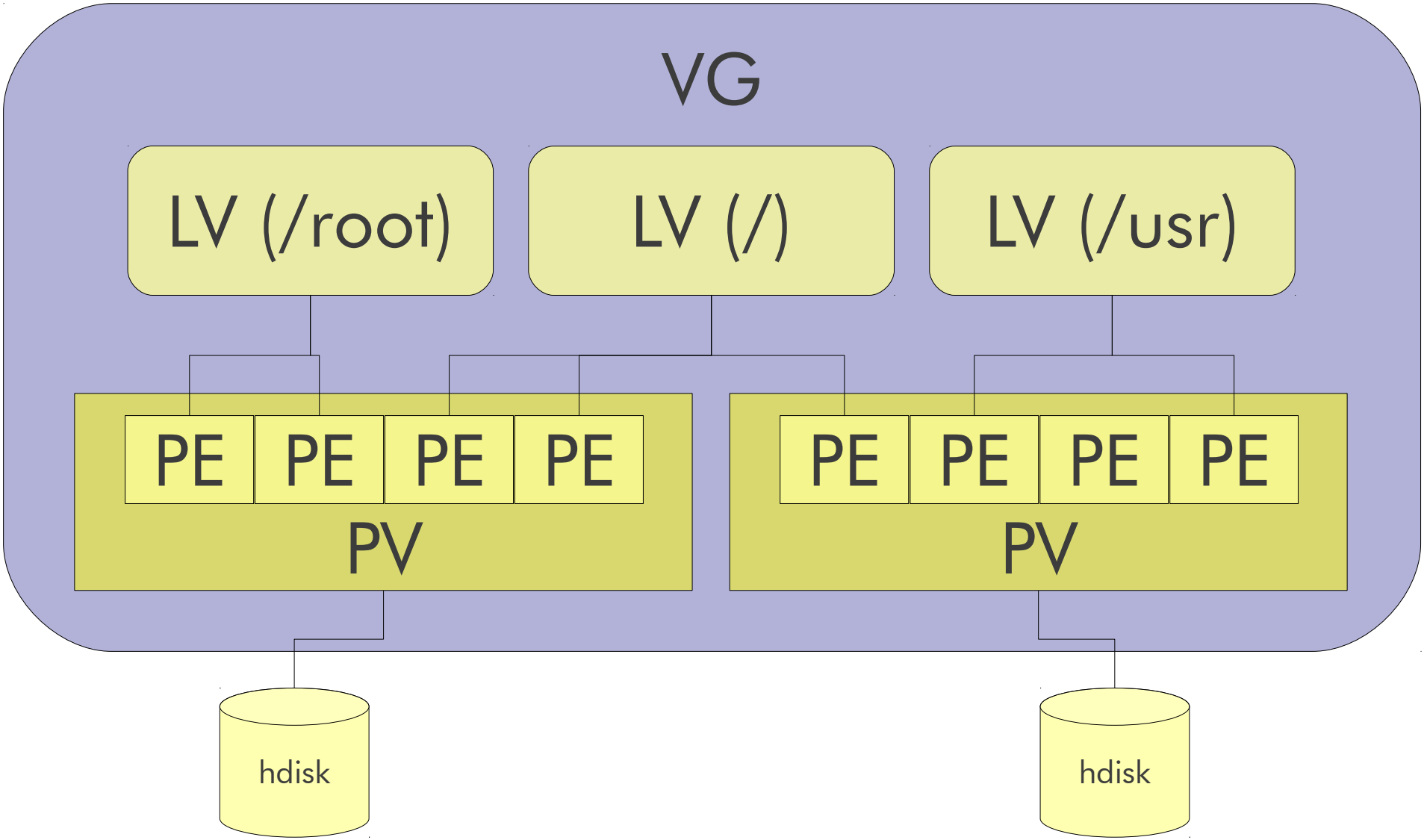


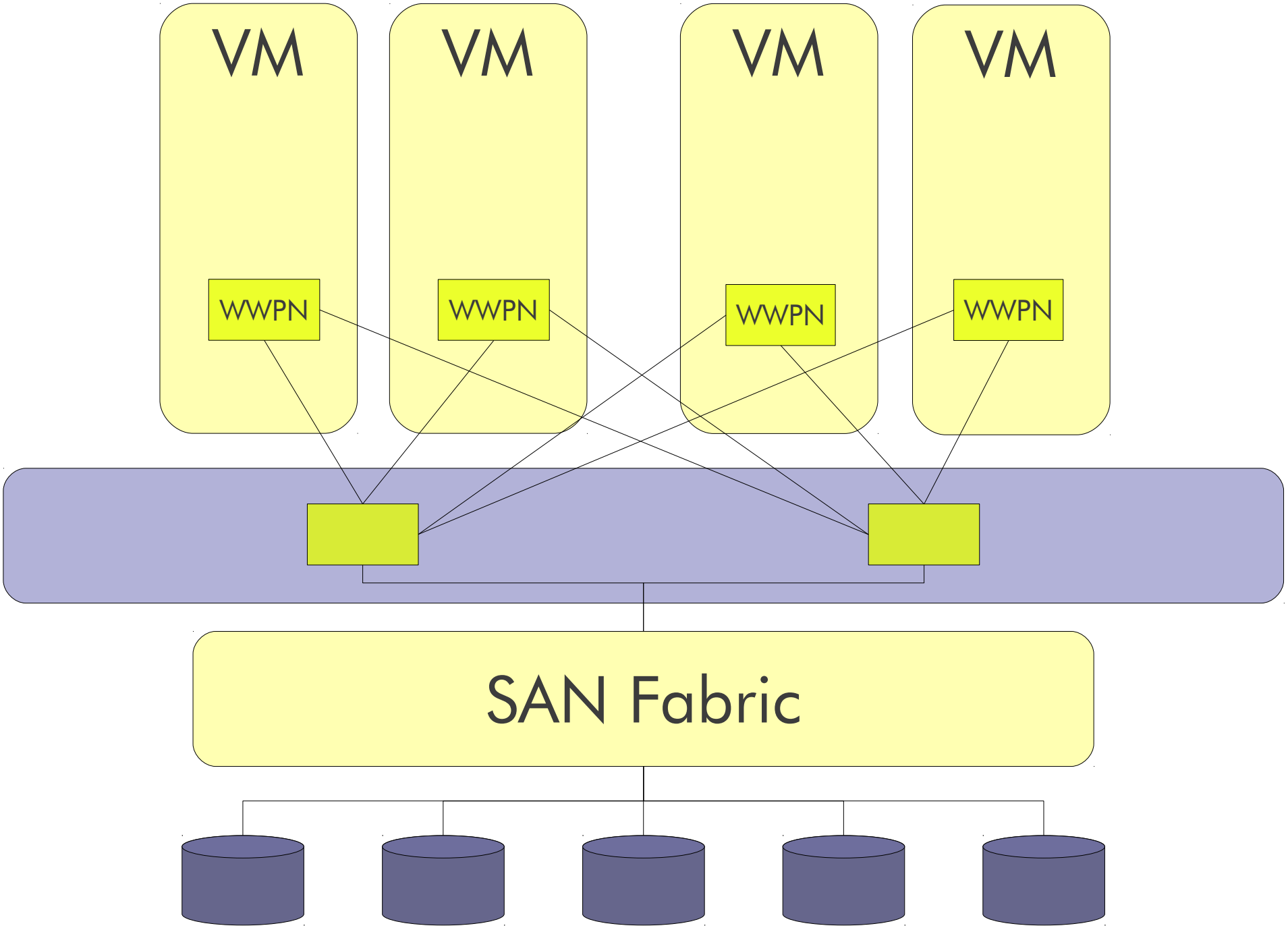

Storage

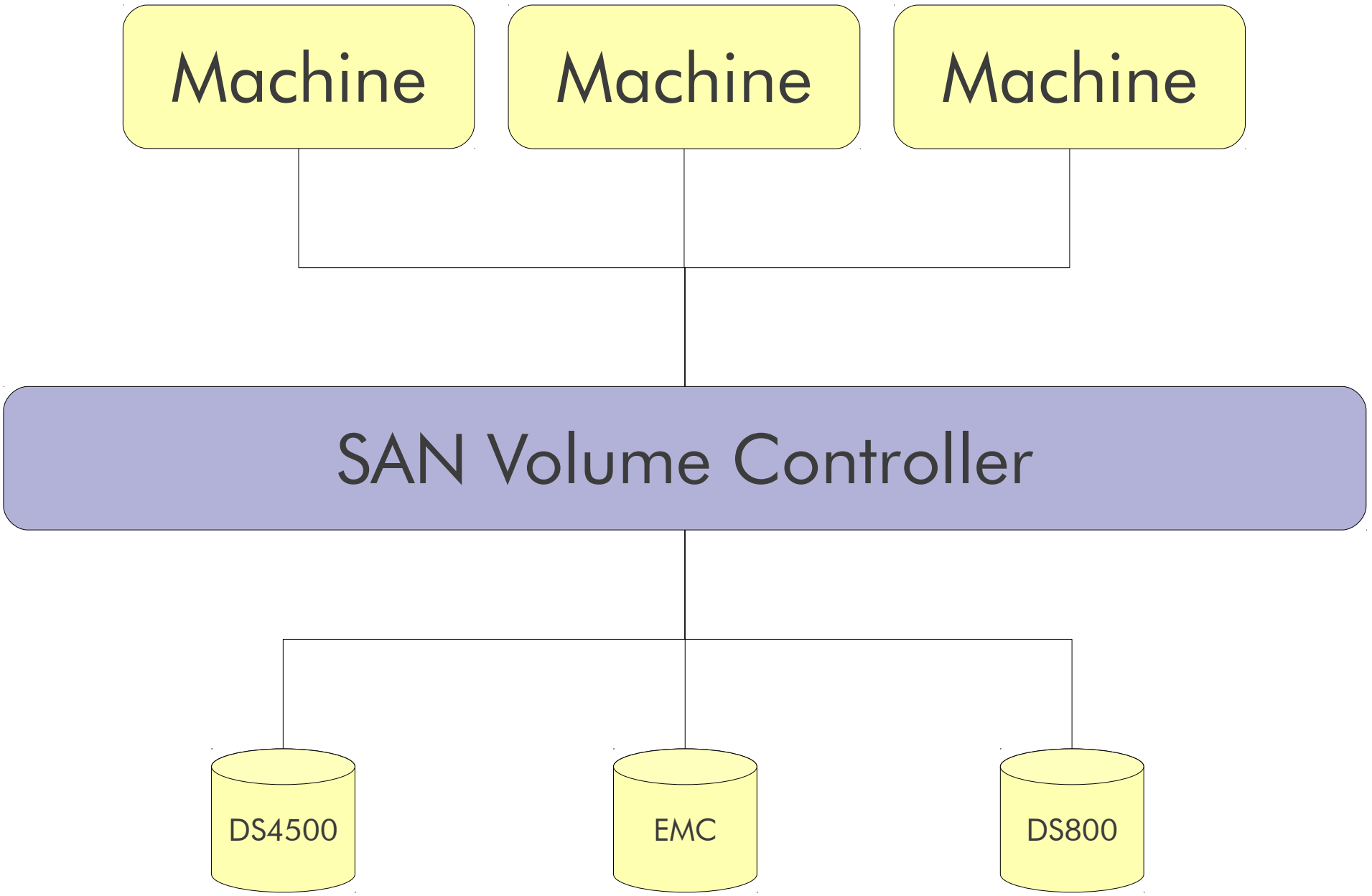
External Terabytes Worldwide



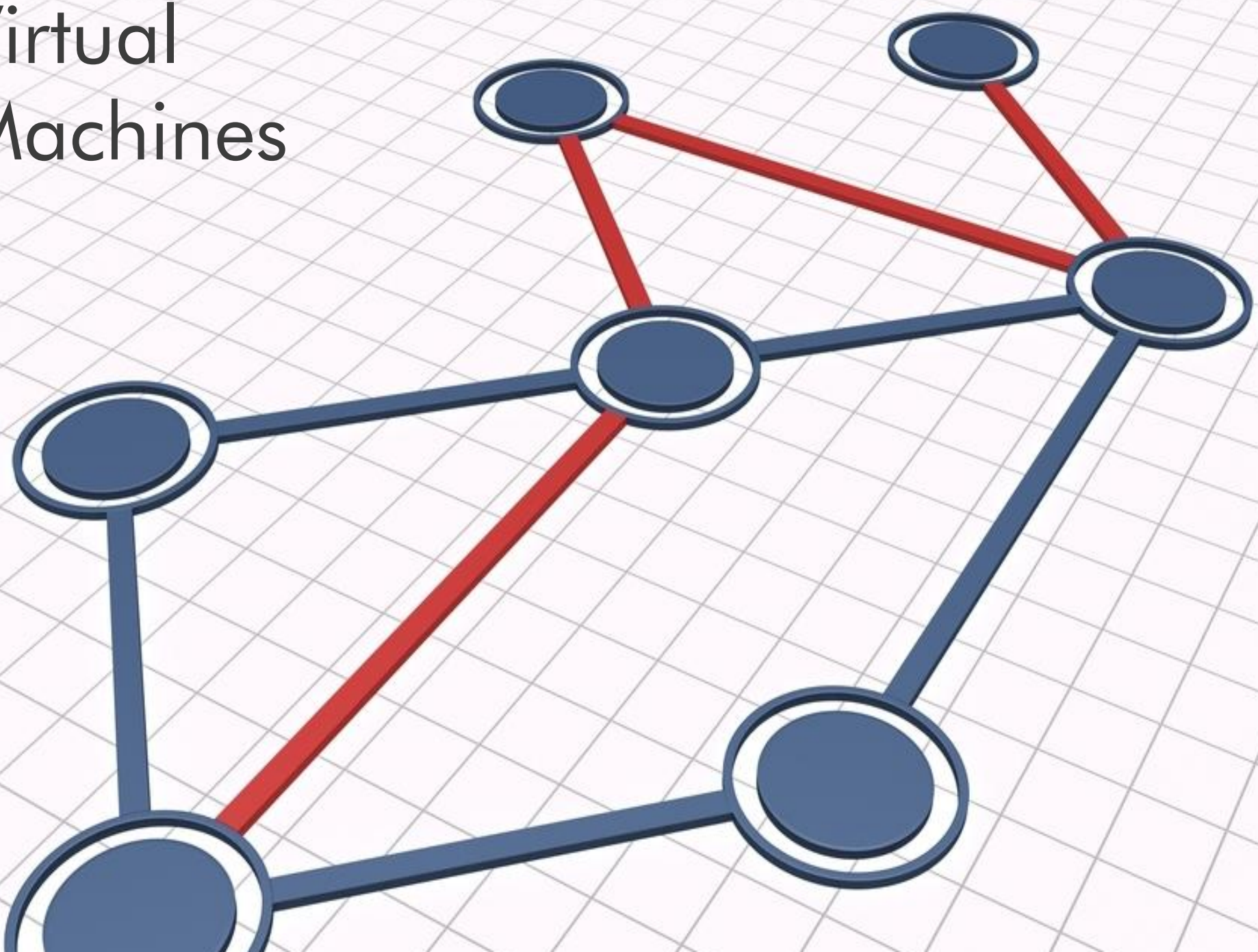
Logical Volume Management




VM
VM
VM
VM
WWPN
WWPN
WWPN
WWPN
SAN Fabric

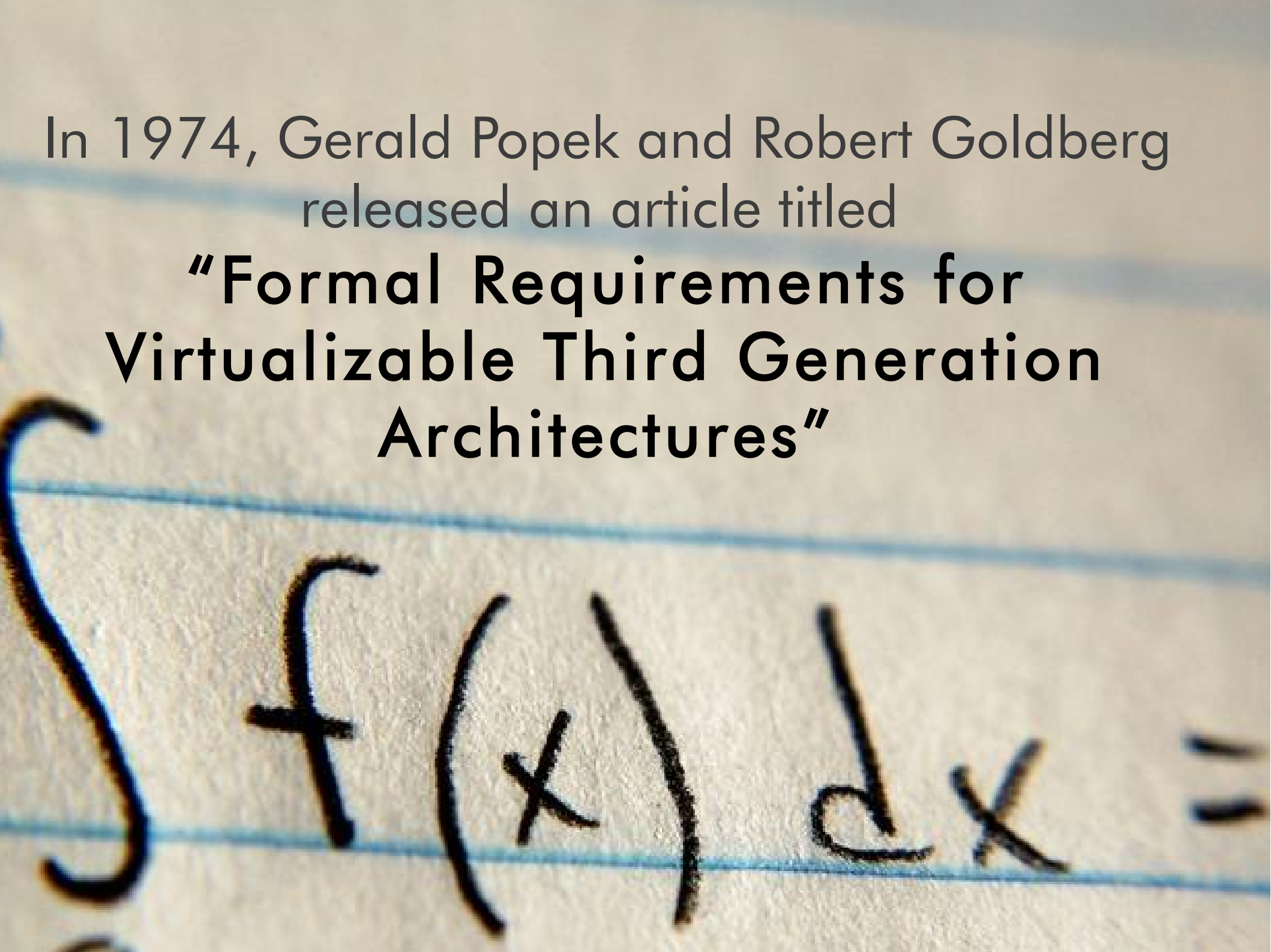

Machine
Machine
Machine
SAN Volume Controller
DS4500
EMC
DS800

Virtual Machines

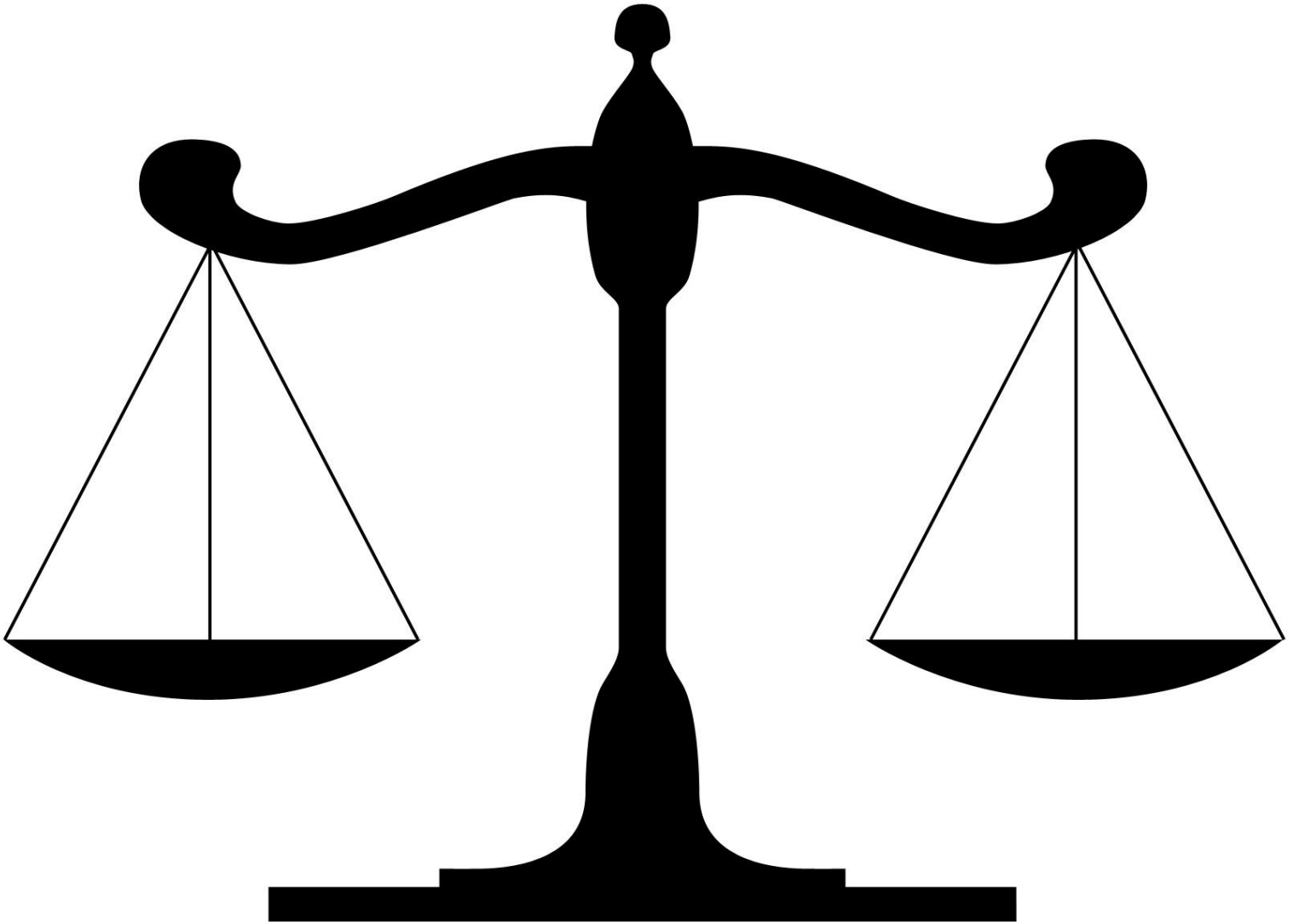


In 1974, Gerald Popek and Robert Goldberg released an article titled

"Formal Requirements for Virtualizable Third Generation Architectures"



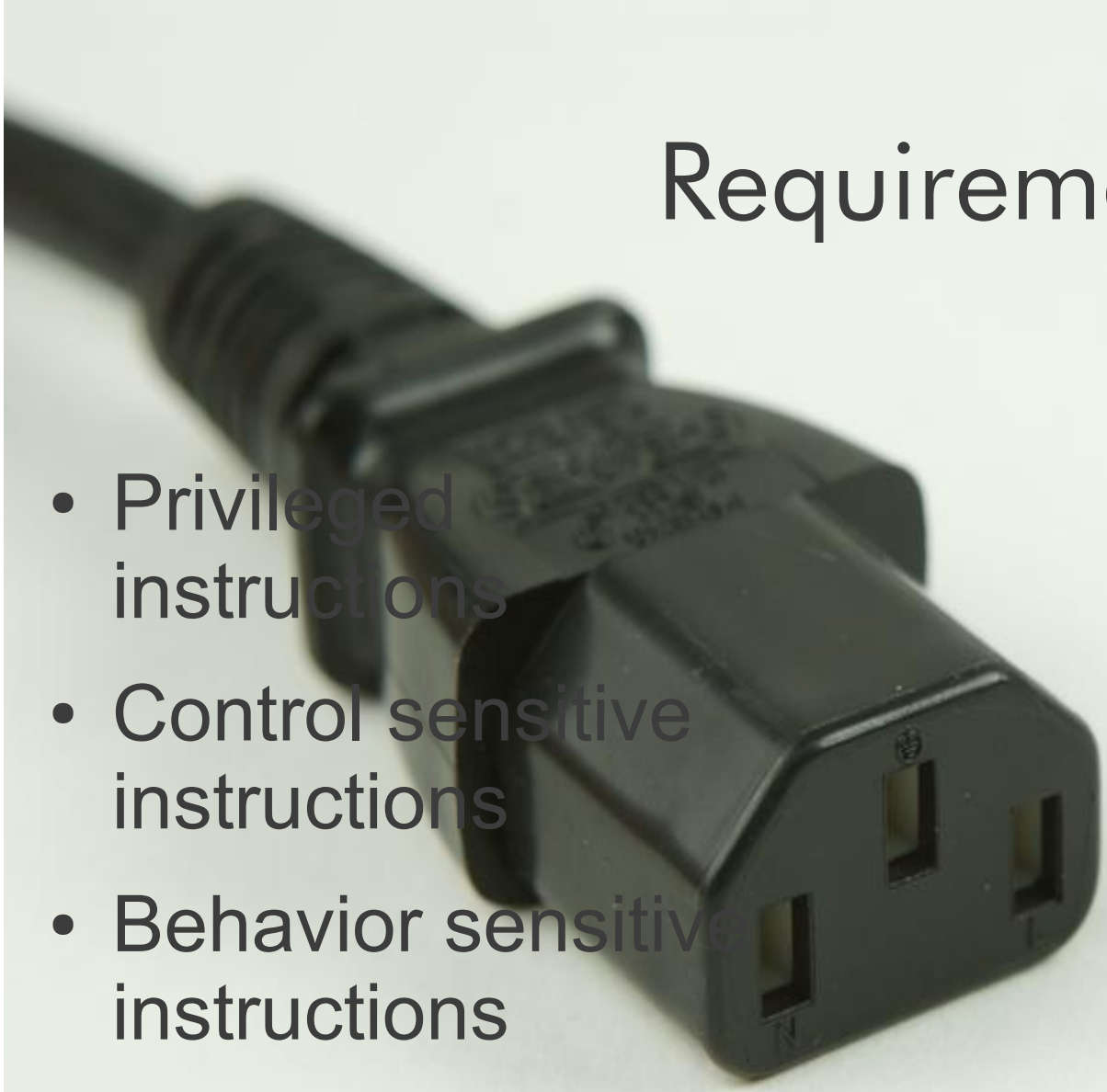
Equivalence



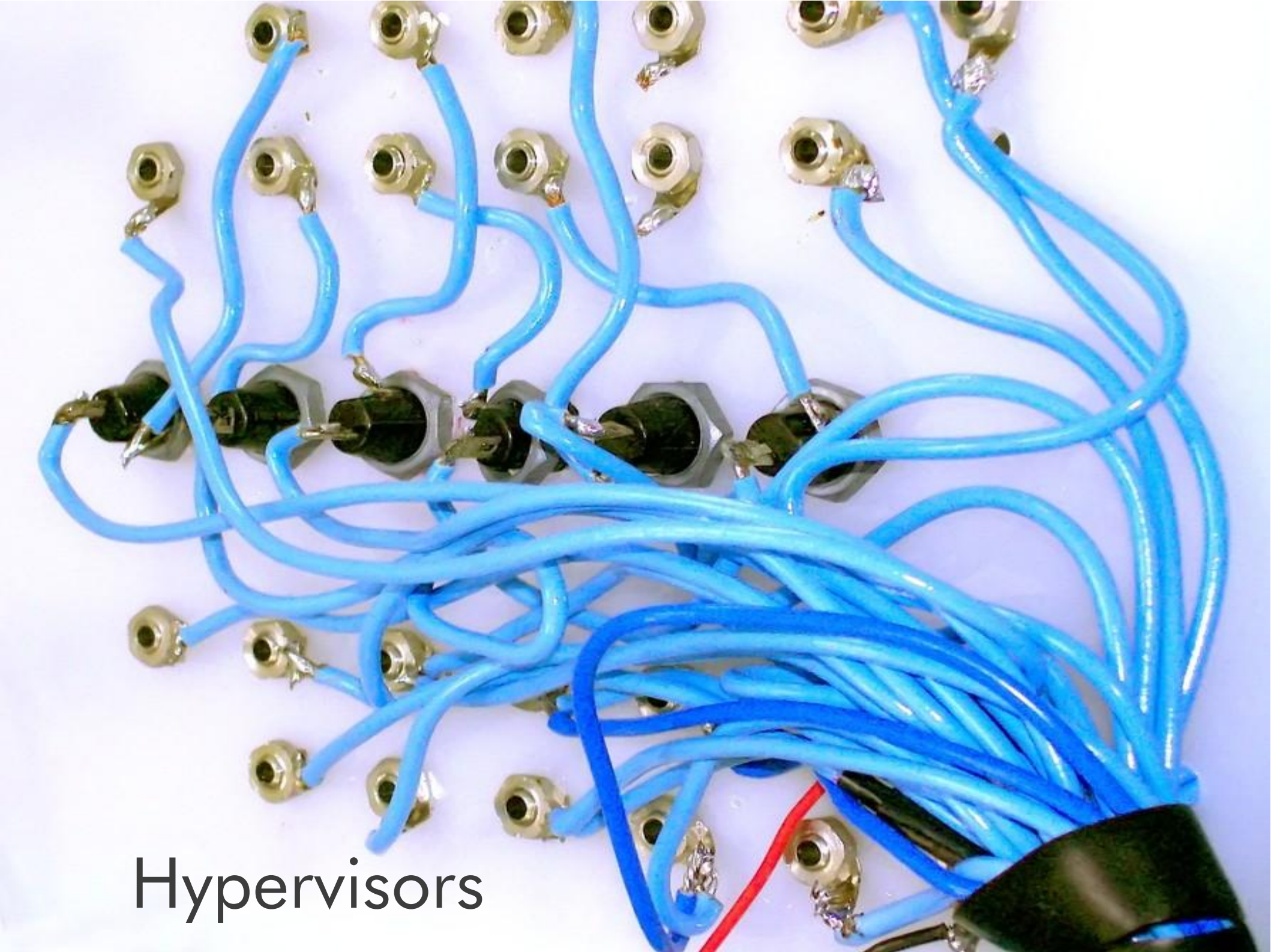

Resource
Control


Efficiency

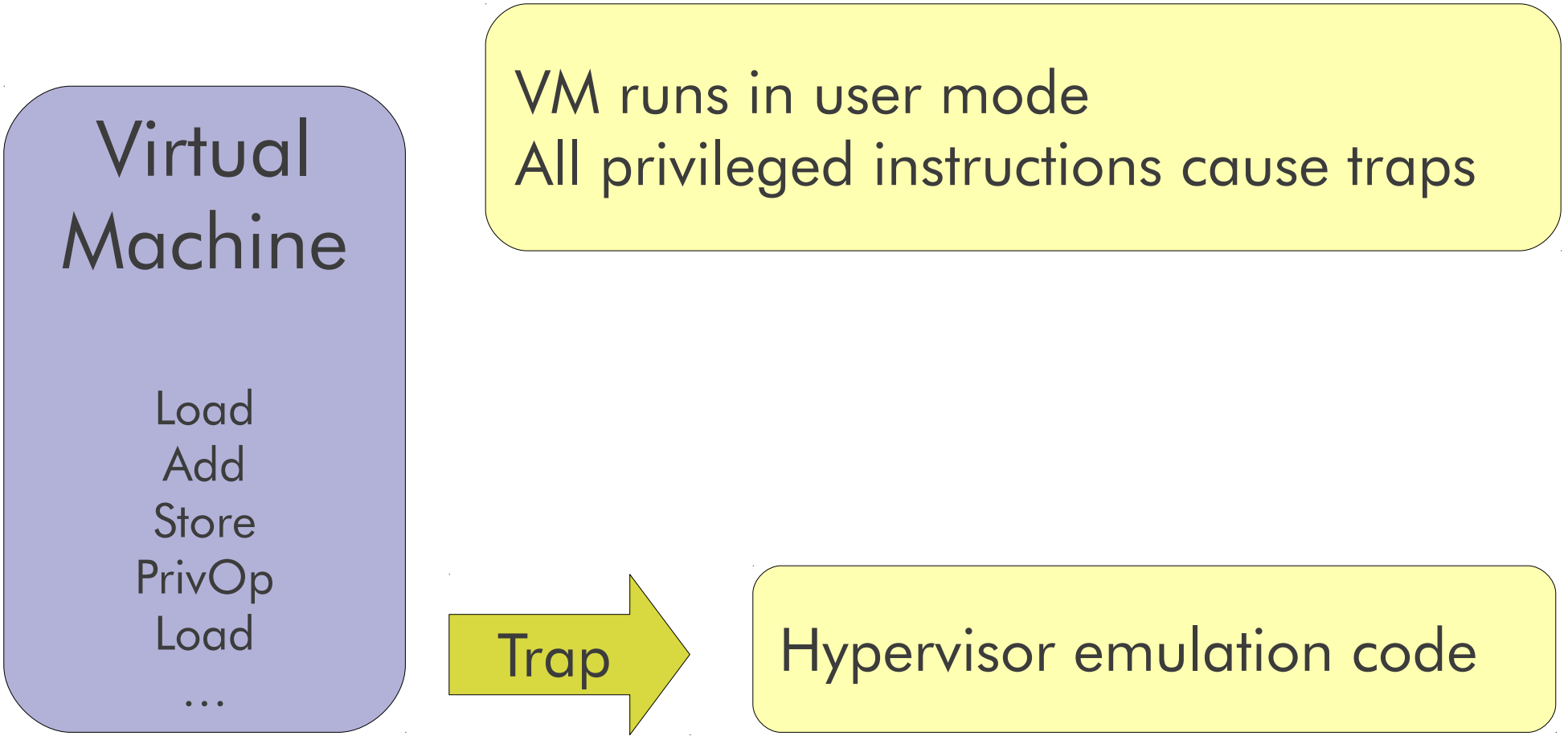
Requirements



- Privileged instructions
- Control sensitive instructions
- Behavior sensitive instructions


Hypervisors

Trap and Emulate

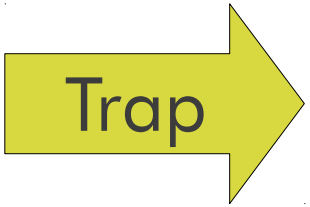


Translate, Trap and Emulate

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

Virtual Machine

Load
Add
Store
TrapOp
Load
…



Hypervisor PrivOP emulation code

Paravirtualization

VM runs in normal mode
OS in VM call hypervisor to access real resources

Virtual Machine

Load
Add
Store
Hcall
Load
...

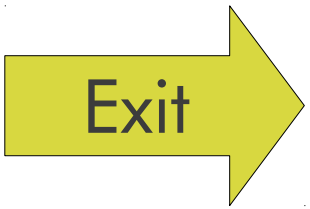
Hypervisor service

Direct Hardware Virtualization

VM runs in normal mode
Hardware provides the virtualization
Hypervisor provides control

Virtual Machine

Load
Add
Store
PrivOp
Load
…



Hypervisor service

Hypervisor calls

System x (Intel)

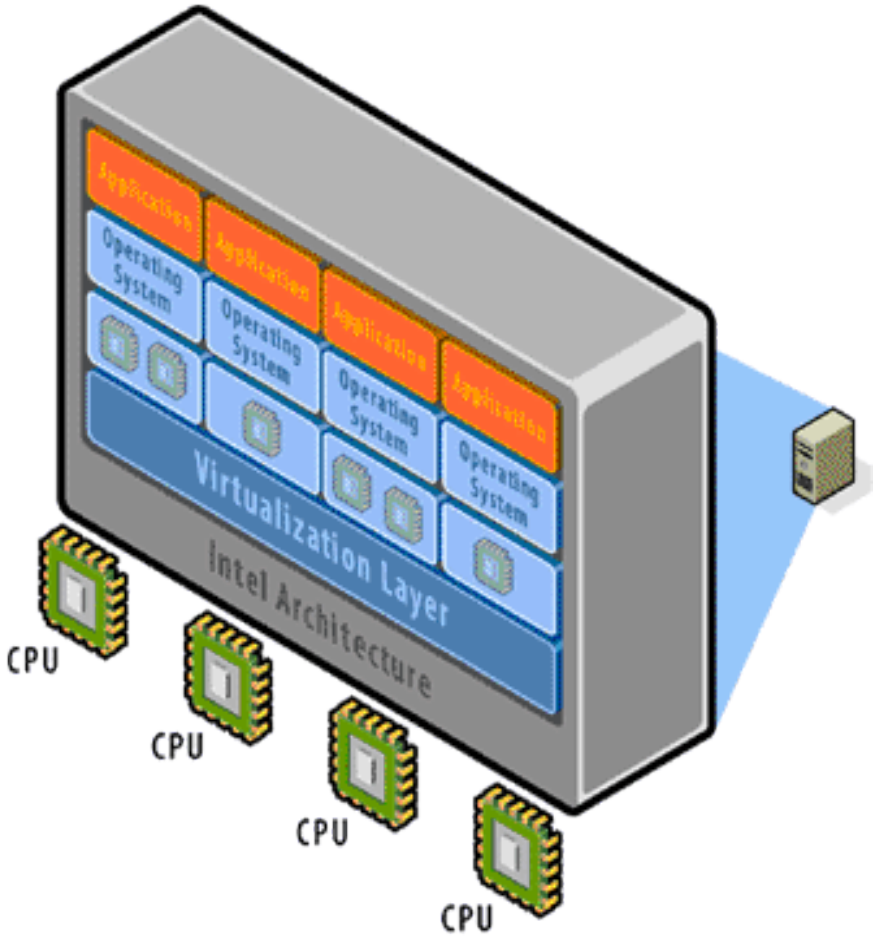
Virtualizing Intel

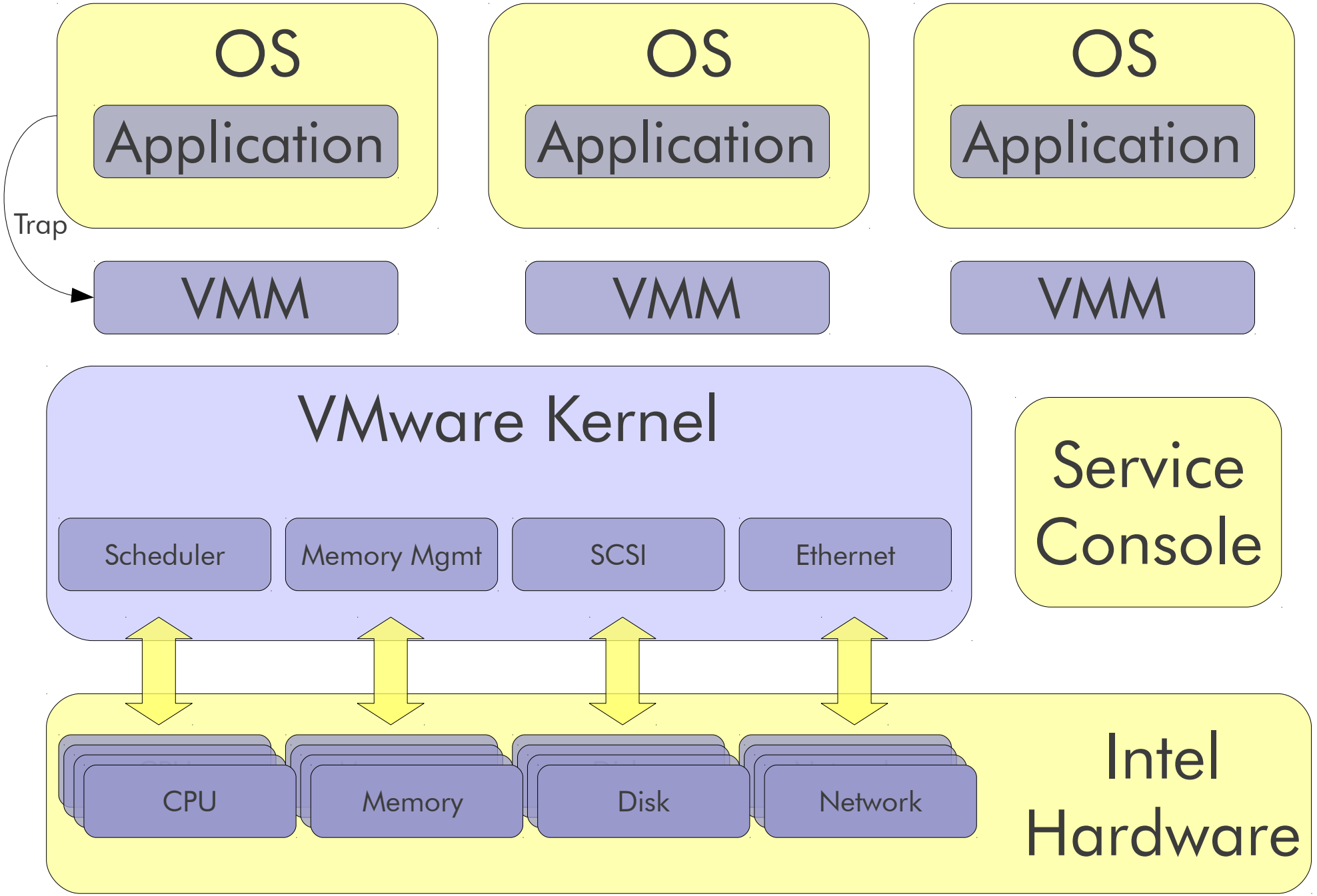
- The IA-32 was not designed to be virtualized
- Many protected instructions are not required to be executed in protected mode
- There are a great deal of devices which must be supported

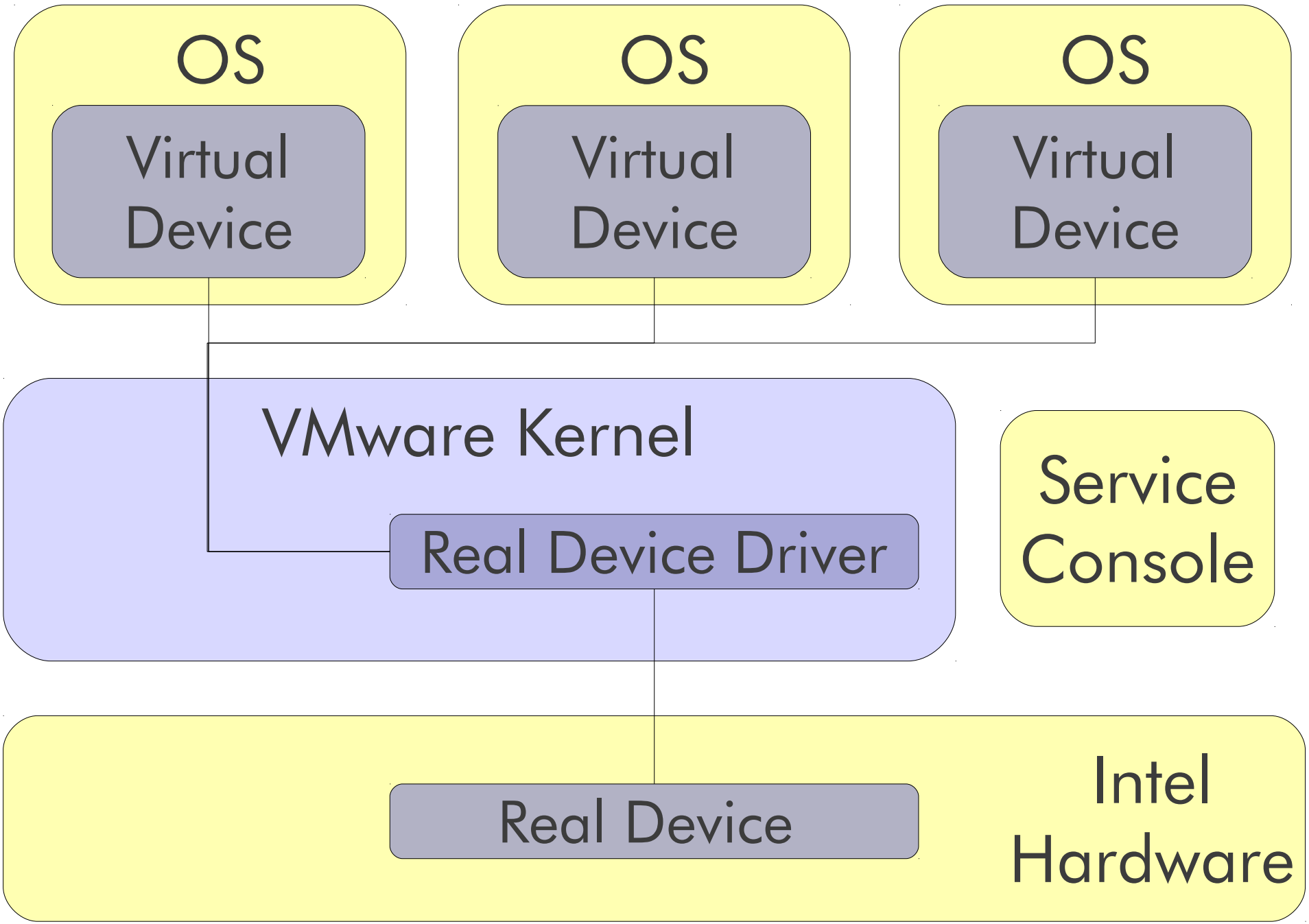

vmware®

VMware Virtualization

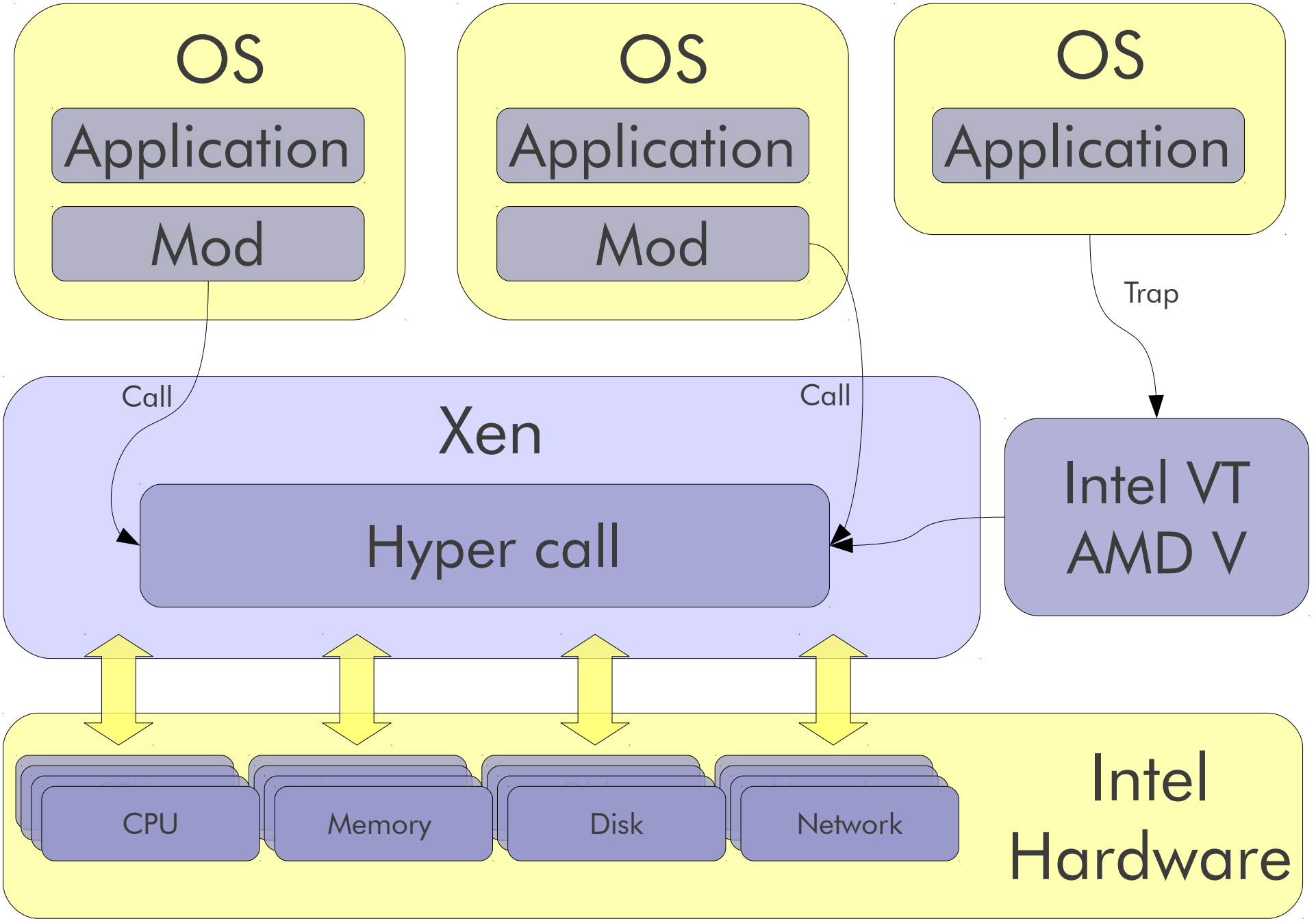
- CPU: Direct Execution w/ Binary Translation
- MEM: Shadow Table w/ Ballooning Driver
- I/O: Hosted Architecture or Limited Support

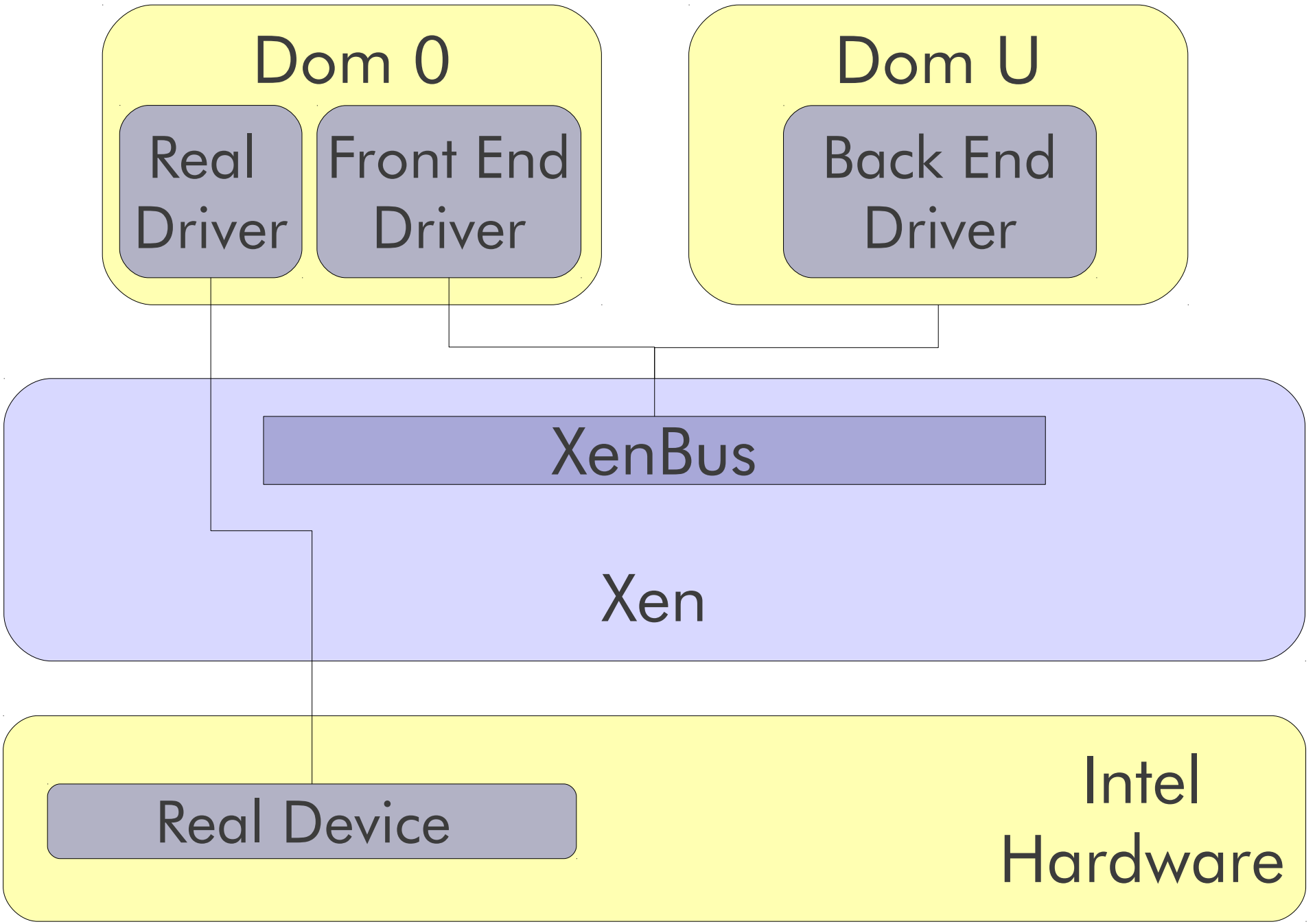



OS
Application
OS
Application
OS
Application
Trap
VMM
VMM
VMM
VMware Kernel
Scheduler
Memory Mgmt
SCSI
Ethernet
Service Console
CPU
Memory
Disk
Network
Intel Hardware

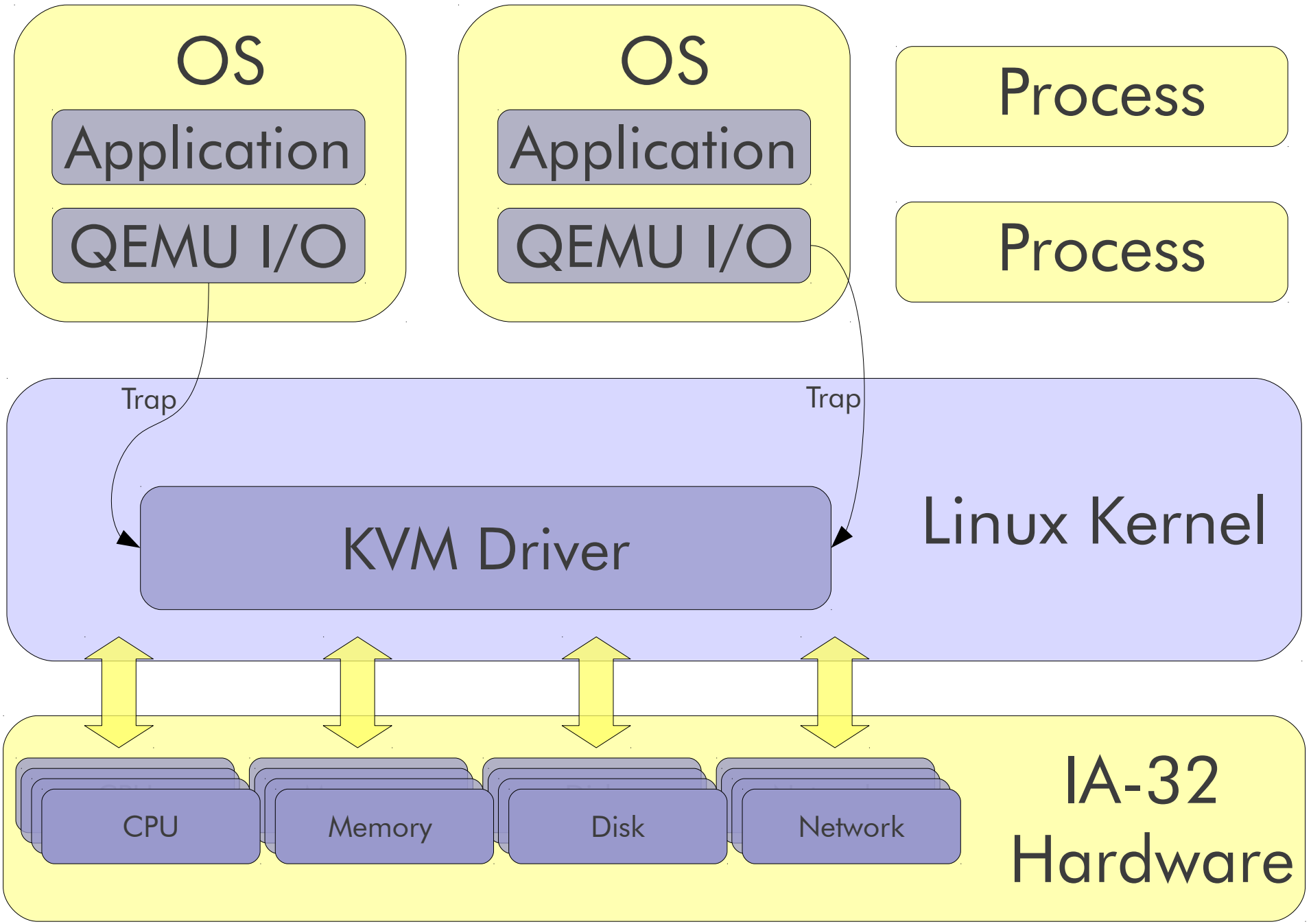

OS
Virtual Device
OS
Virtual Device
OS
Virtual Device
VMware Kernel
Real Device Driver
Service Console
Real Device
Intel Hardware

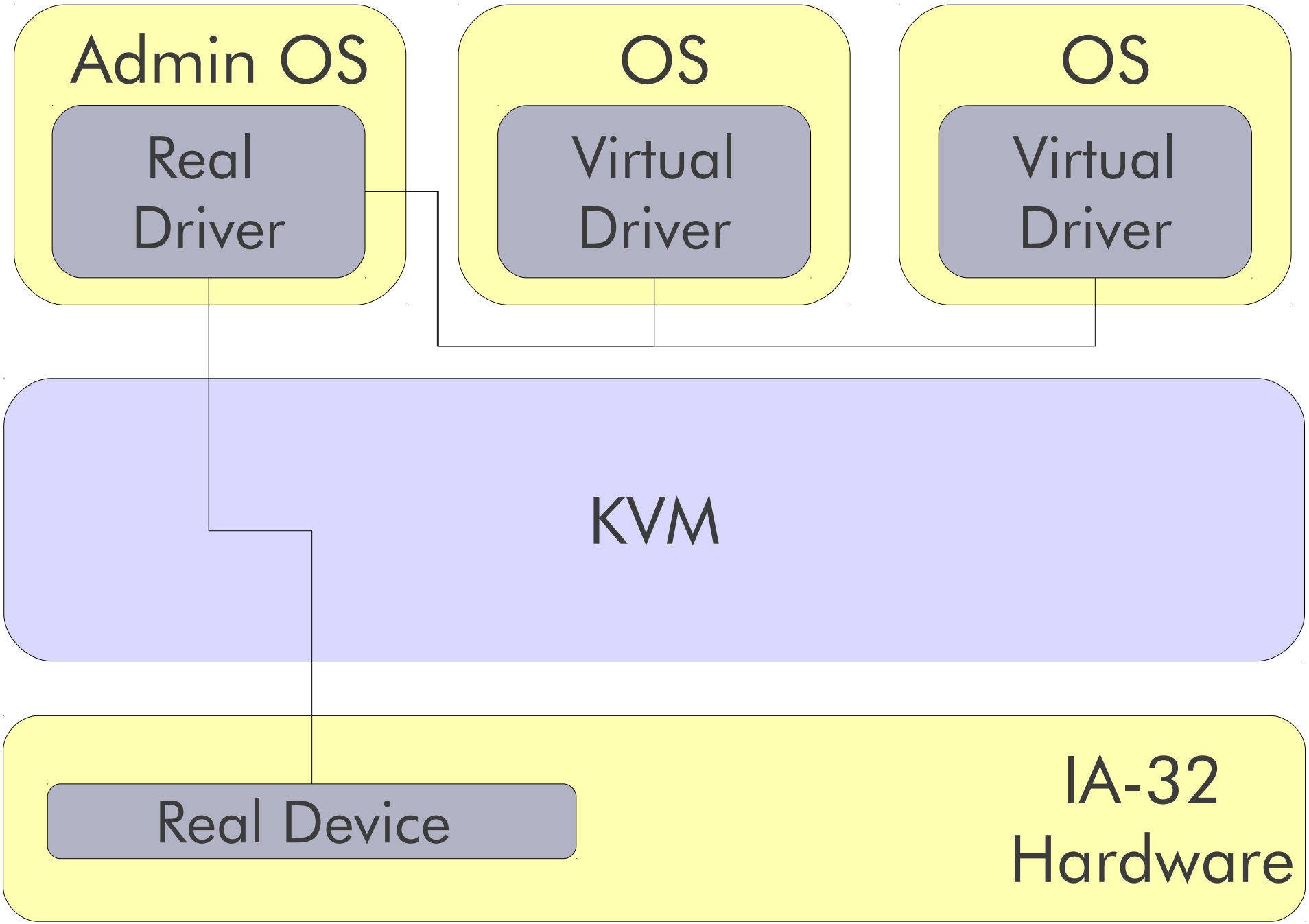

Xen
Source™


OS
Application
Mod
OS
Application
Mod
OS
Application
Trap
Call
Call
Xen
Hyper call
Intel VT
AMD V
CPU
Memory
Disk
Network
Intel
Hardware


Dom 0
Real Driver
Front End Driver
Dom U
Back End Driver
XenBus
Xen
Real Device
Intel Hardware

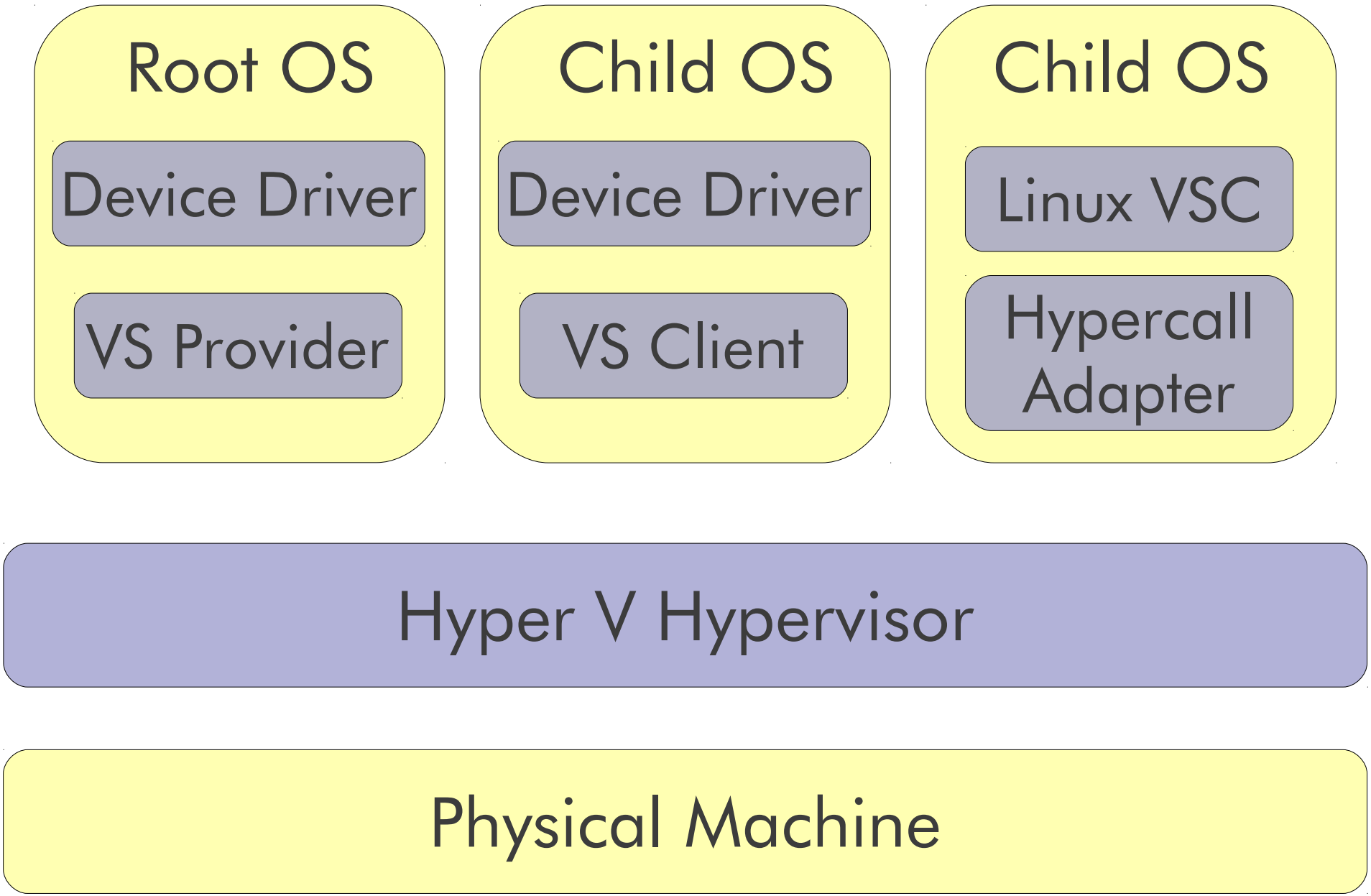

KVM

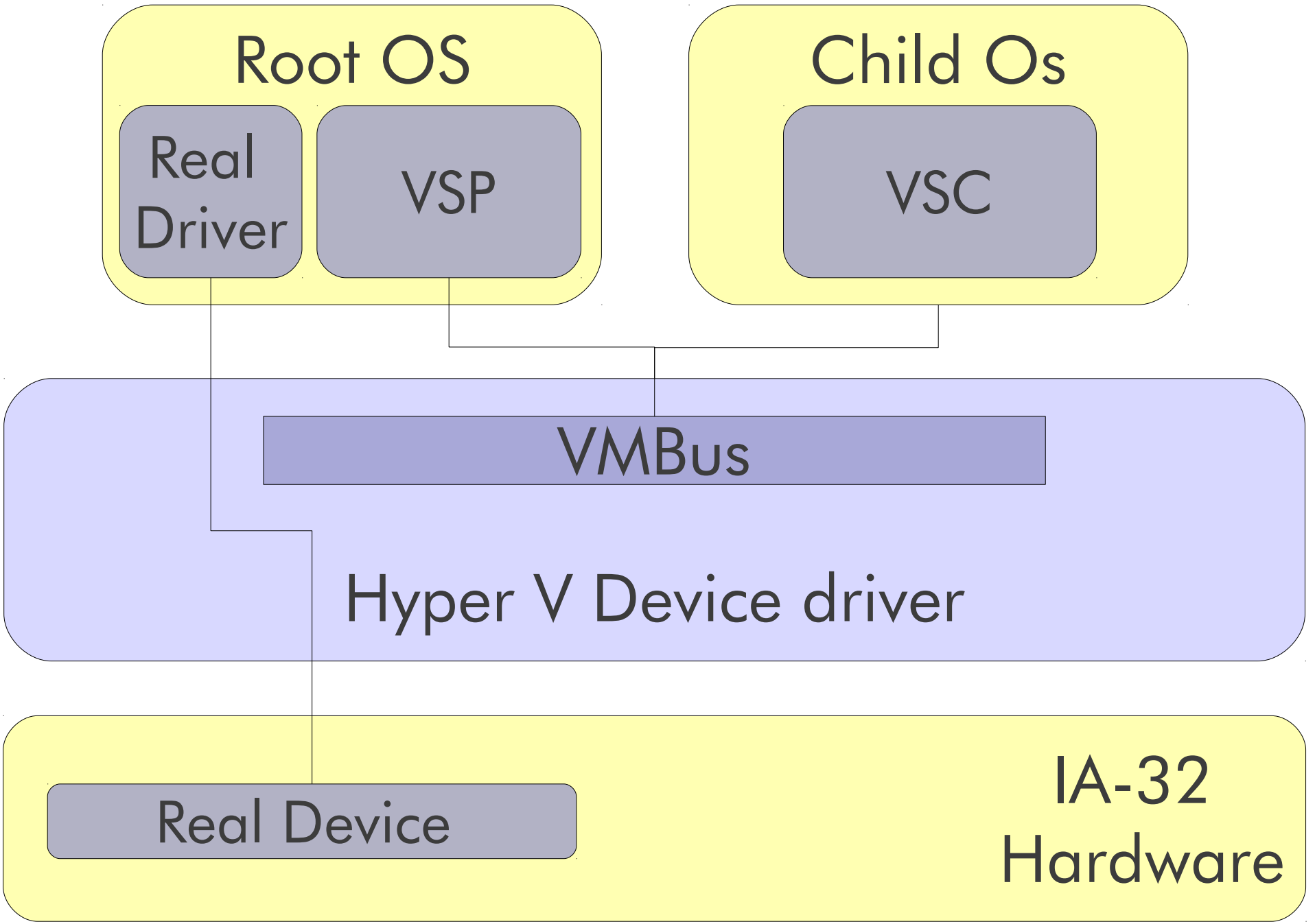

OS
Application
QEMU I/O
OS
Application
QEMU I/O
Process
Process
Trap
Trap
KVM Driver
Linux Kernel
CPU
Memory
Disk
Network
IA-32
Hardware


Admin OS
Real Driver
OS
Virtual Driver
OS
Virtual Driver
KVM
Real Device
IA-32 Hardware

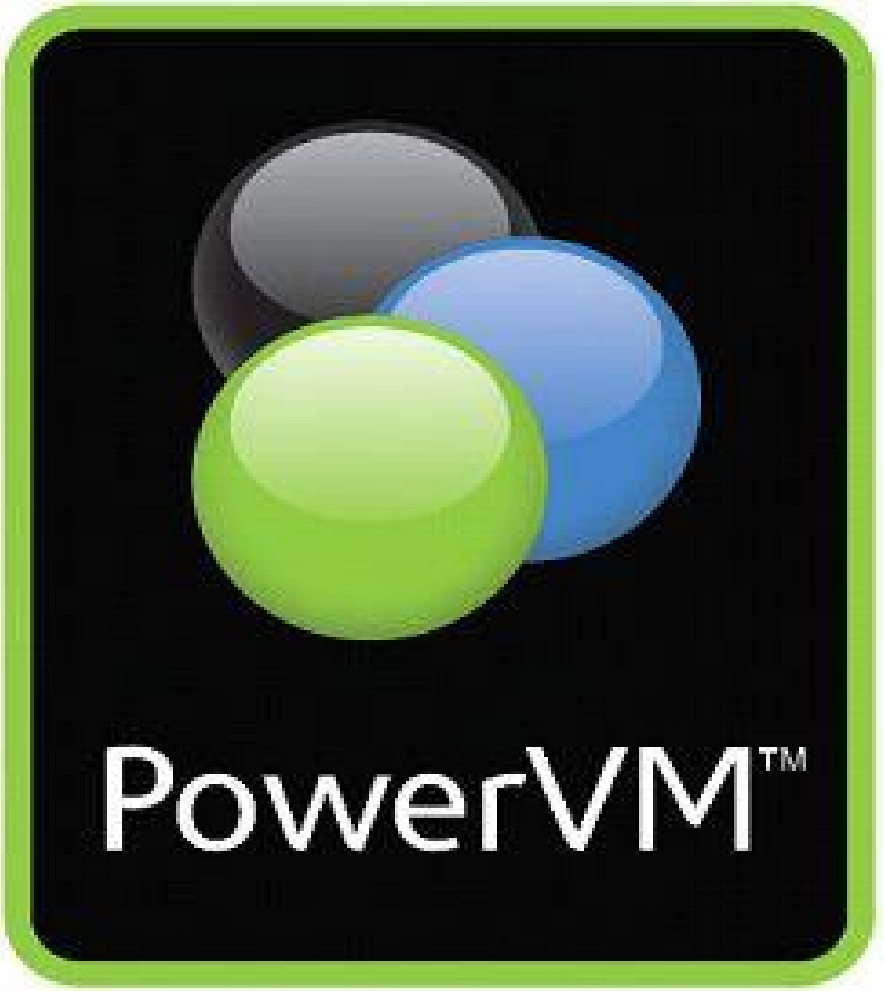
Windows Server® 2008

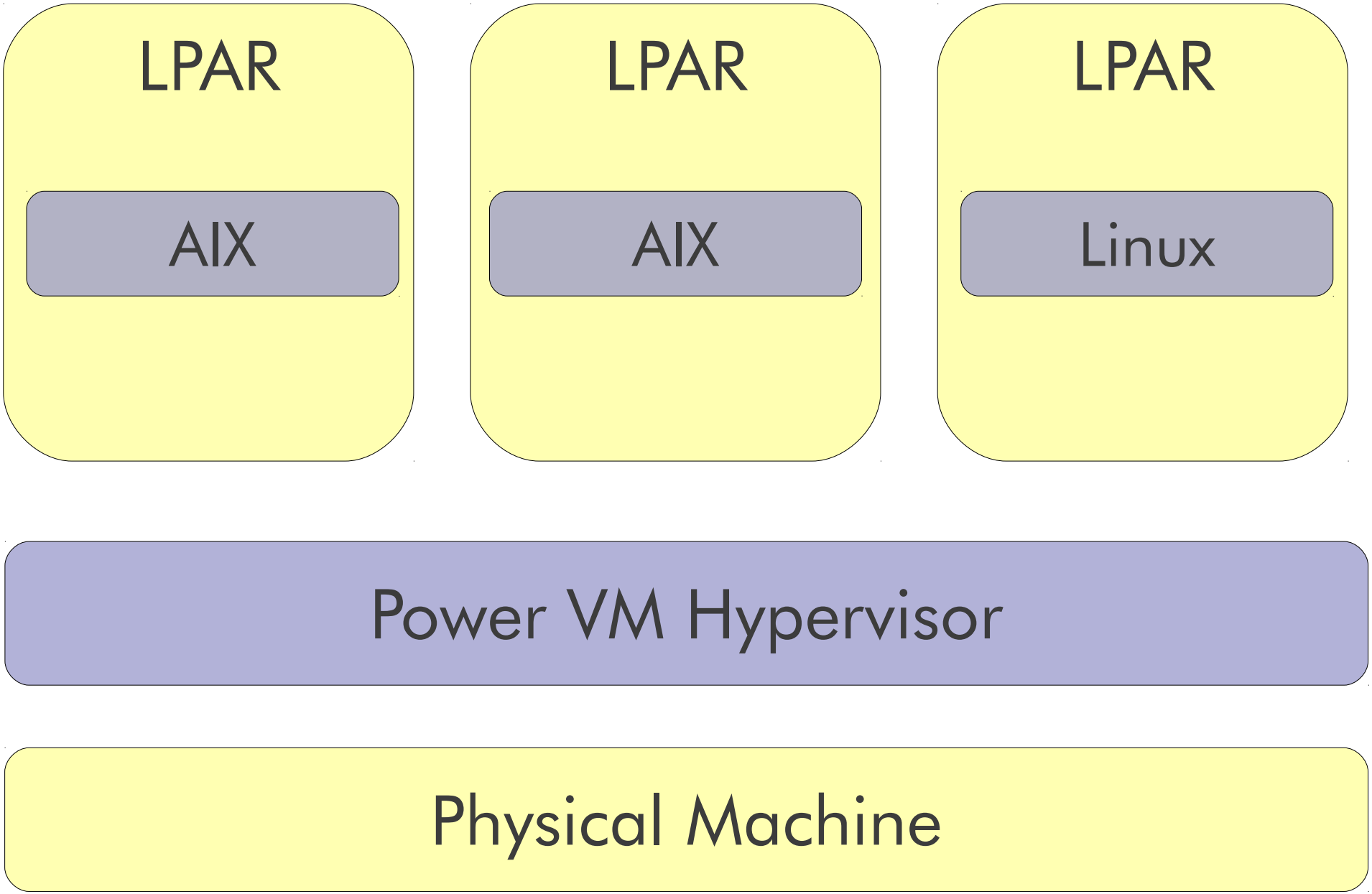
Hyper-V™

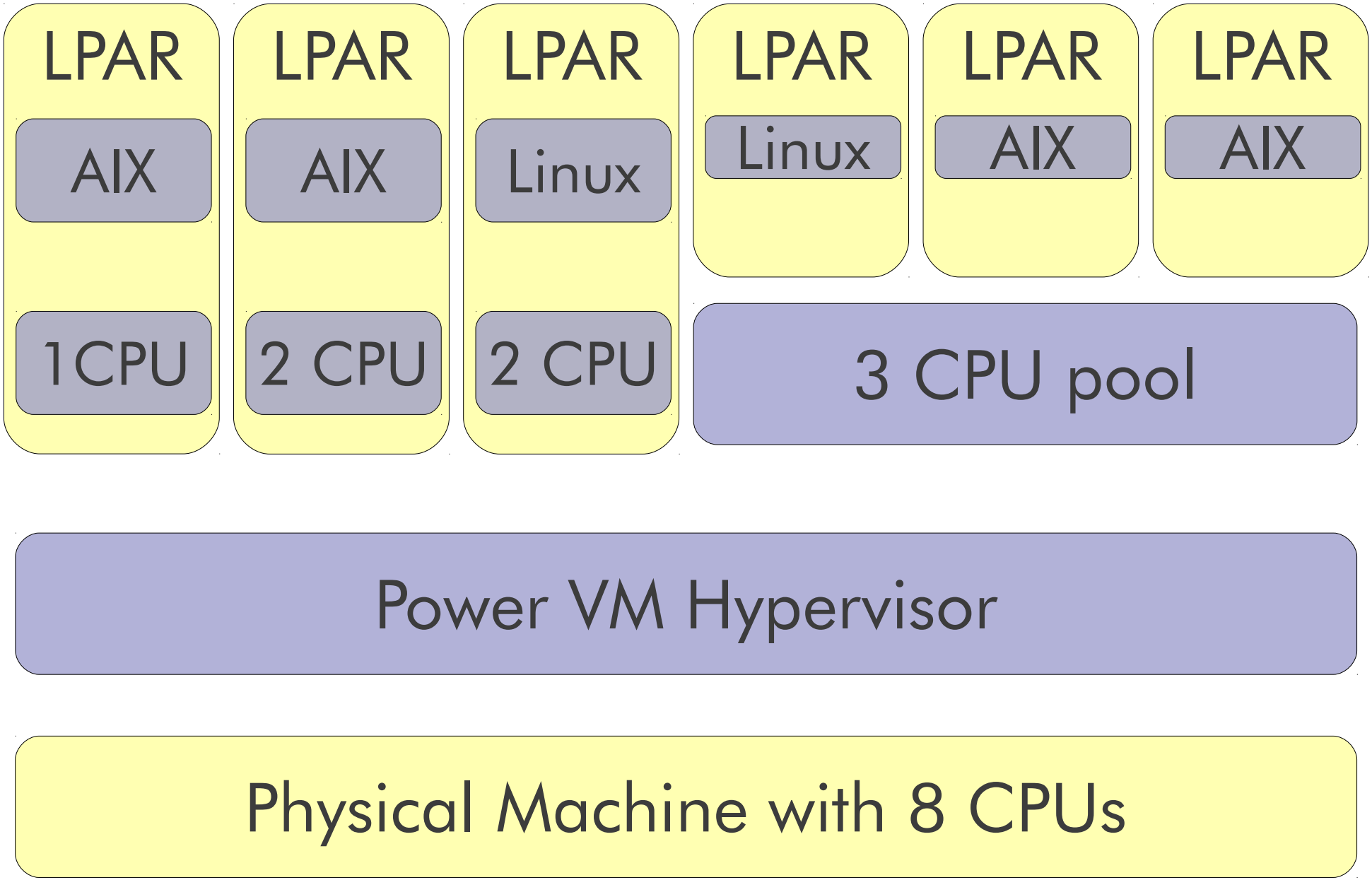

Root OS
Device Driver
VS Provider
Child OS
Device Driver
VS Client
Child OS
Linux VSC
Hypercall Adapter
Hyper V Hypervisor
Physical Machine

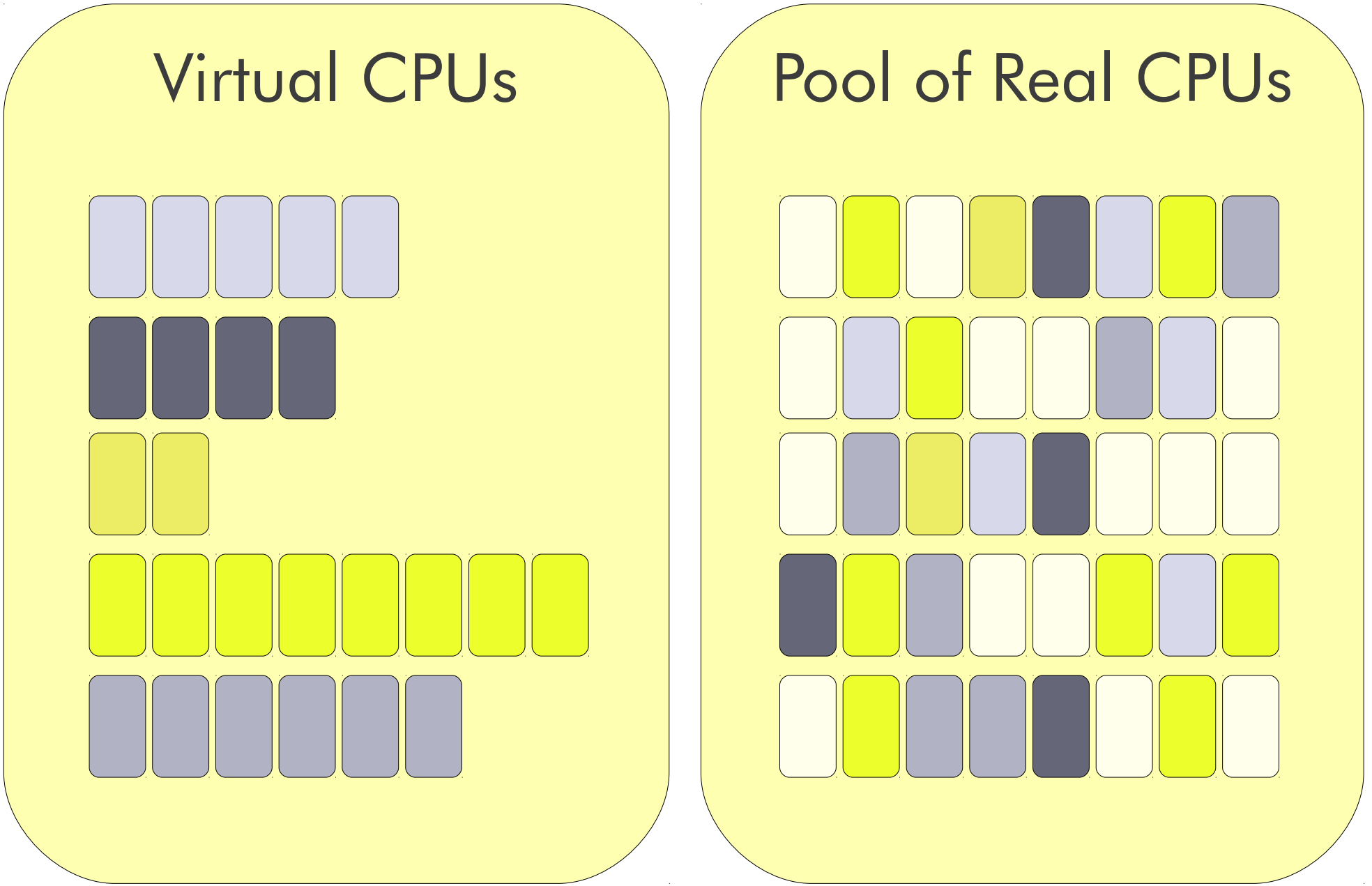

Root OS
Real Driver
VSP
Child Os
VSC
VMBus
Hyper V Device driver
Real Device
IA-32 Hardware

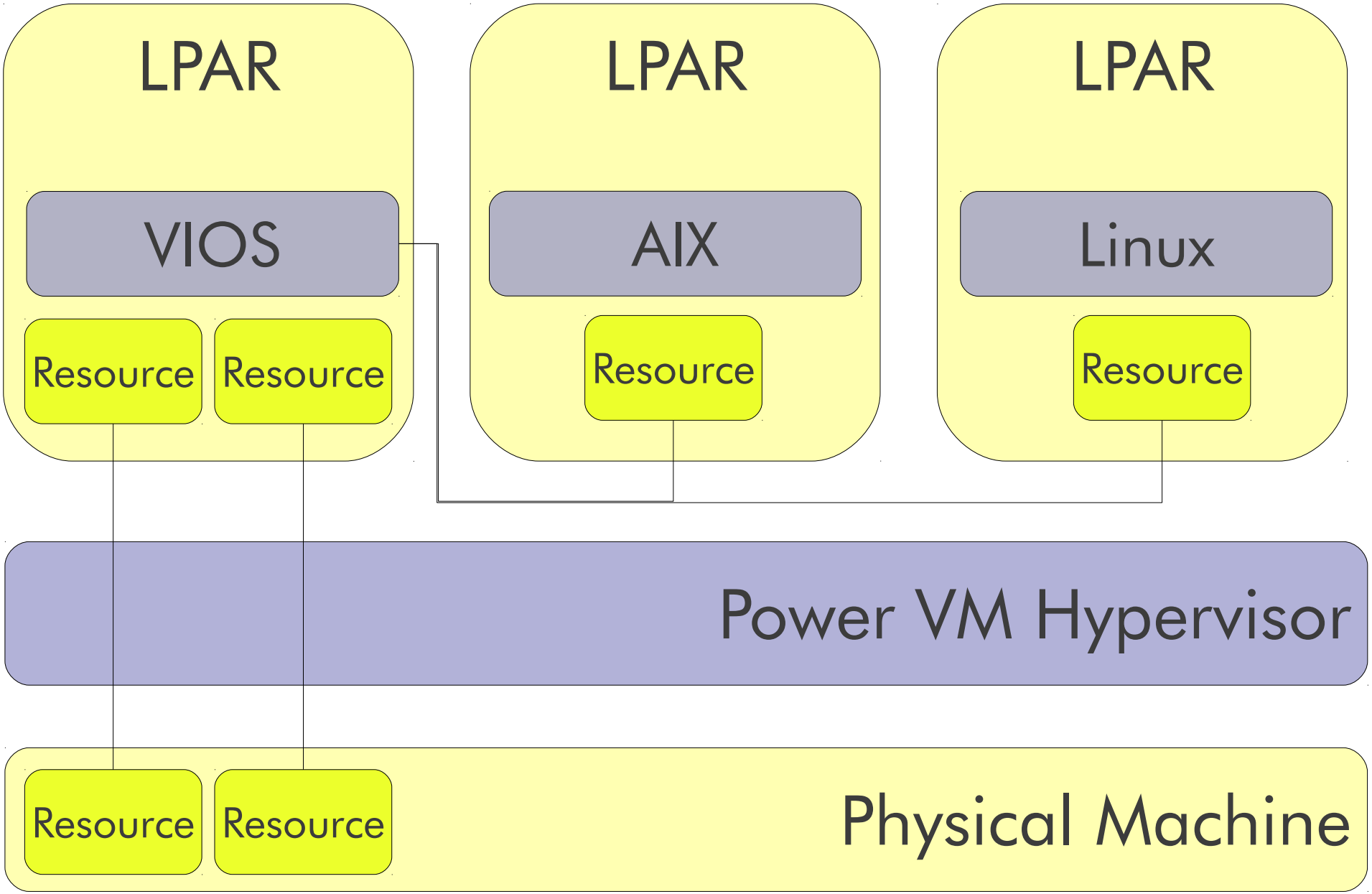
System p

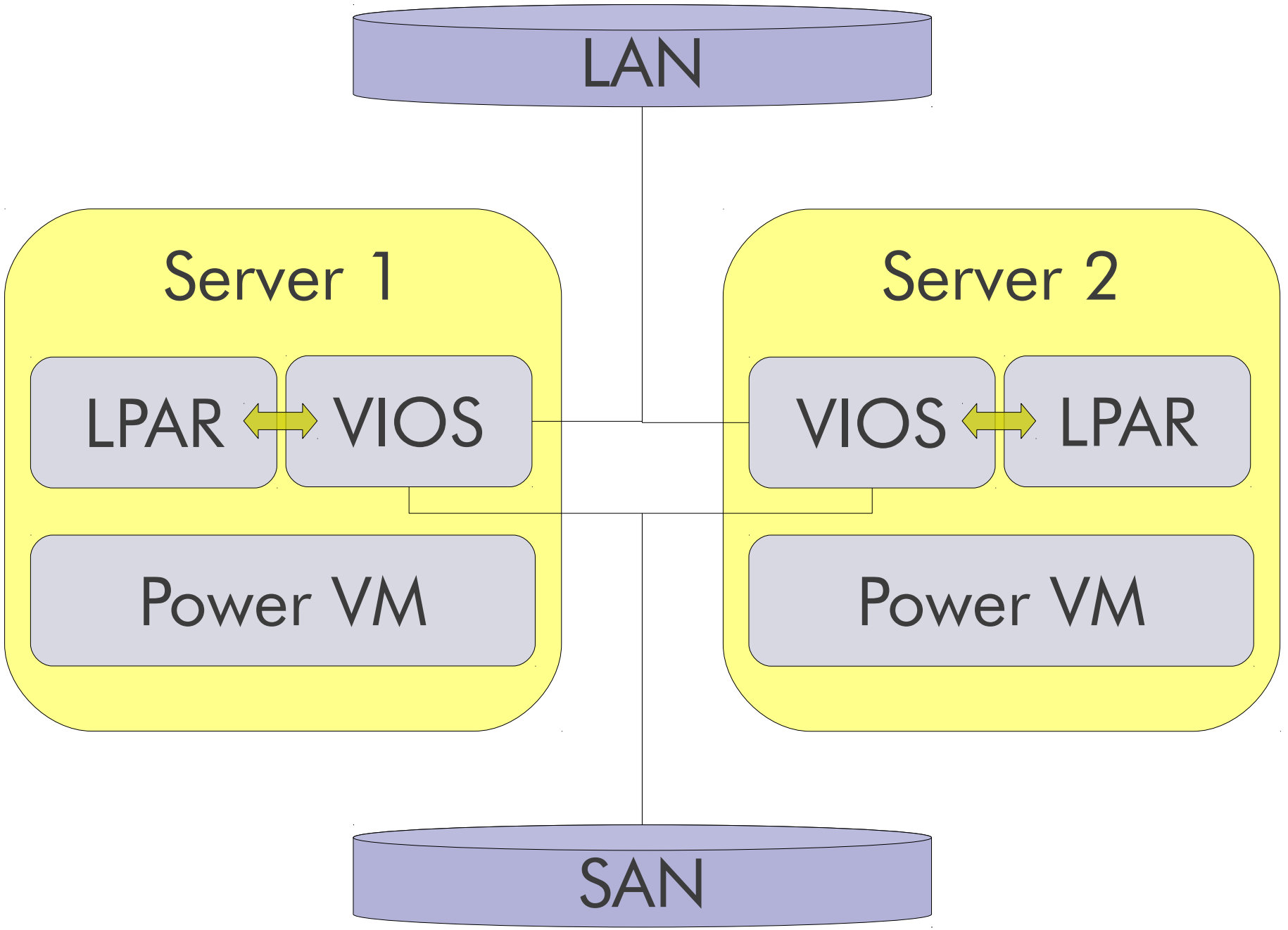

PowerVM™

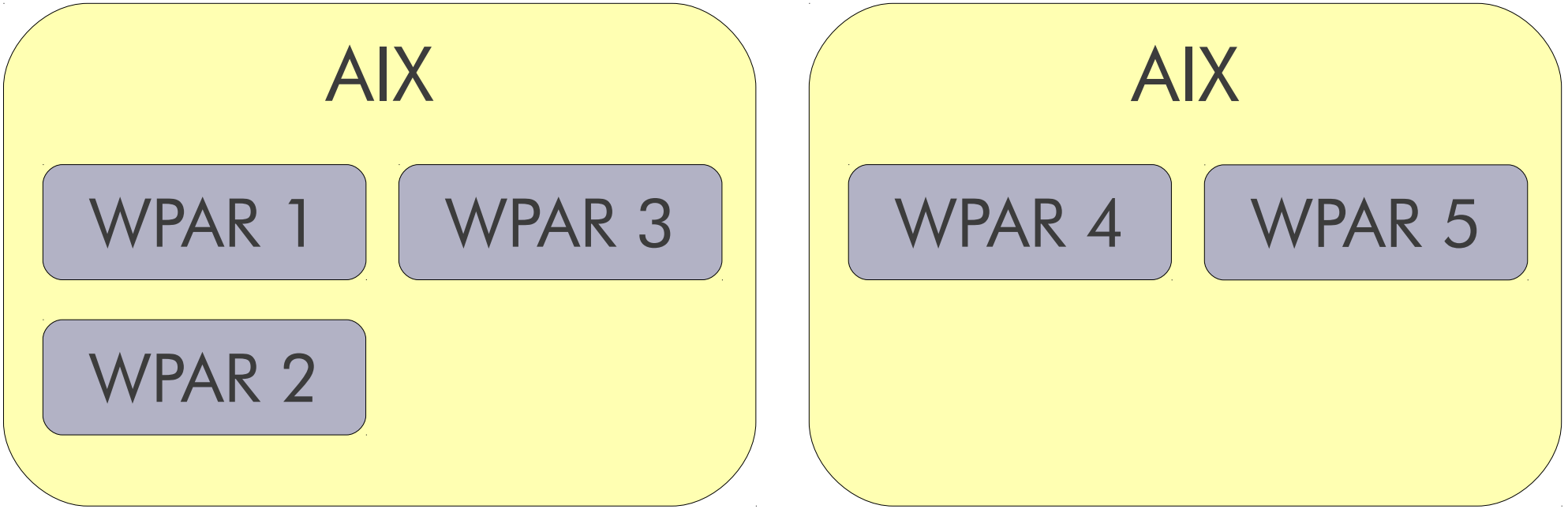

LPAR
AIX
LPAR
AIX
LPAR
Linux
Power VM Hypervisor
Physical Machine

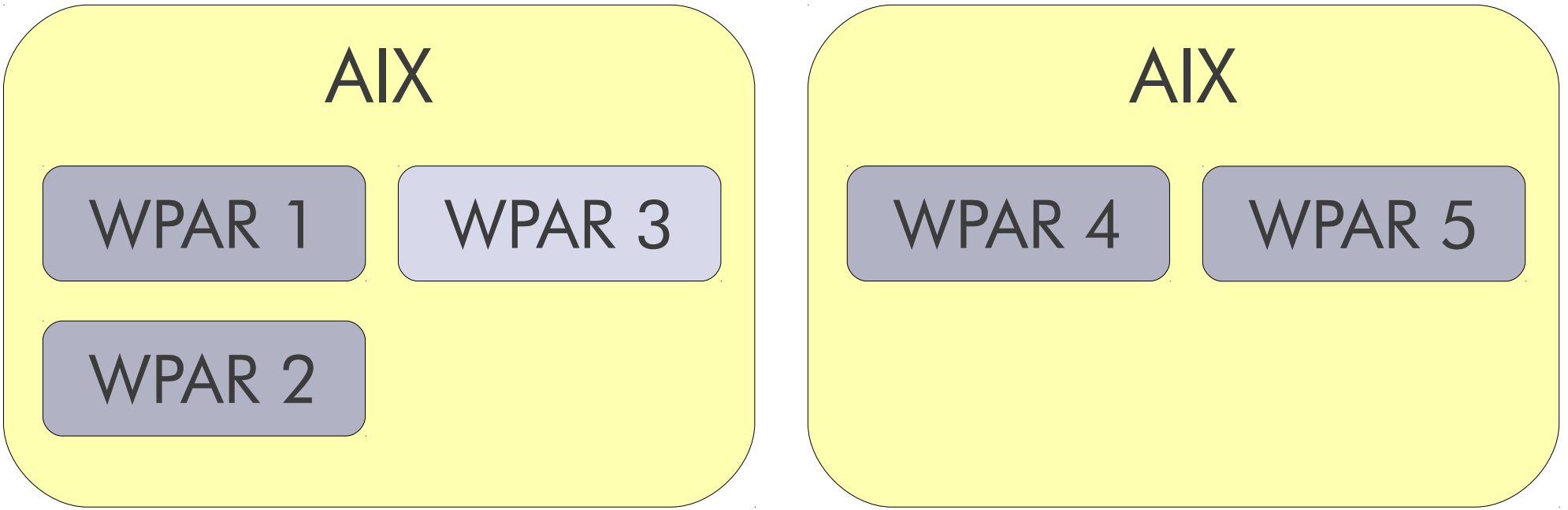

LPAR
AIX
1CPU
LPAR
AIX
2 CPU
LPAR
Linux
2 CPU
LPAR
Linux
LPAR
AIX
LPAR
AIX
3 CPU pool
Power VM Hypervisor
Physical Machine with 8 CPUs

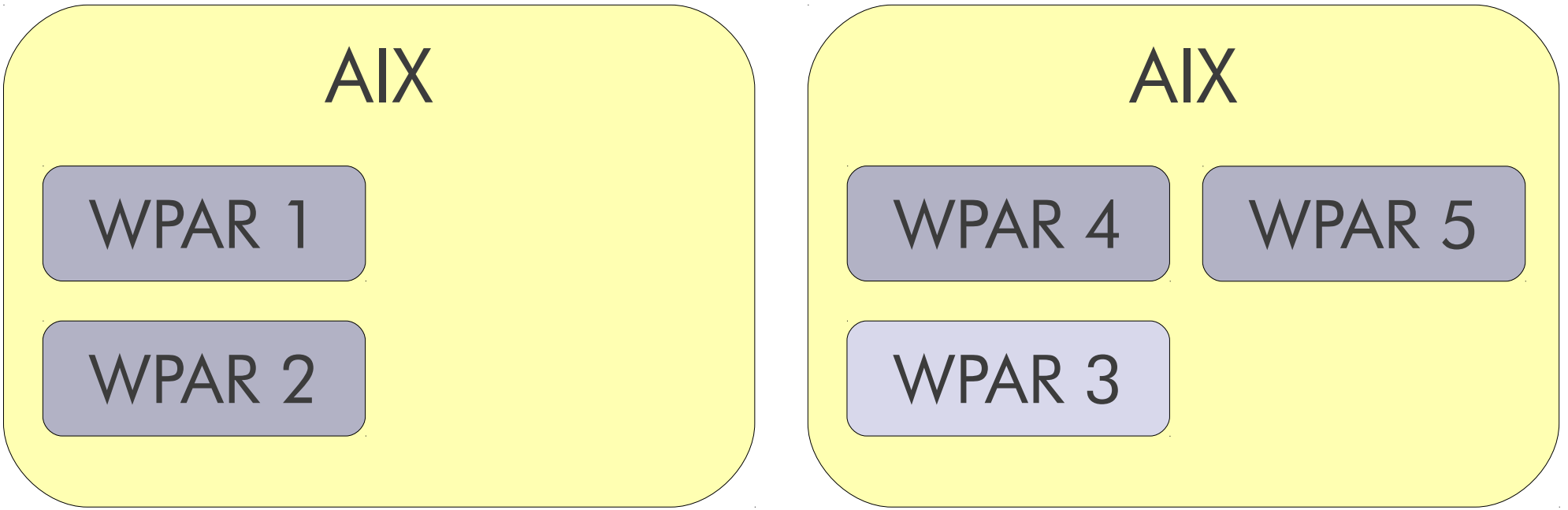

Virtual CPUs
Pool of Real CPUs

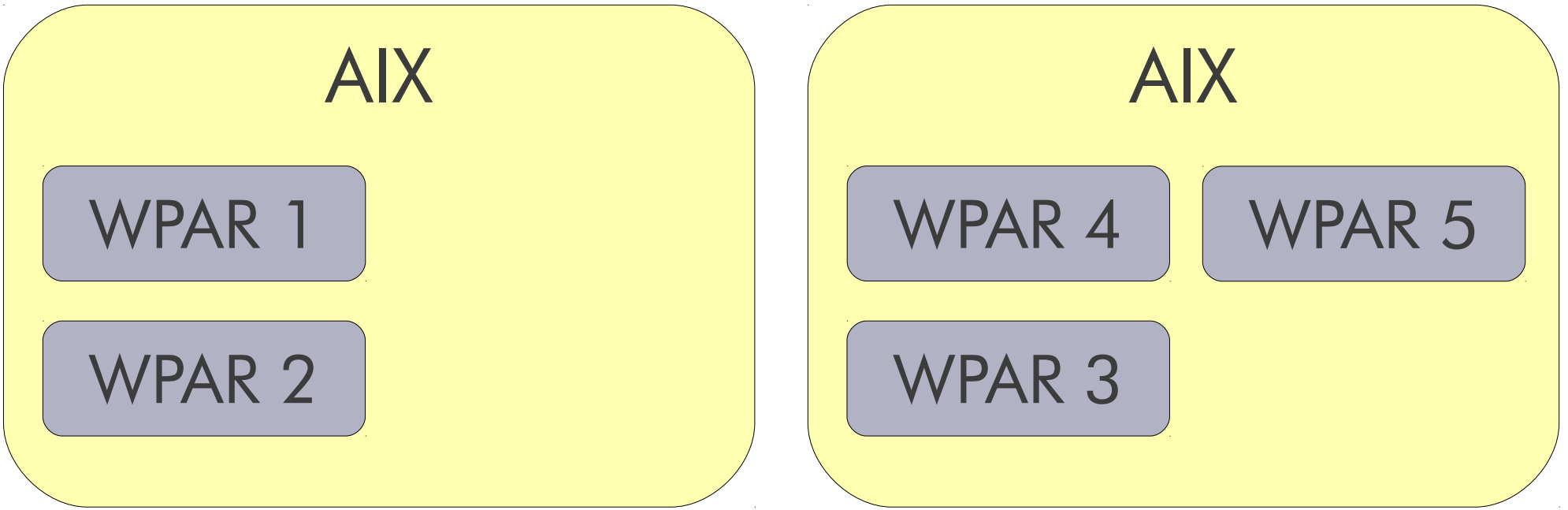

LPAR
LPAR
LPAR
VIOS
AIX
Linux
Resource
Resource
Resource
Resource
Power VM Hypervisor
Resource
Resource
Physical Machine


LAN
Server 1
LPAR
VIOS
Power VM
Server 2
VIOS
LPAR
Power VM
SAN


AIX
WPAR 1
WPAR 3
WPAR 2
AIX
WPAR 4
WPAR 5

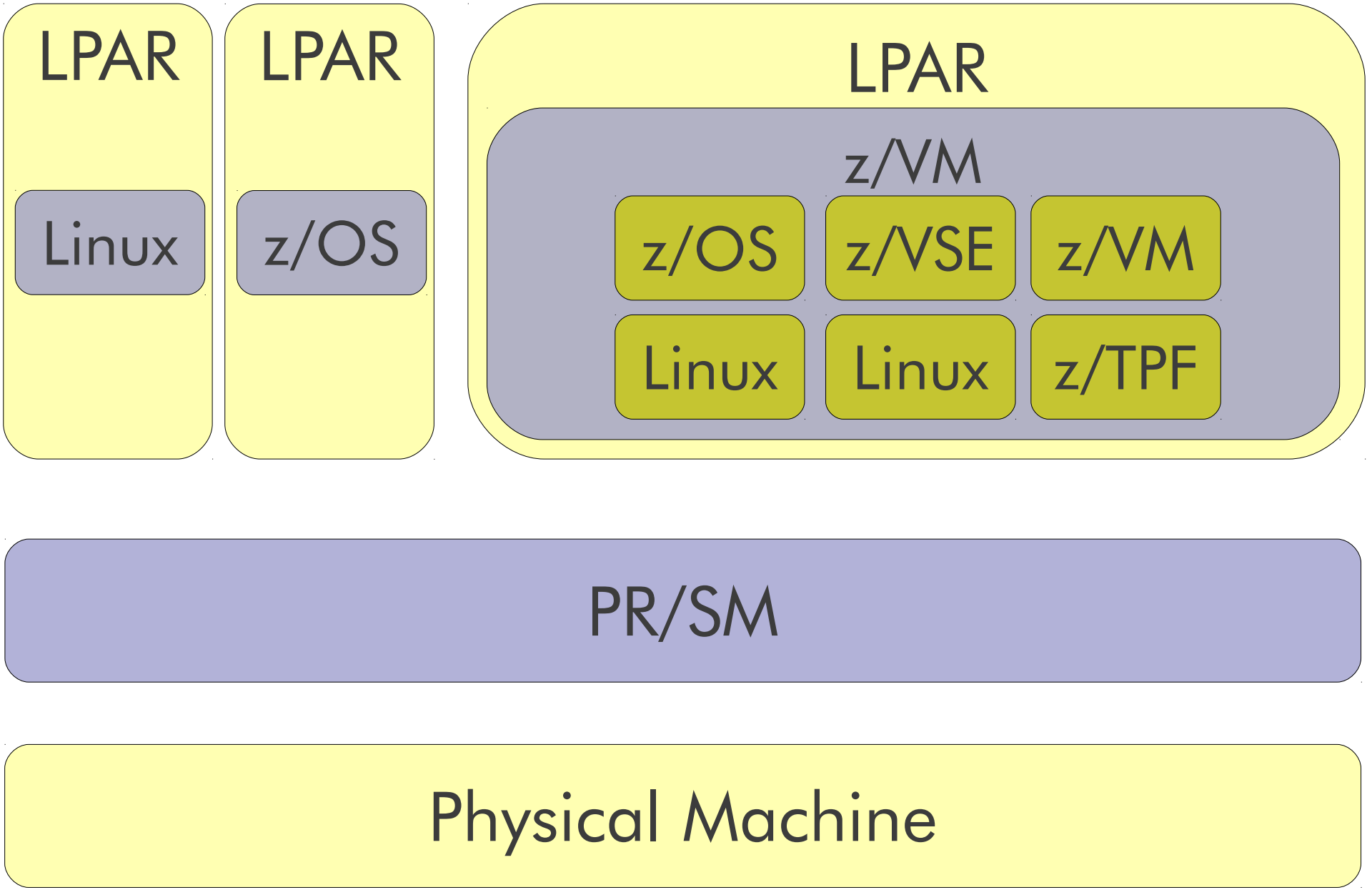

AIX
WPAR 1
WPAR 3
WPAR 2
AIX
WPAR 4
WPAR 5


AIX
WPAR 1
WPAR 2
AIX
WPAR 4
WPAR 5
WPAR 3

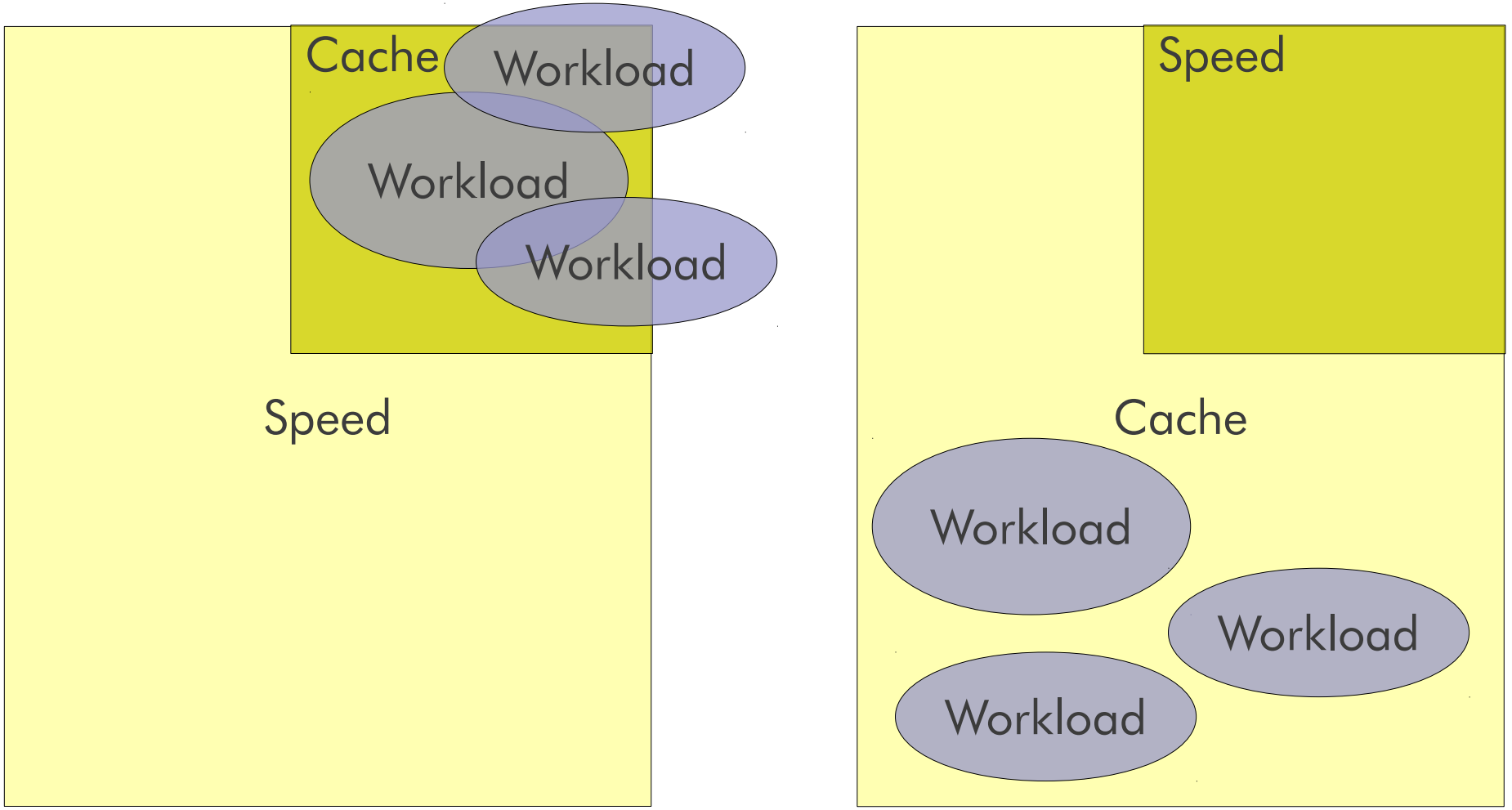

AIX
WPAR 1
WPAR 2
AIX
WPAR 4
WPAR 5
WPAR 3

System z

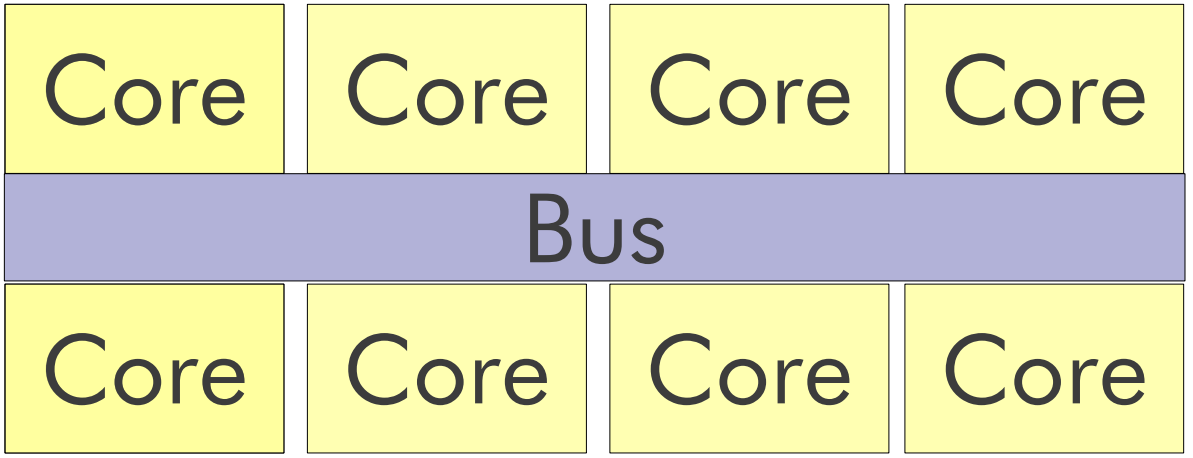
Virtualizing System/360

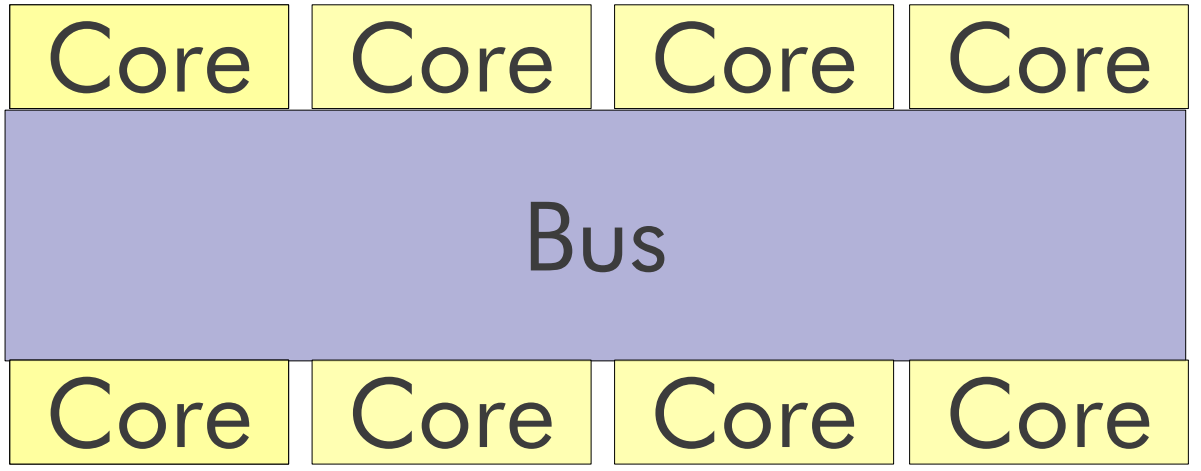

LPAR
LPAR
LPAR
Linux
z/OS
z/VM
z/OS
z/VSE
z/VM
Linux
Linux
z/TPF
PR/SM
Physical Machine

Chip Design Affects Virtualization

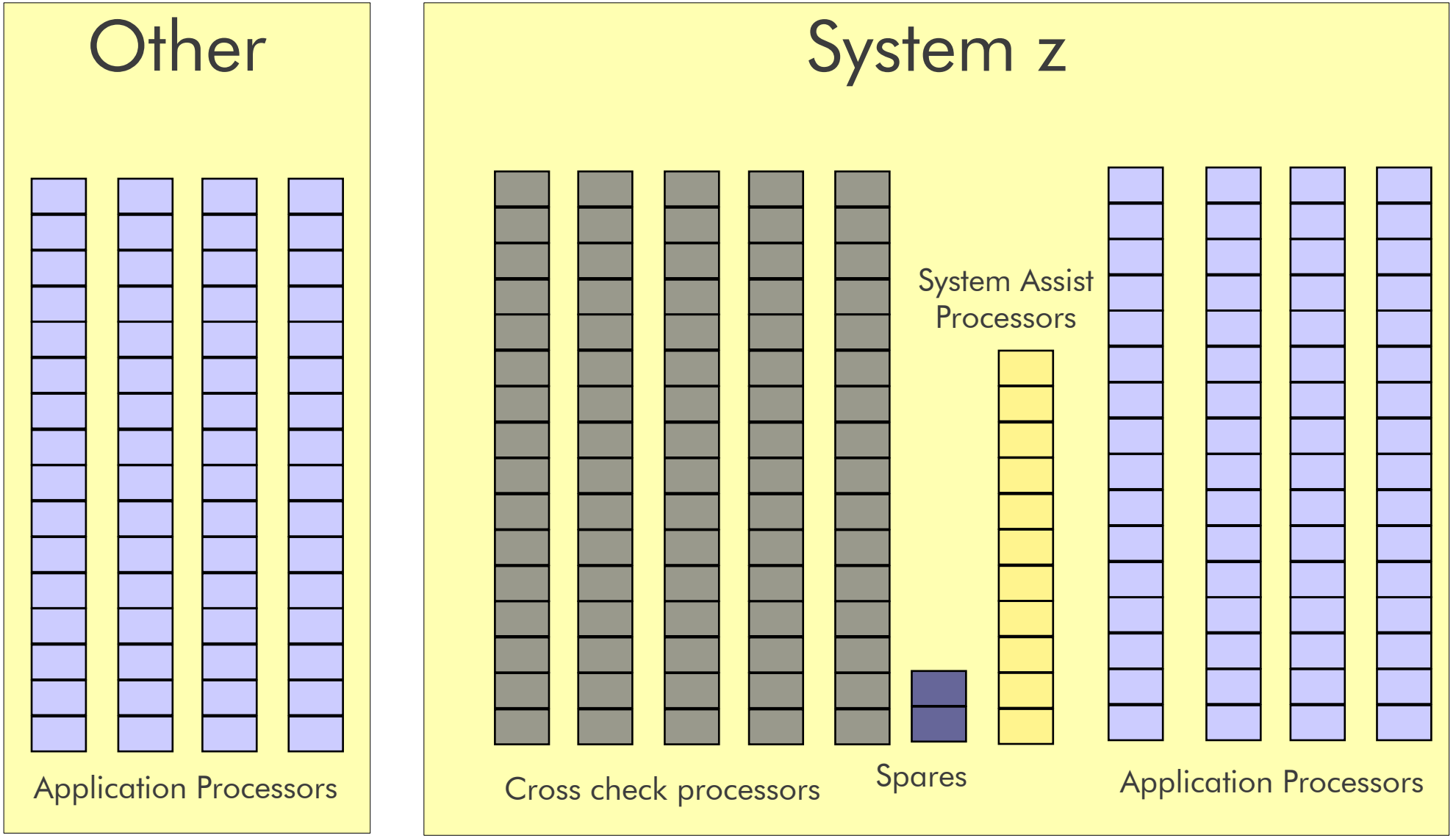


Design Differences

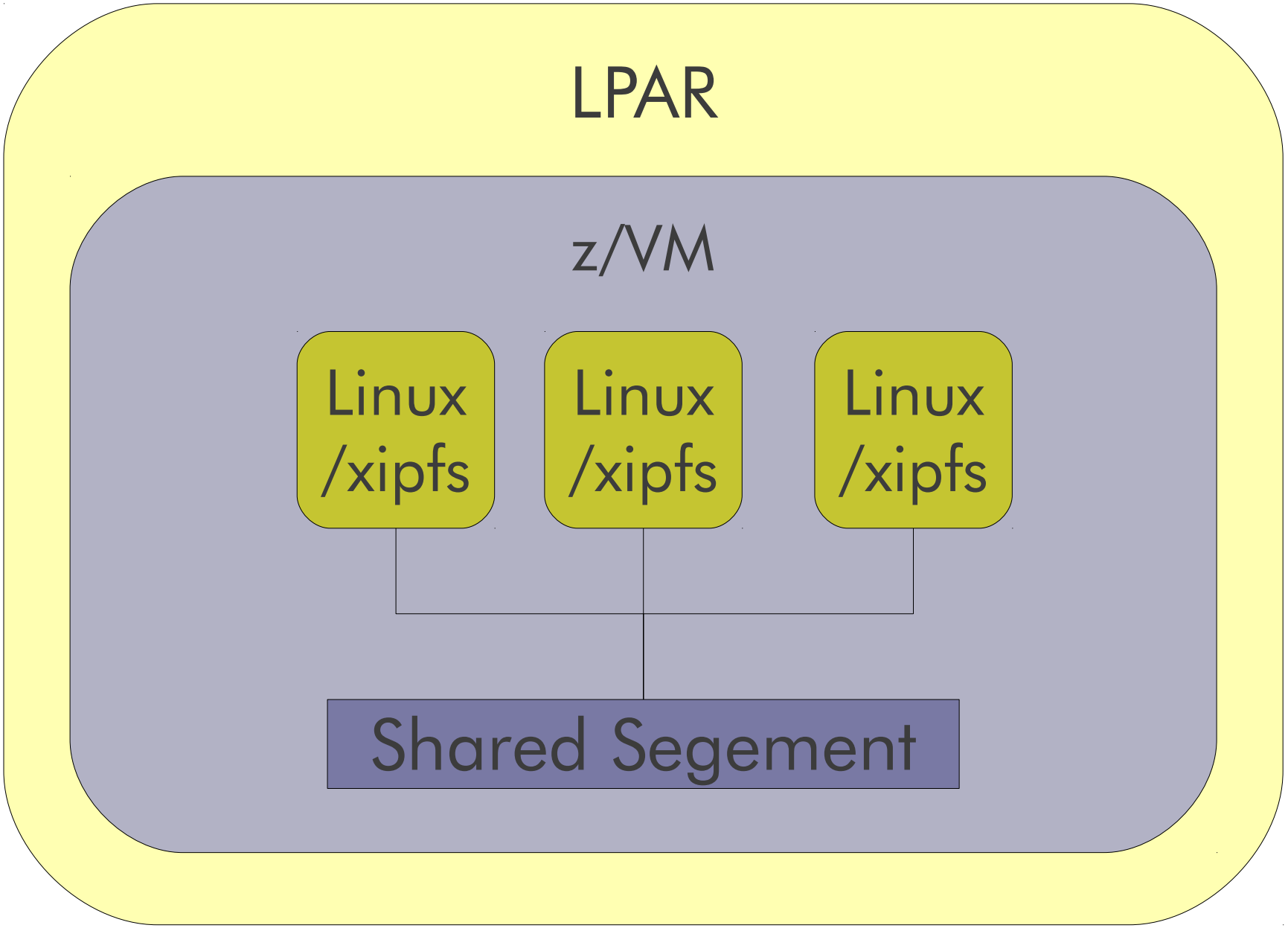




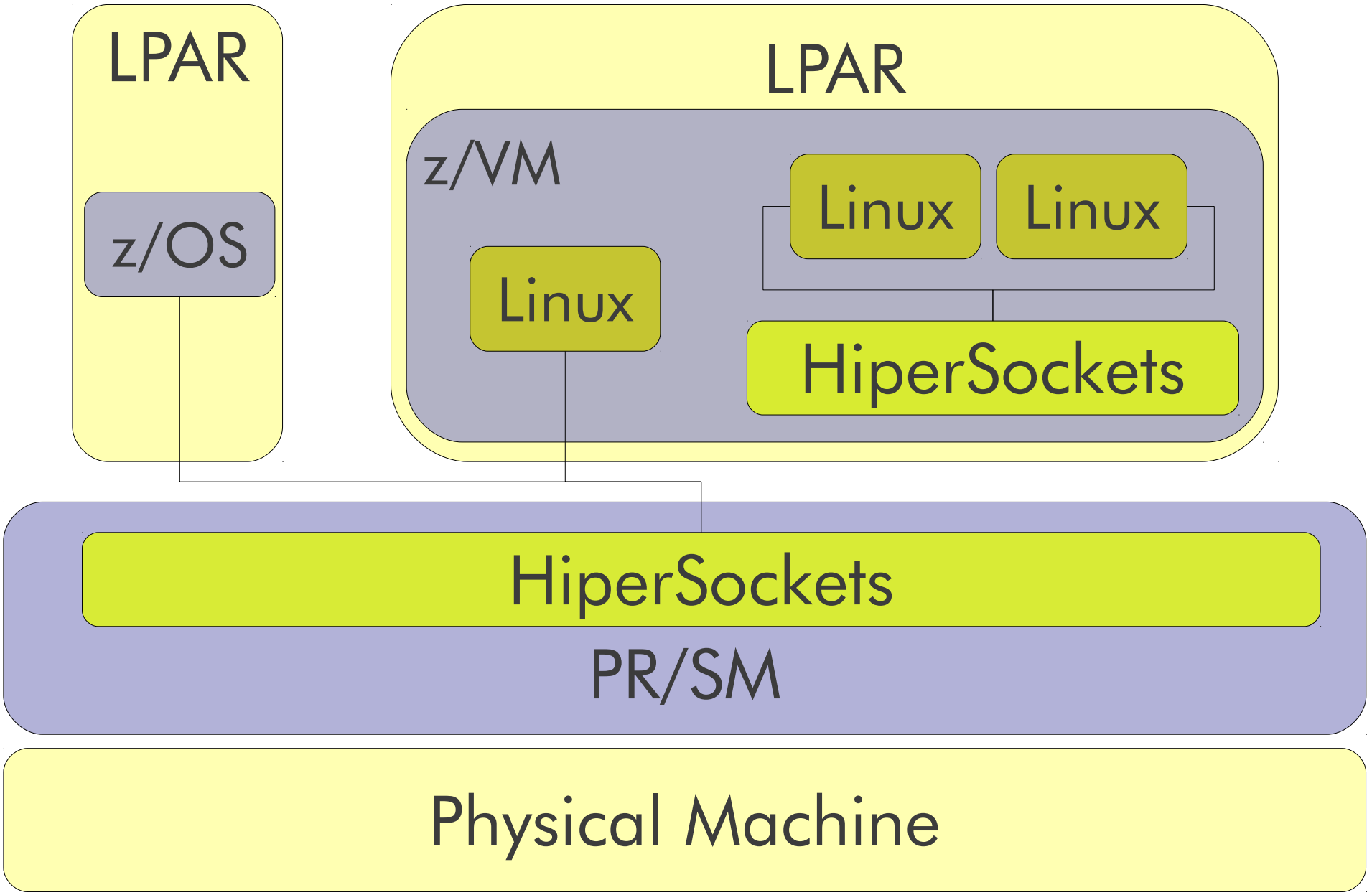
Comparison of 64-way Machines



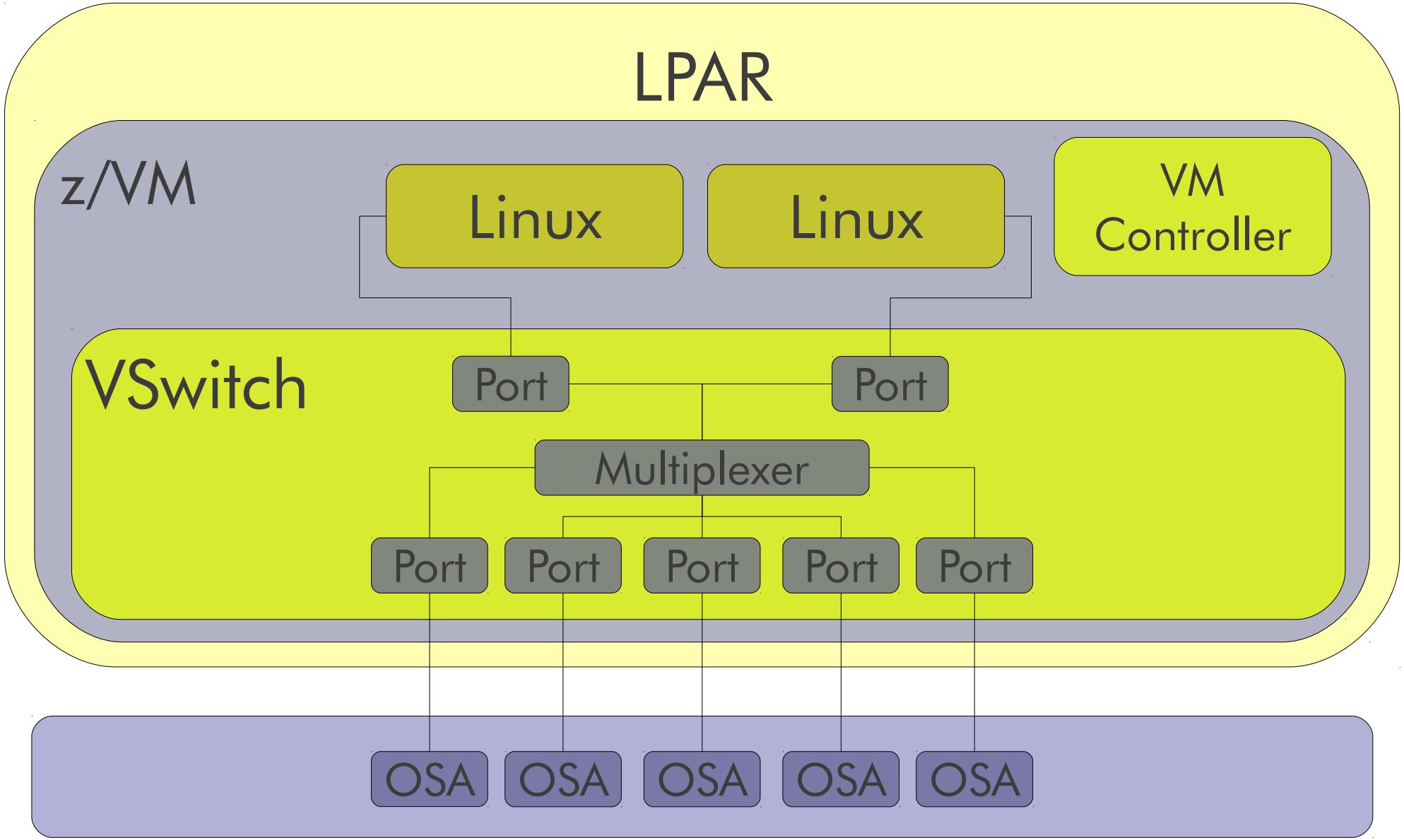
DCSS and XIP

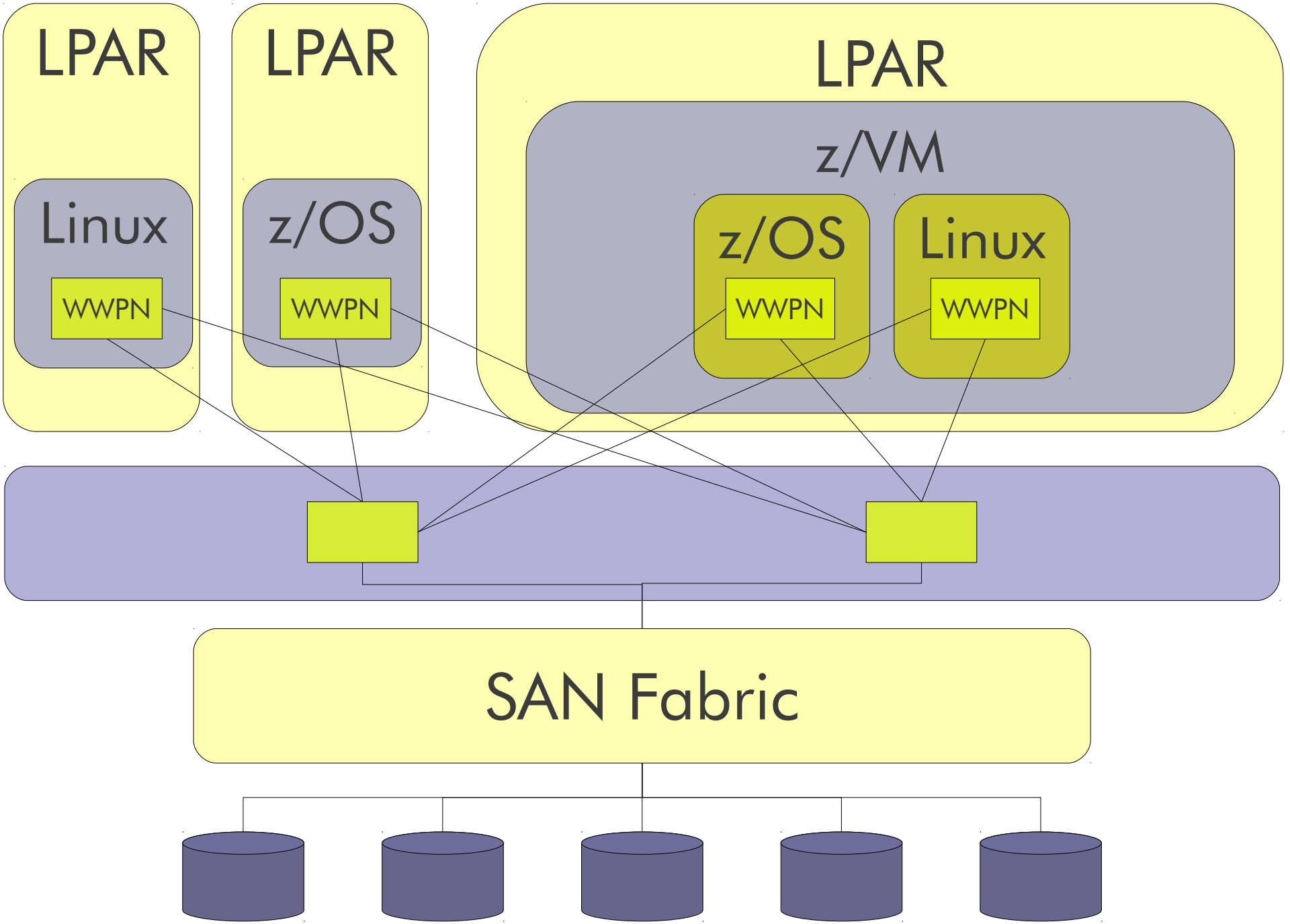

LPAR
z/VM
Linux /xipfs
Linux /xipfs
Linux /xipfs
Shared Segement

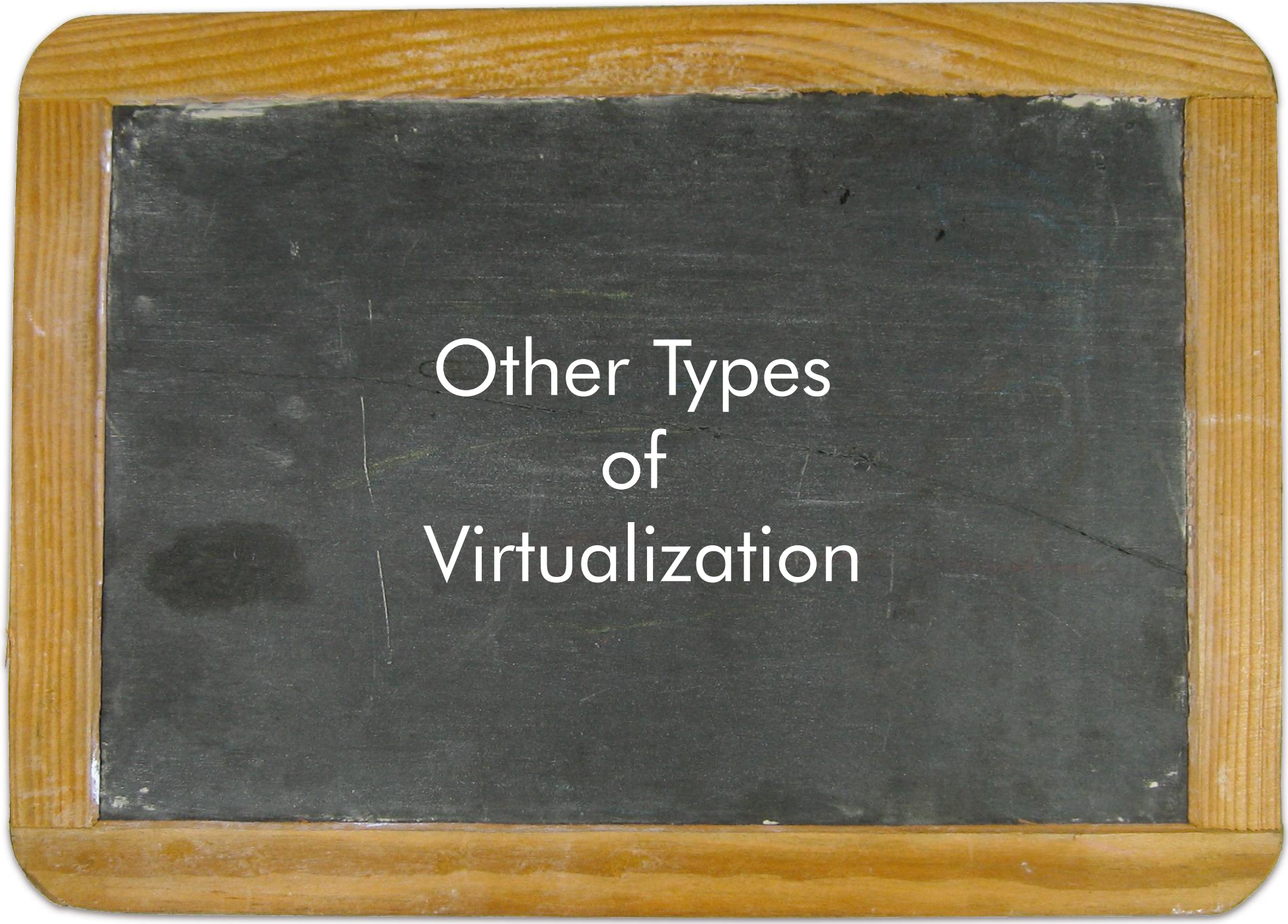
HiperSockets



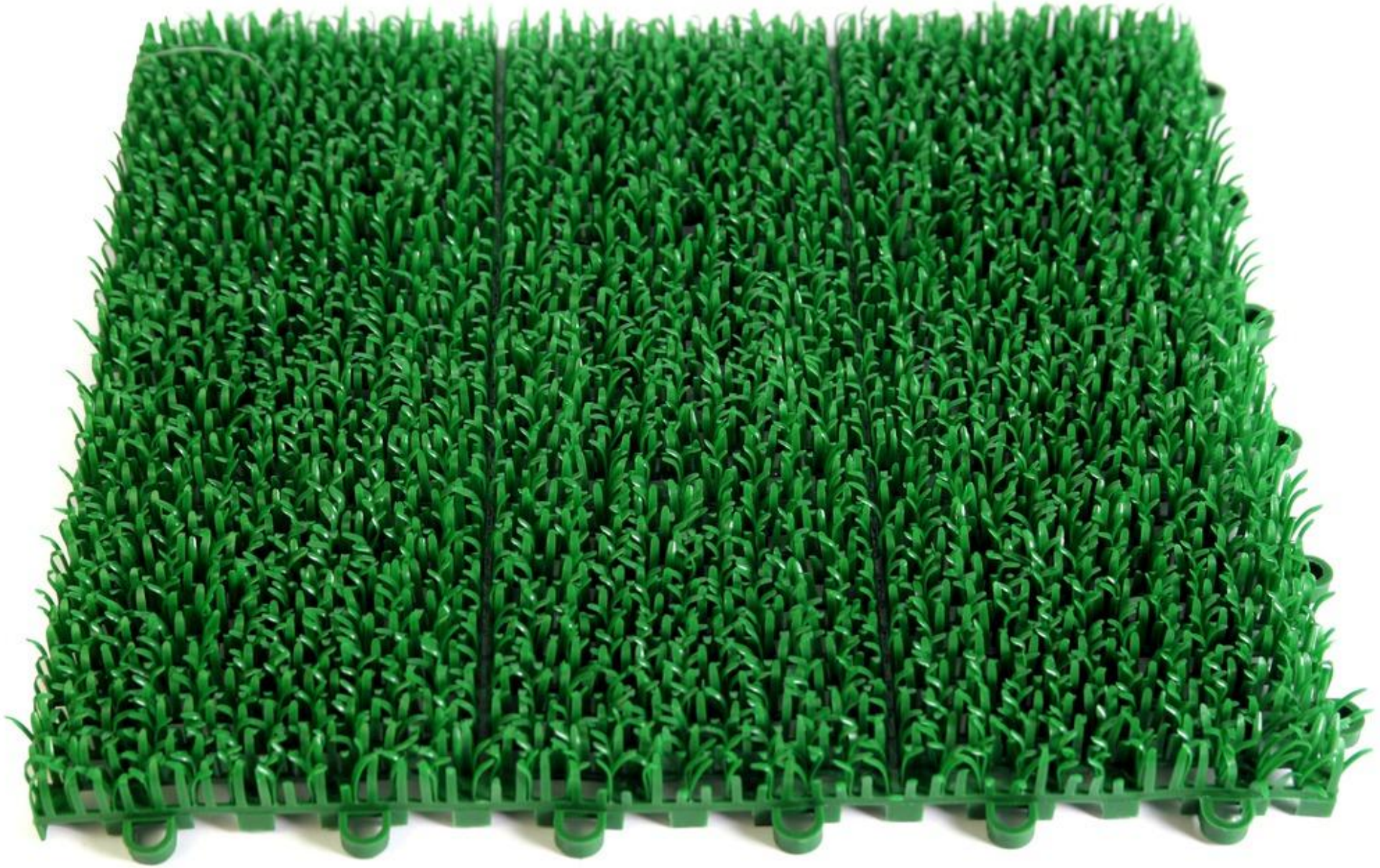
VSwitch


LPAR
z/VM
Linux
Linux
VM Controller
VSwitch
Port
Port
Multiplexer
Port
Port
Port
Port
Port
OSA
OSA
OSA
OSA
OSA


LPAR
LPAR
LPAR
z/VM
Linux
z/OS
z/OS
Linux
WWPN
WWPN
WWPN
WWPN
SAN Fabric


Other Types
of
Virtualization

Emulation

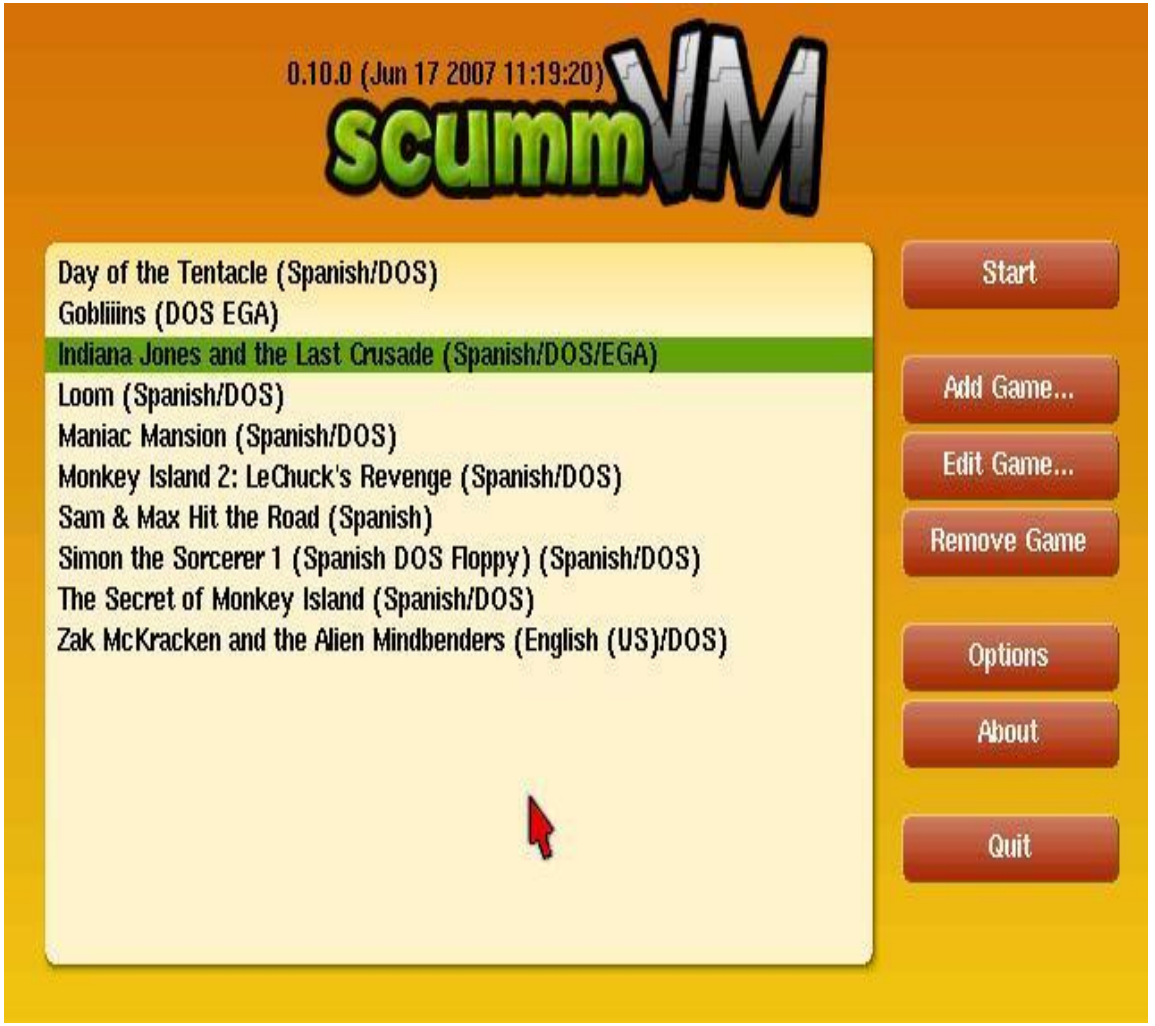


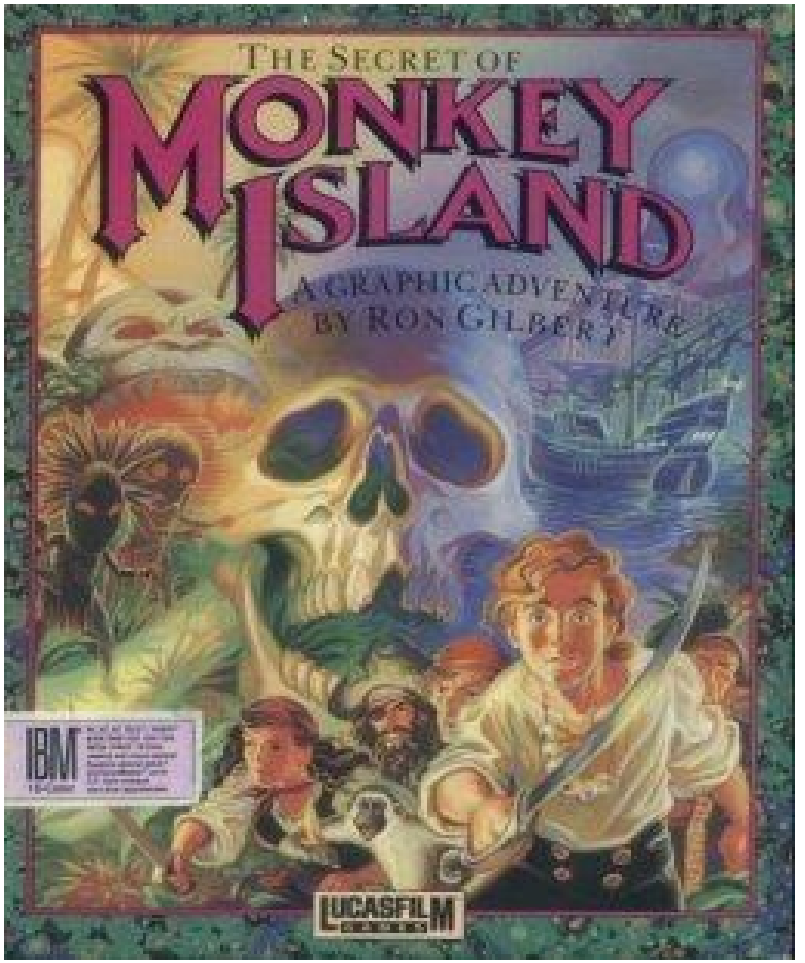
3270


Vista Session A
File Edit Font Transfer Macro Options Window Help
Have Fun!
IP Address = 69.121.152.51
YTAM 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
MA
0.0 09/14/07.257 04:26PM 192.86.32.16
a
24,1

ScummVM

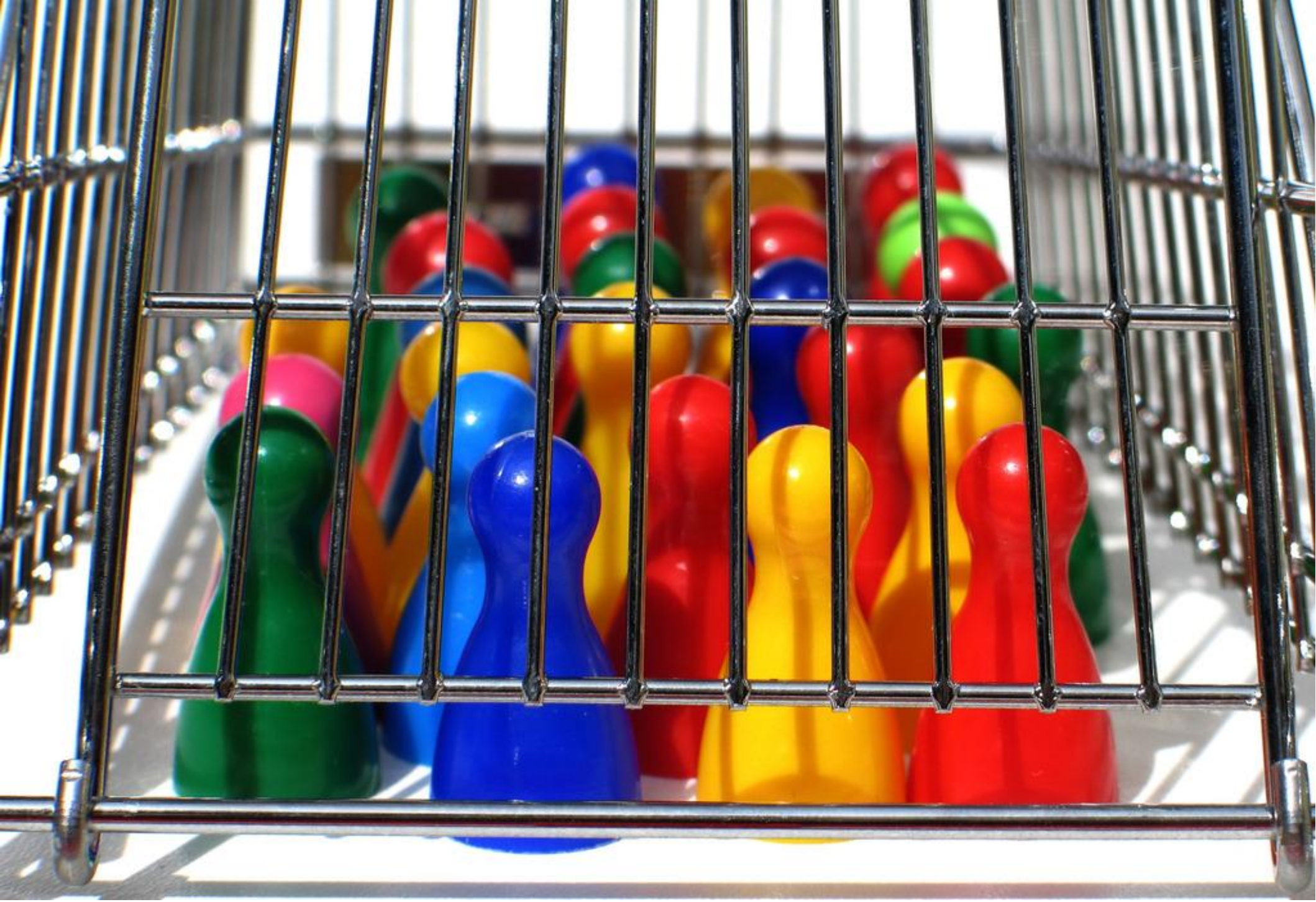
Script Creation Utility for Maniac Mansion



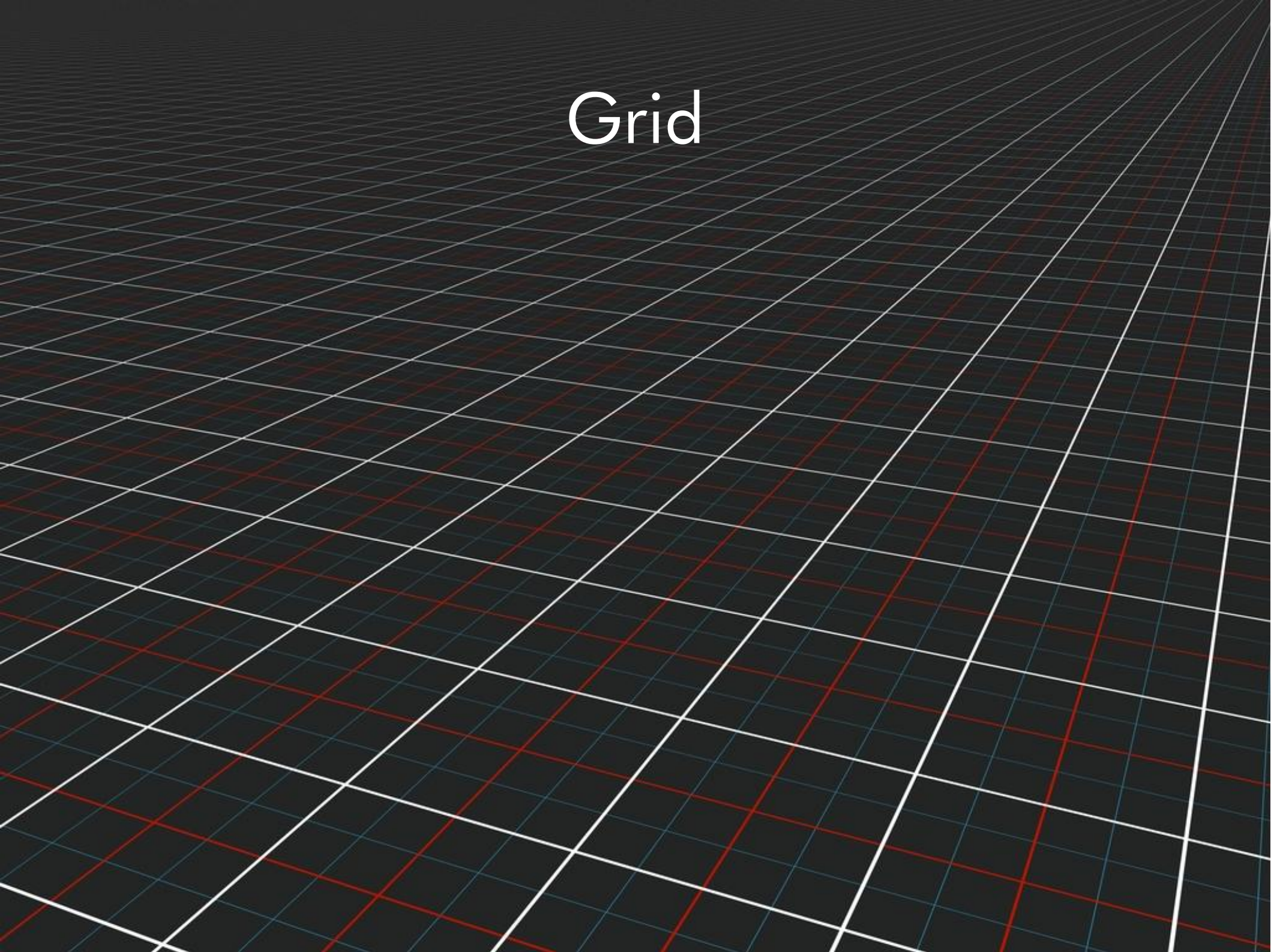


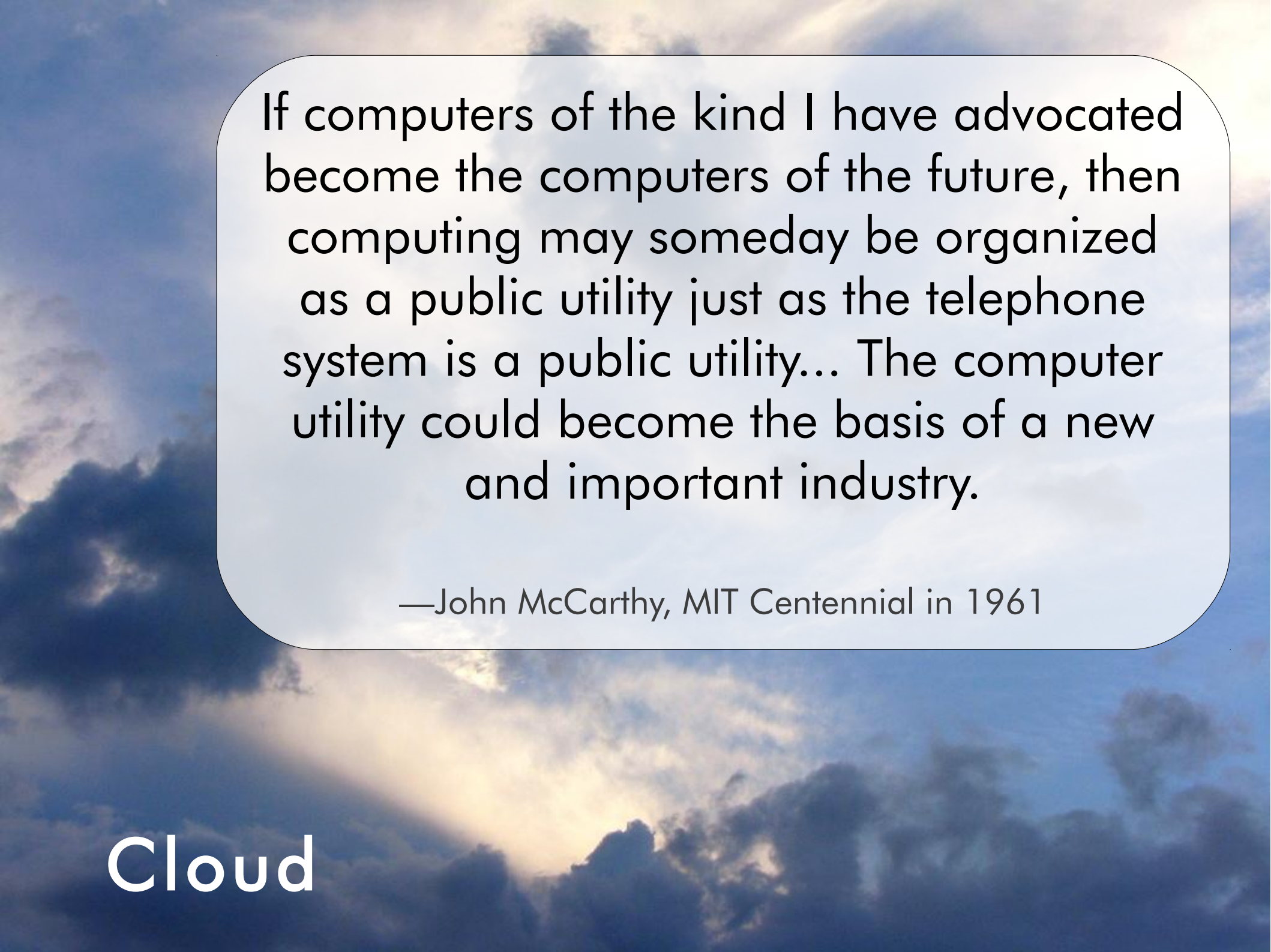
Wii

Wine



Chroot (jail)


Grid


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

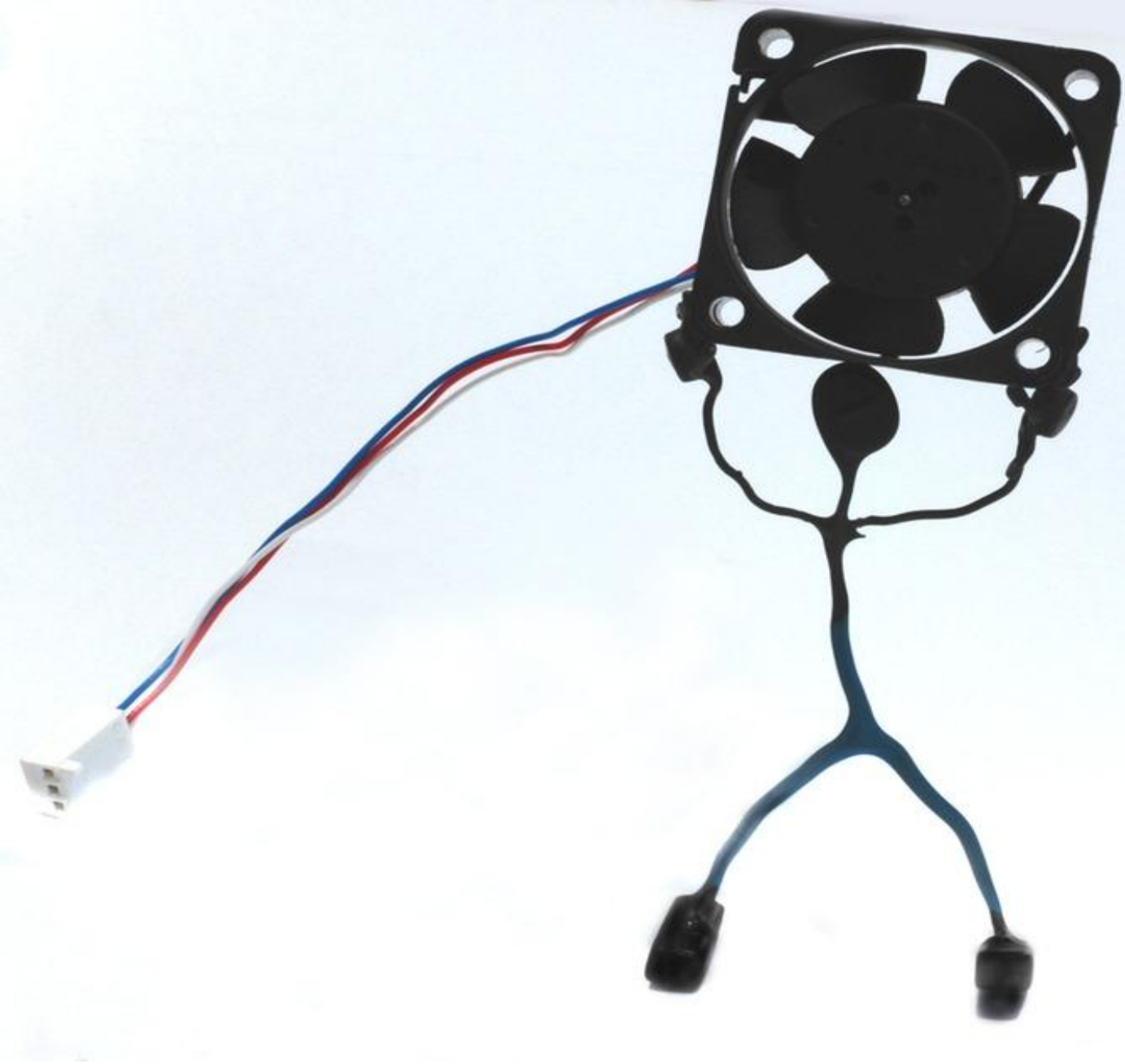
How do I choose?


What are
the pain
points

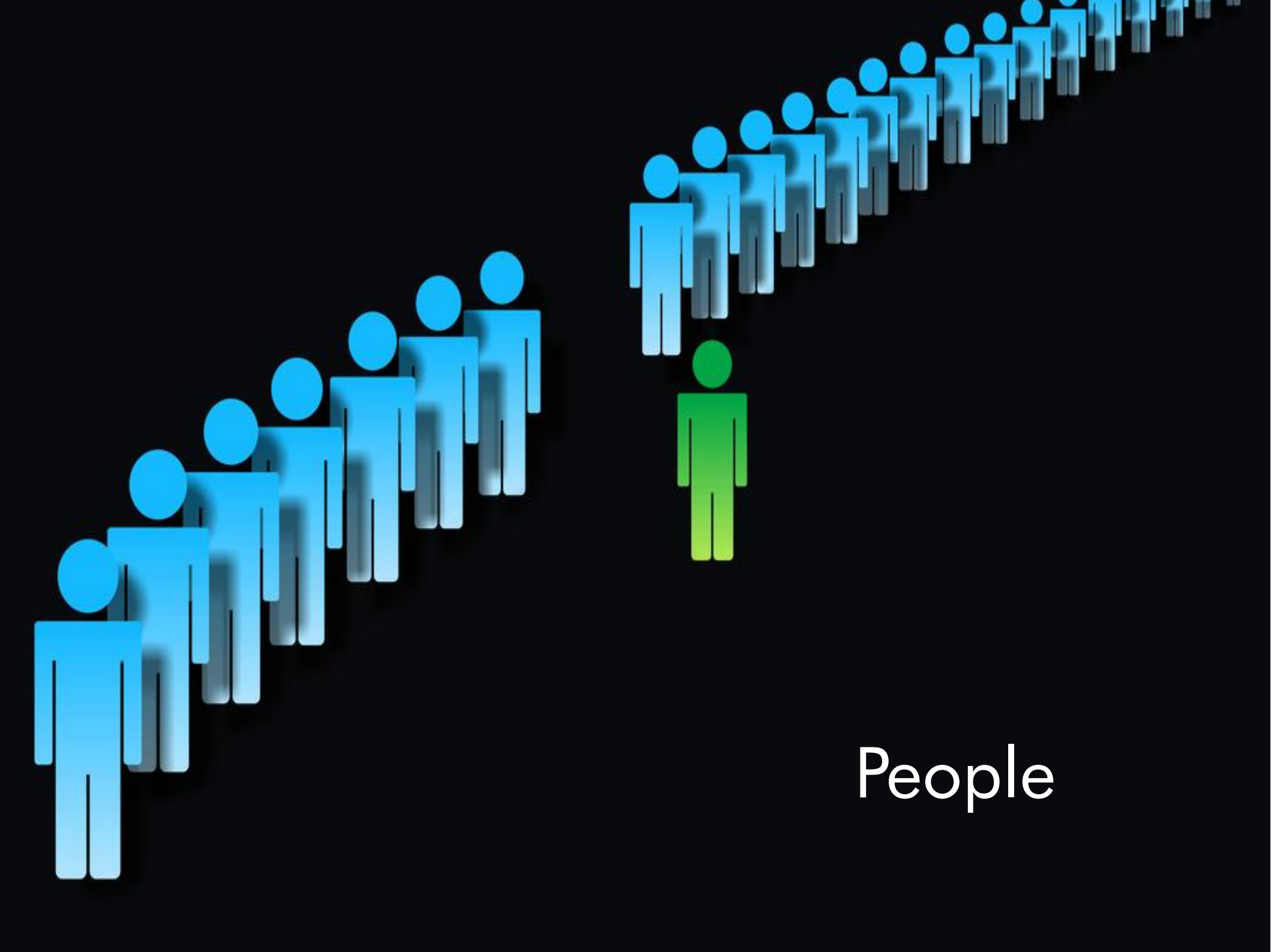
Cost


Complexity

Power



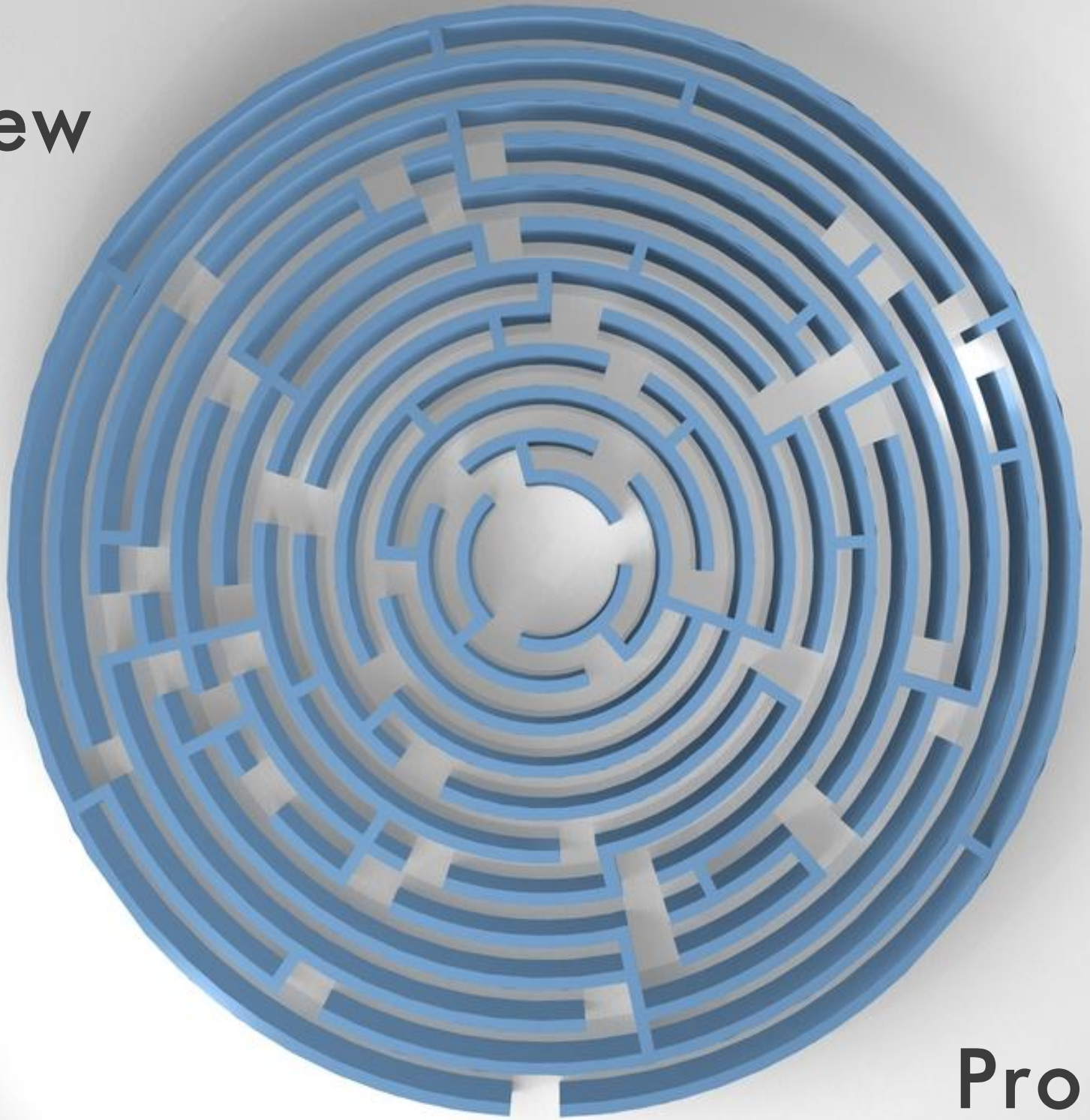
Cooling


People


DEAD
END
MINIMUM
MAINTENANCE
ROAD
TRAVEL AT YOUR OWN RISK

Know your workload

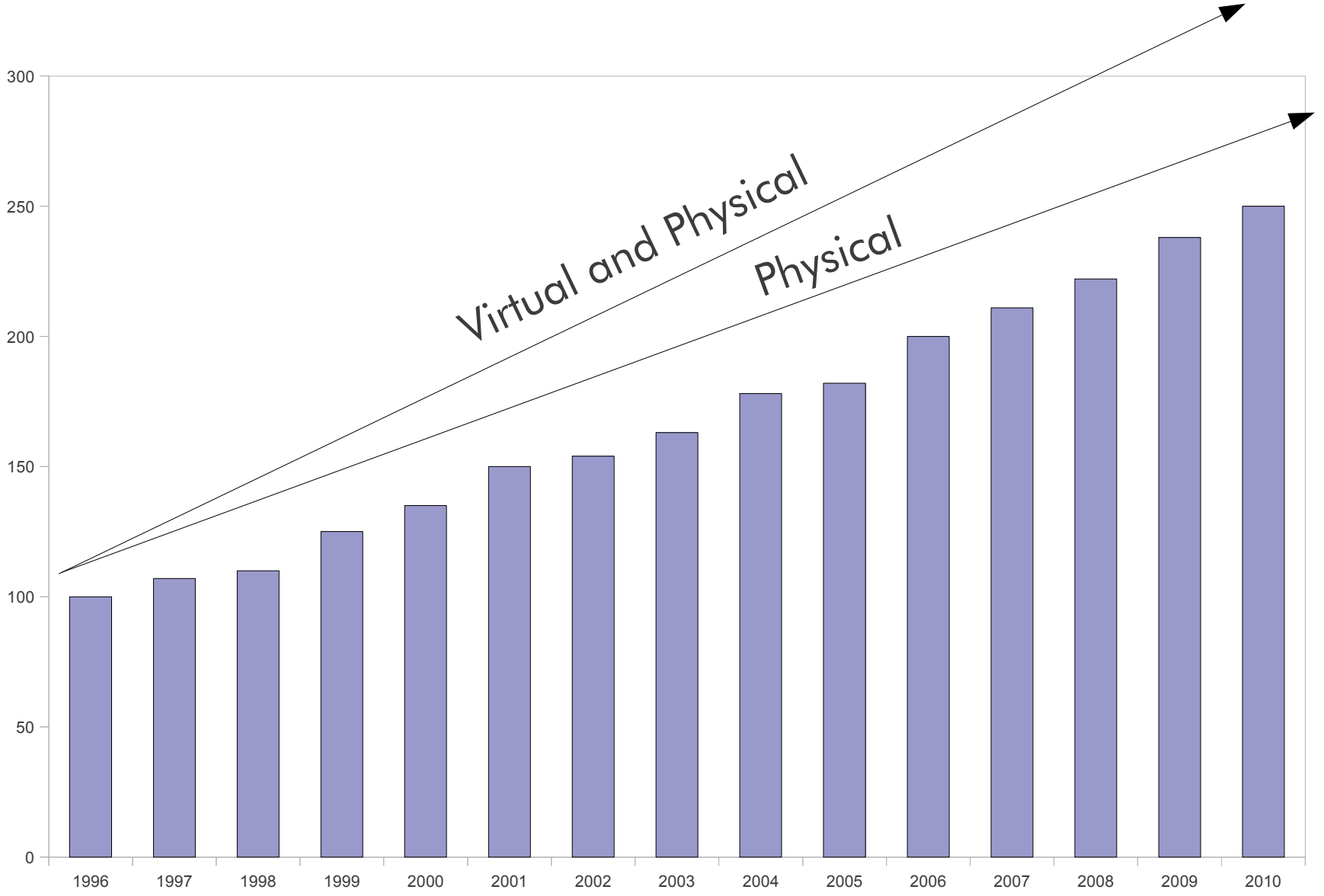
New



Problems

Utilization

Management Challenges




Agent
Proliferation

Where is my data ?

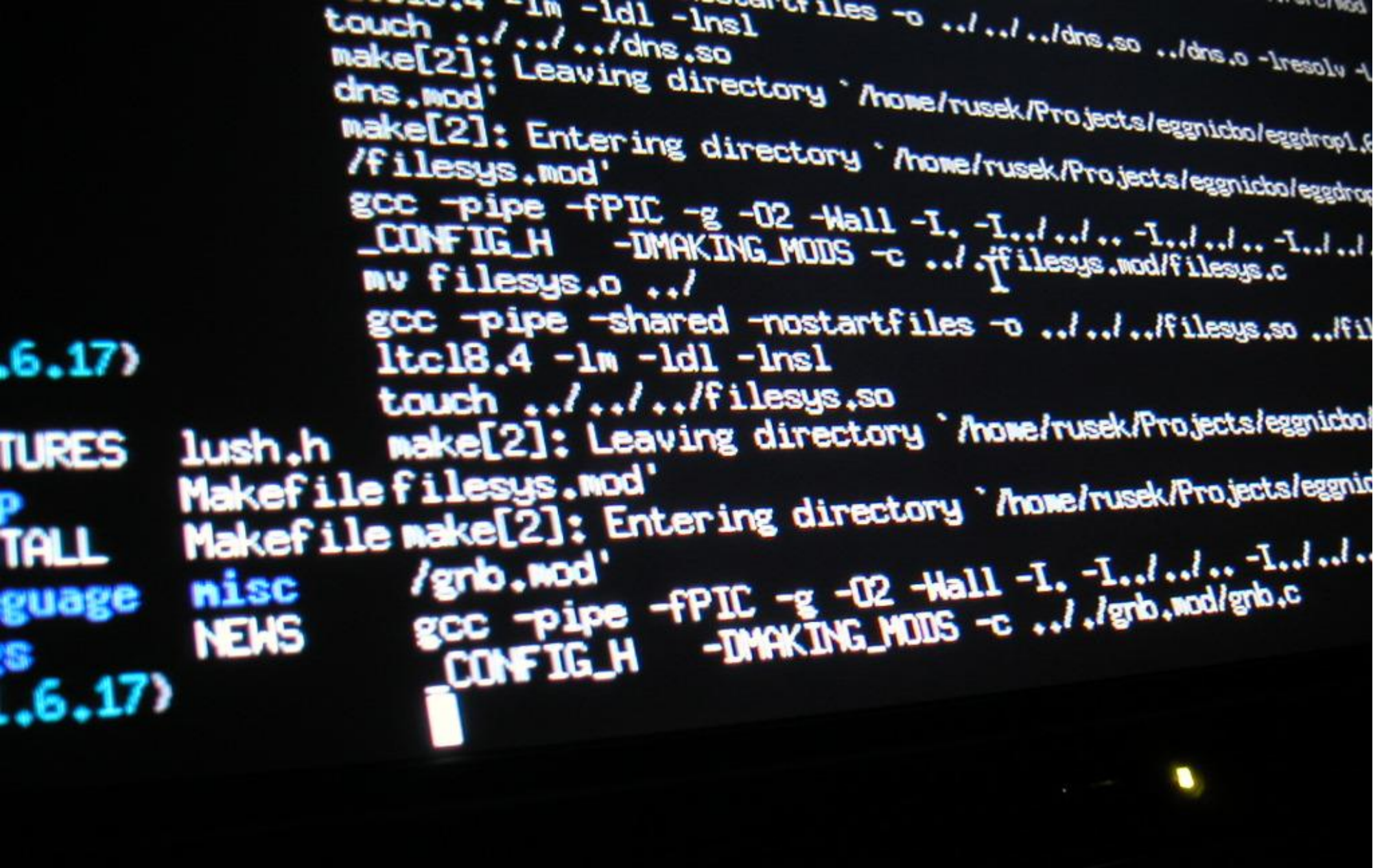
Accountability


Security


Image
Management

Where to start ?


Go for the low hanging fruit



Development


Test
BOMEX
80 ml
APPROX.
60
40
20

Quality Assurance

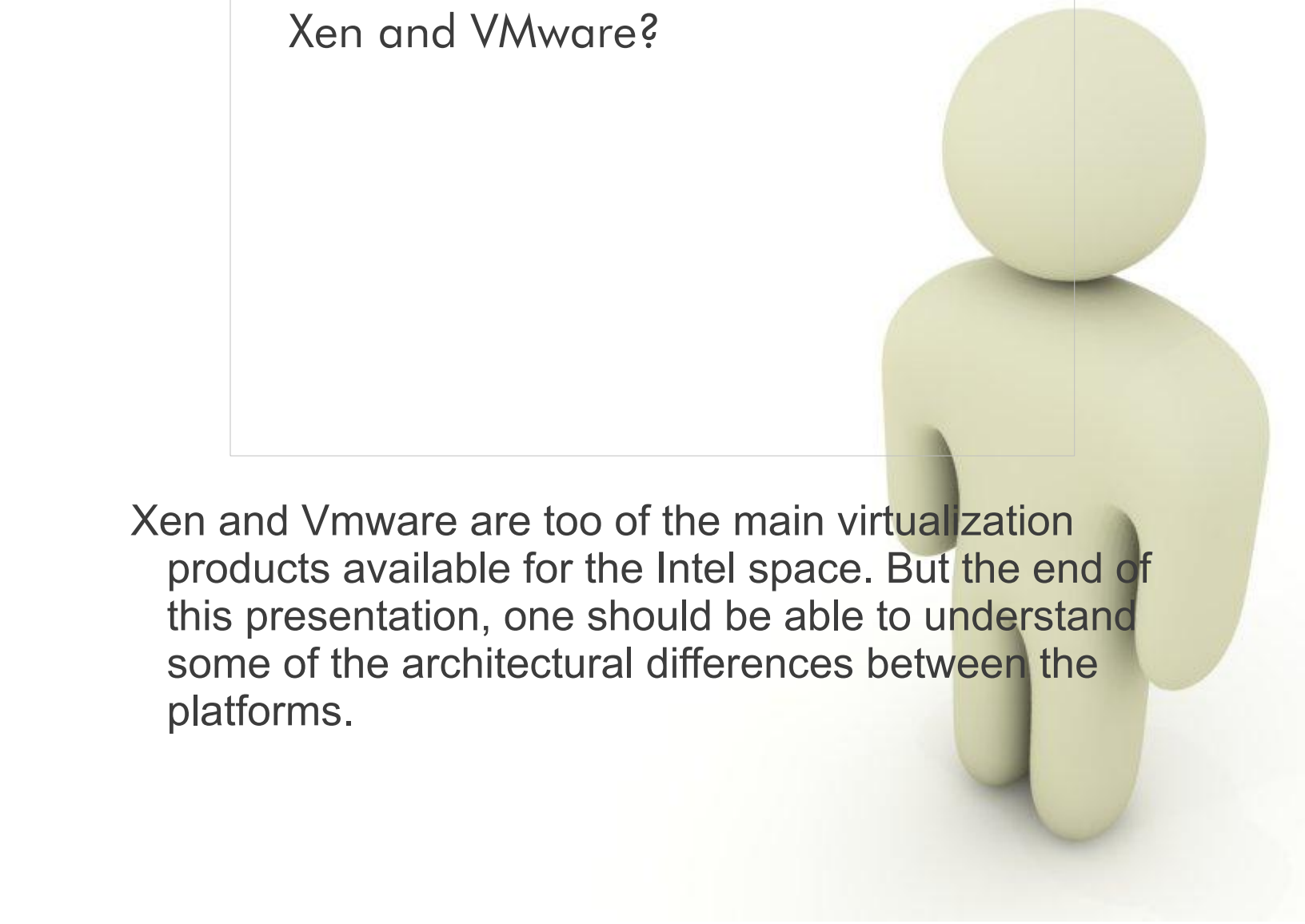

04972G072
1,000,000
100,000
10,000
1,000
PER REV
CUBIC
FEET
Production

The End

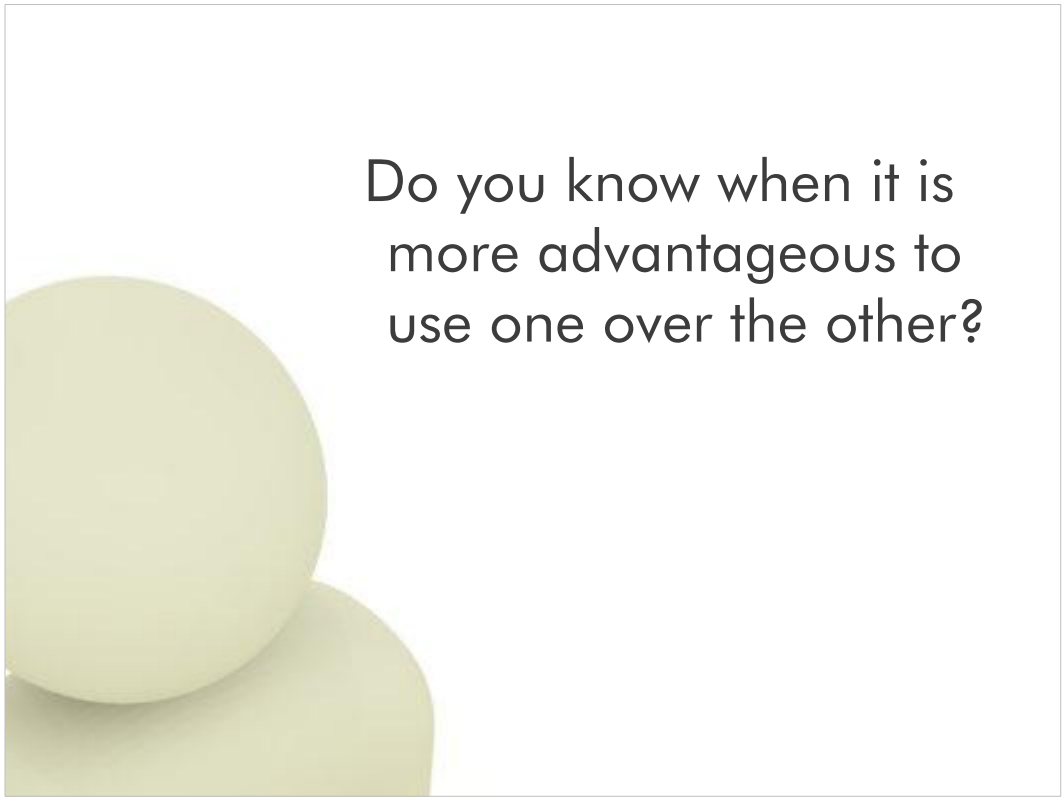
Comparing and Contrasting Virtualization Technologies

12359

Do you know the differences between
Xen and VMware?



Xen and Vmware are too of the main virtualization products available for the Intel space. But the end of this presentation, one should be able to understand some of the architectural differences between the platforms.



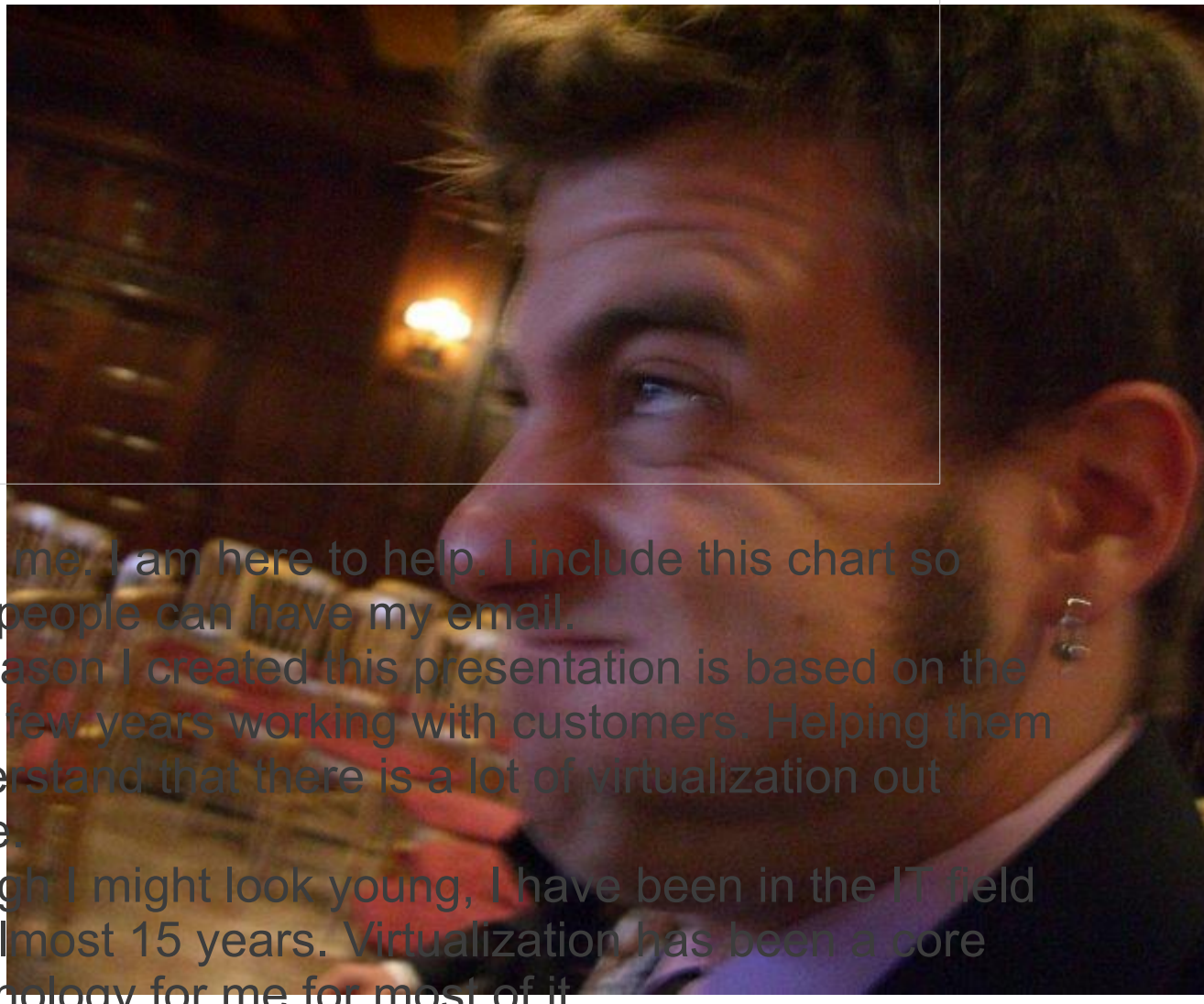
No company uses a single hardware platform. Therefore, there may be times when a company would want to use more than one hardware virtualizatoin technology. This presentation should help understand which platform provided the most advantage for a given application or workload.

Virtualization can be a complicated subject with many different facets.

It is not always easy to choose the strategy that best fits your needs.

Making a choice is hard when you don't have the knowledge to make an informed decision. This pretension should help frame the question such that one can become better suited to make that decision.

I am here to help
buzzetti@us.ibm.com



- This is me. I am here to help. I include this chart so that people can have my email.
- The reason I created this presentation is based on the past few years working with customers. Helping them understand that there is a lot of virtualization out there.
- Although I might look young, I have been in the IT field for almost 15 years. Virtualization has been a core technology for me for most of it.

Worldwide Design Centers

Montpellier, France
design.center@fr.ibm.com

Makuhari, Japan
design@jp.ibm.com

Boeblingen, Germany
design@de.ibm.com

The different places where I work. This slide is to show that the Design Center has a world wide presence, and that we can help customers in almost all geographies.

Because we have such a wide reach, we have have insight into a wide variety of customer profiles. From insurance to banking and universities, the Design Center can help solve even the most complex problems.

Worldwide Design Centers

Our mission – To architect & design innovative end to end solutions with selected worldwide clients that leverage leading-edge IBM technologies to accelerate their IT transformation

- **Analysis**
  Define and analyze various solution options that meet existing and future requirements
- **Assessment**
  Review and validate a planned solution, strategy, IT transformation, or architecture, and provide recommendations and roadmaps
- **Design**
  Define an architecture and high level design for an IT infrastructure to meet business requirements

What we do
This slide shows 3 different engagement types.
I have this slide here to show how we can help with different amounts of effort.

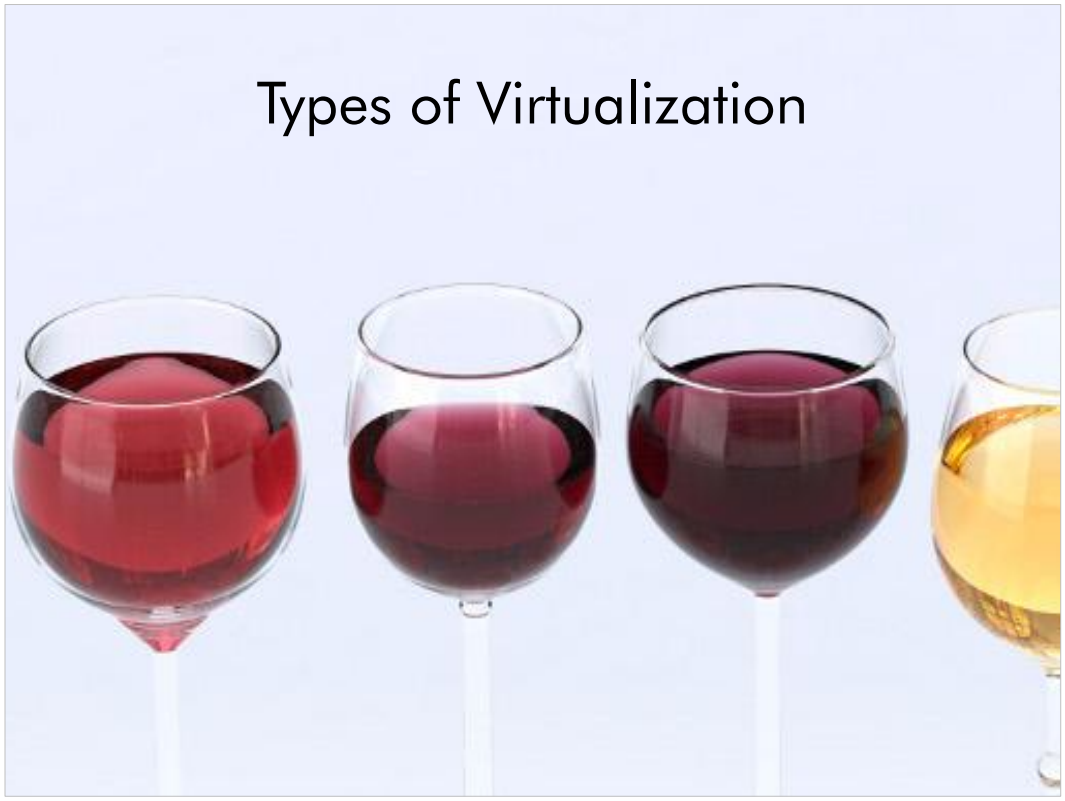
It is important to show that I work with customers for extended periods of time, and that the knowledge contained in this presentation comes from working with and preparing for these working session.

There is no free lunch

Popularized by science fiction writer Robert A. Heinlein in his 1966 novel The Moon Is a Harsh Mistress, which discusses the problems caused by not considering the eventual outcome of an unbalanced economy. In order to avoid a double negative, the acronym "TINSTAAFL" is sometimes used instead, meaning "There Is No Such Thing As A Free Lunch".

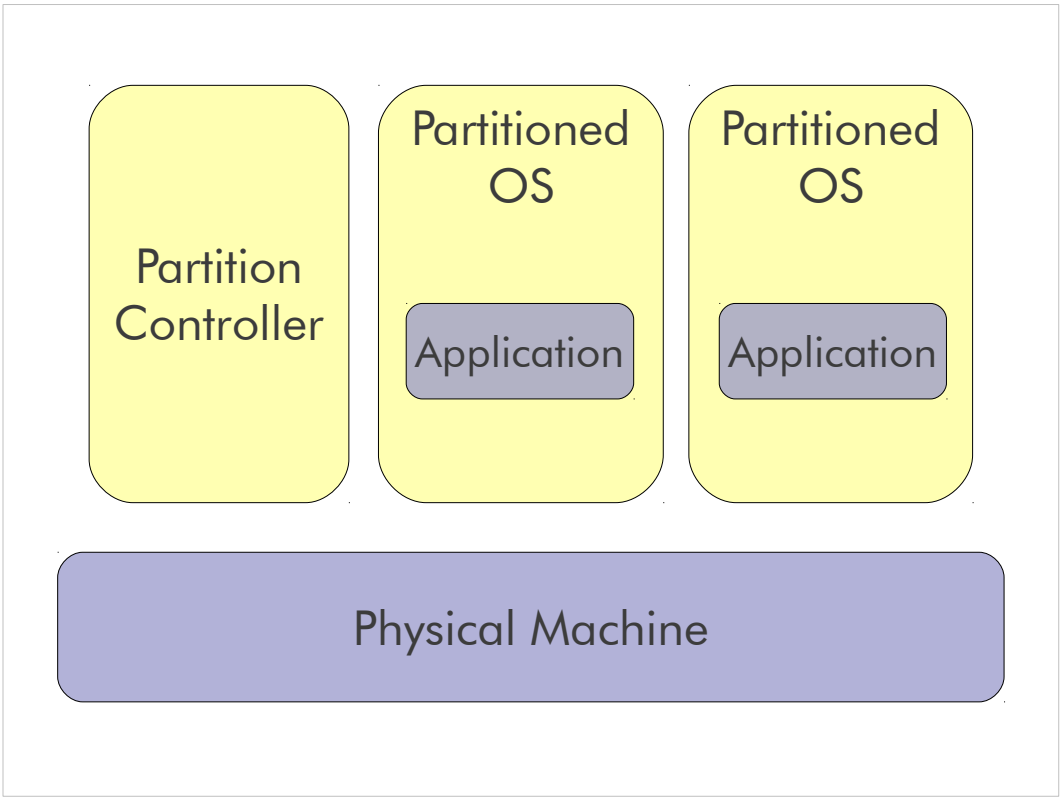
Greg Mankiw described the concept as: "To get one thing that we like, we usually have to give up another thing that we like. Making decisions requires trading off one goal against another."

When dealing with virtualization, trading of something like raw clock speed for larger cache, is an importatn choice.



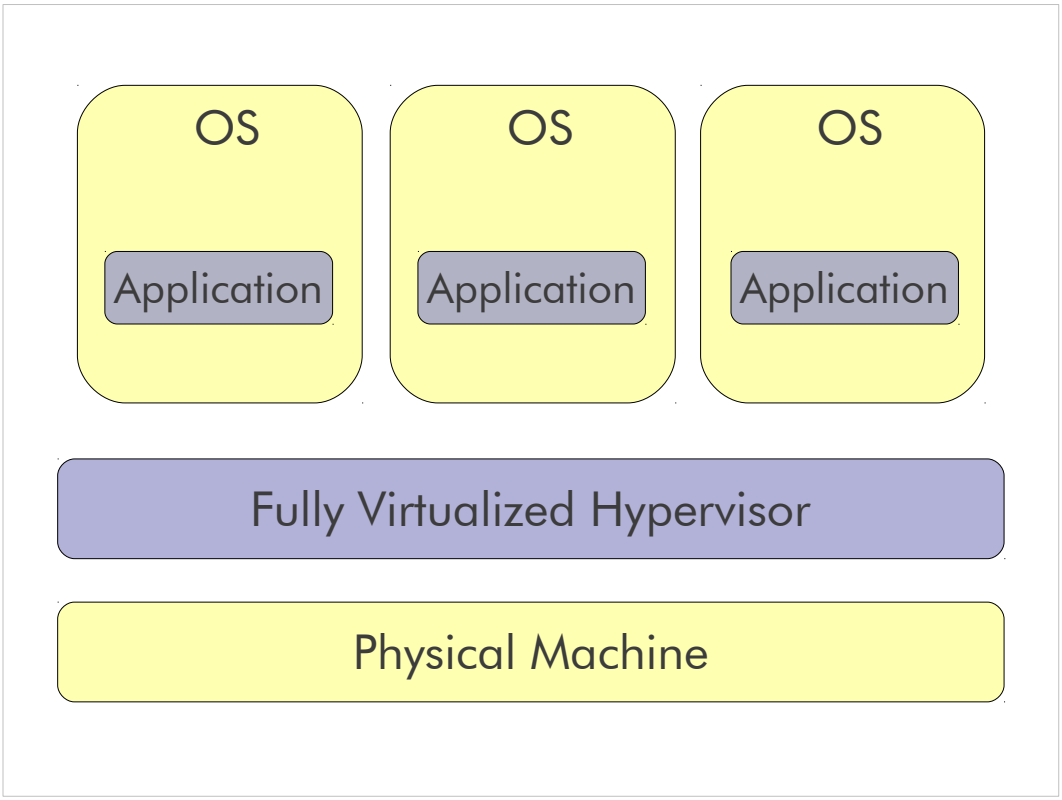
There are many types of virtualization. Here we talk about a few so we can expand on them later in the presentation.

The main point of this slide is to get people to understand that there is more than just machine virtualization, like there is more than one type of wine, and one type of wine glass.



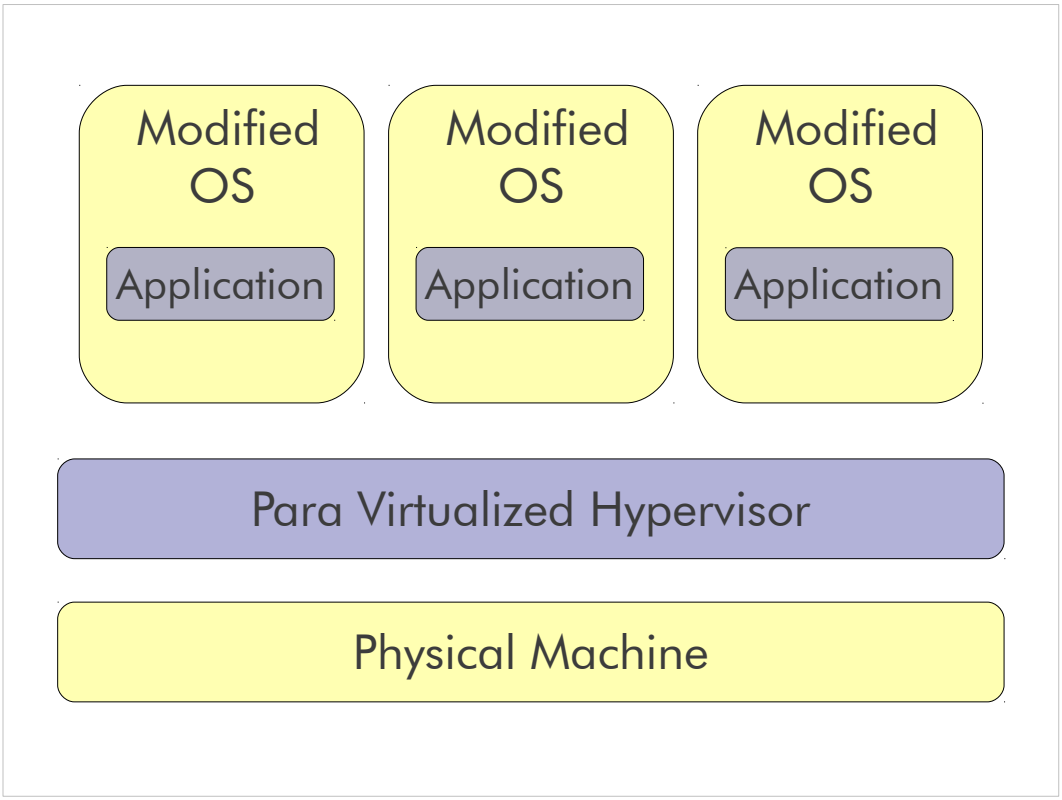
System p LPAR
System z LPAR
Sun Domains,
HP nPartitions
HP vPartitions

Physical machine virtualization. The hardware and/or firmware provides machine separation.



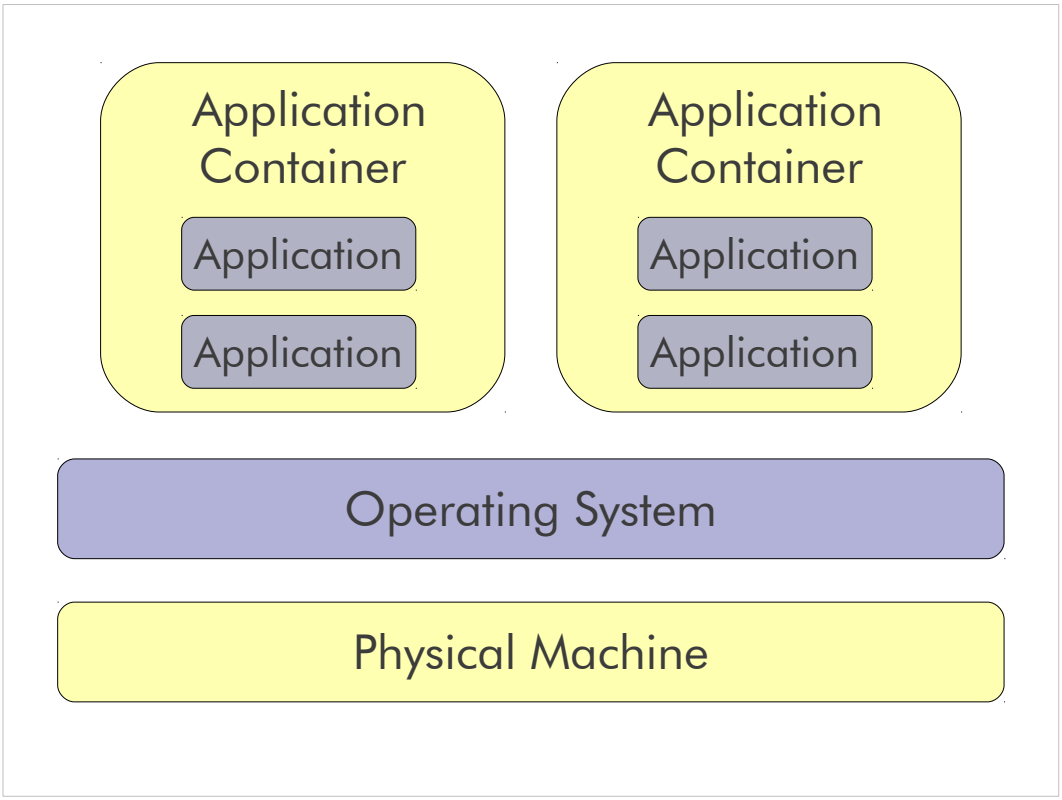
The hyper visor is normally a fully functional operating system. Each of the virtualized operating systems are none modified and do not know that they are nor running on bar metal.

VMware.



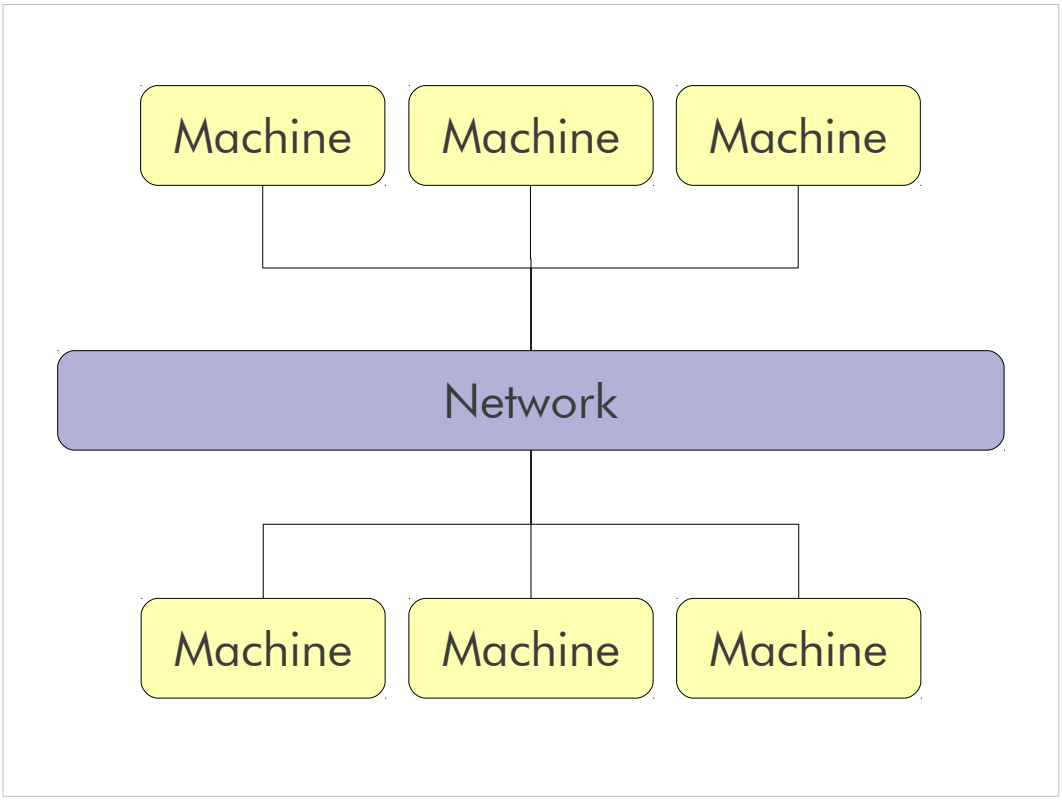
Para virtualization requires that the virtualized operating system is modified.

Xen
KVM

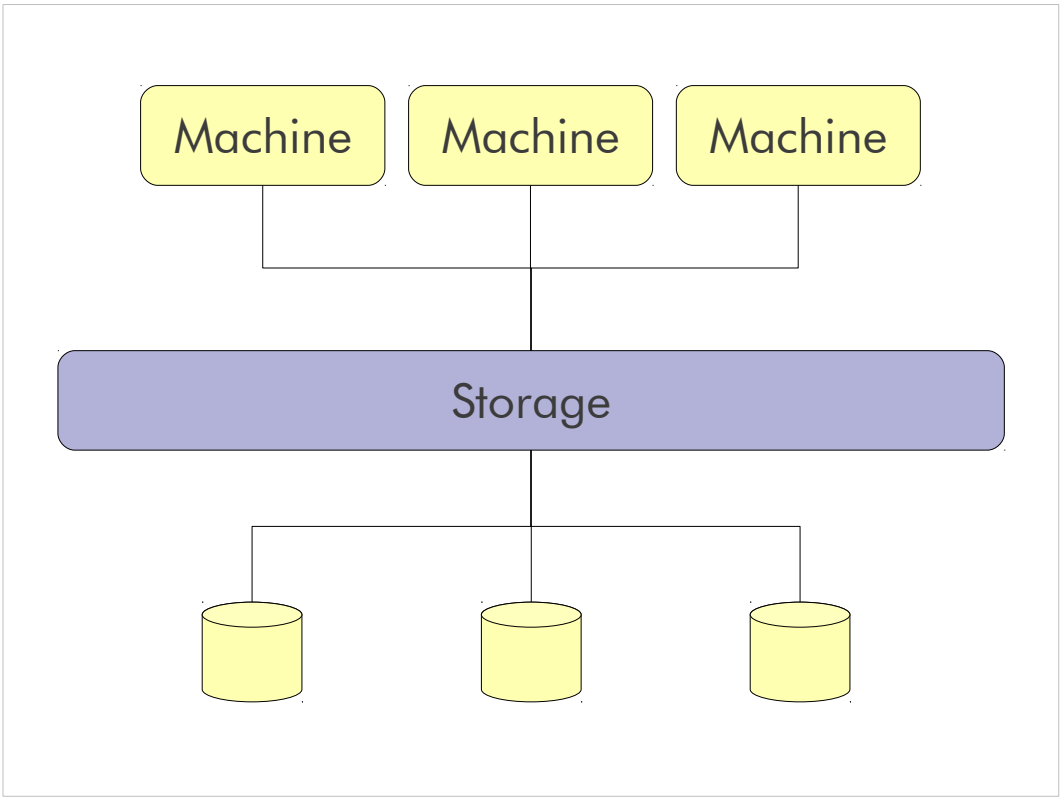


The operating system may or may not be virtualized, but in this case, the virtualization has moved up the stack to the application level. This is interesting because it can reduce the amount of system management. For example, there is only one version of the operating system, but there is 4 applications.

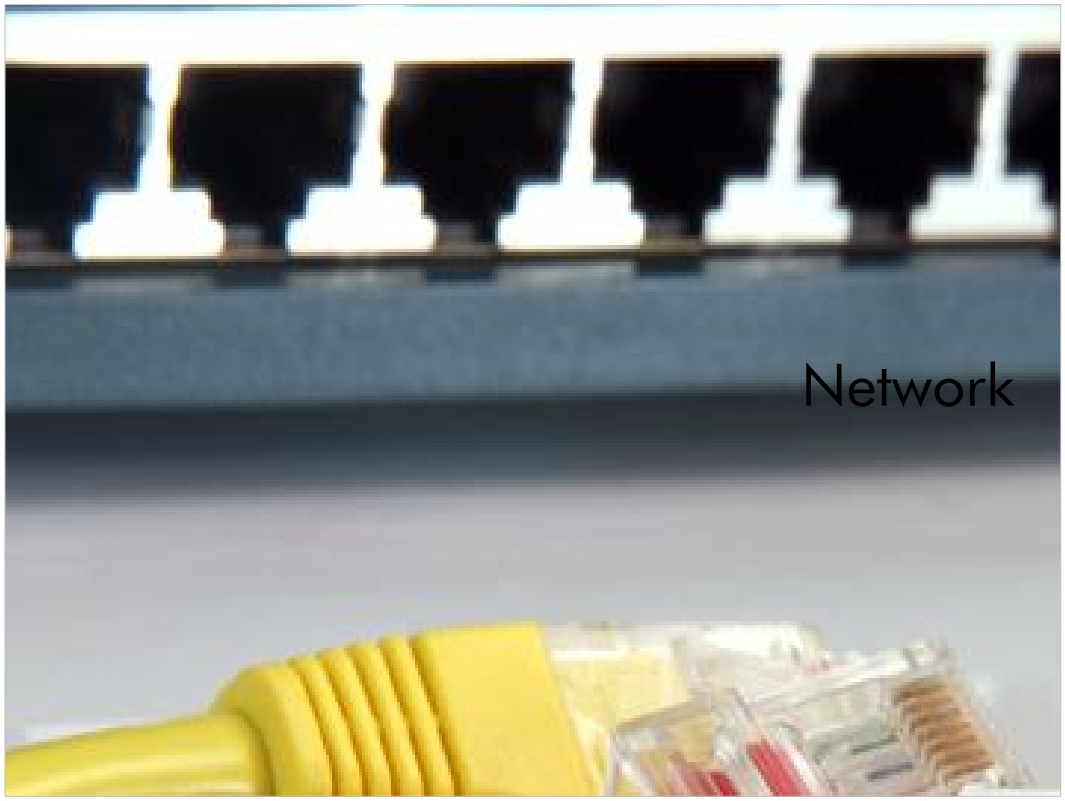
WPAR
Chroot jail
Vhost

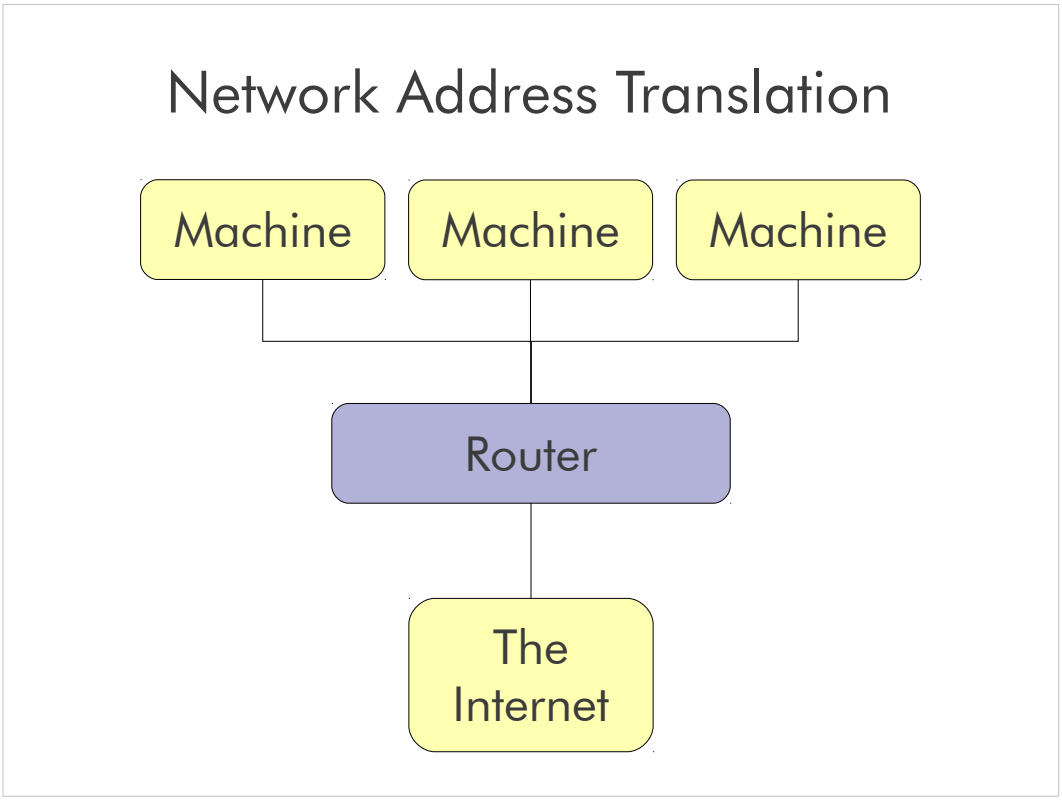


Each of these machines may be on different networks. With network based virtualization, these networks scan exists on the same physical wire. Or, not on a wire at all.



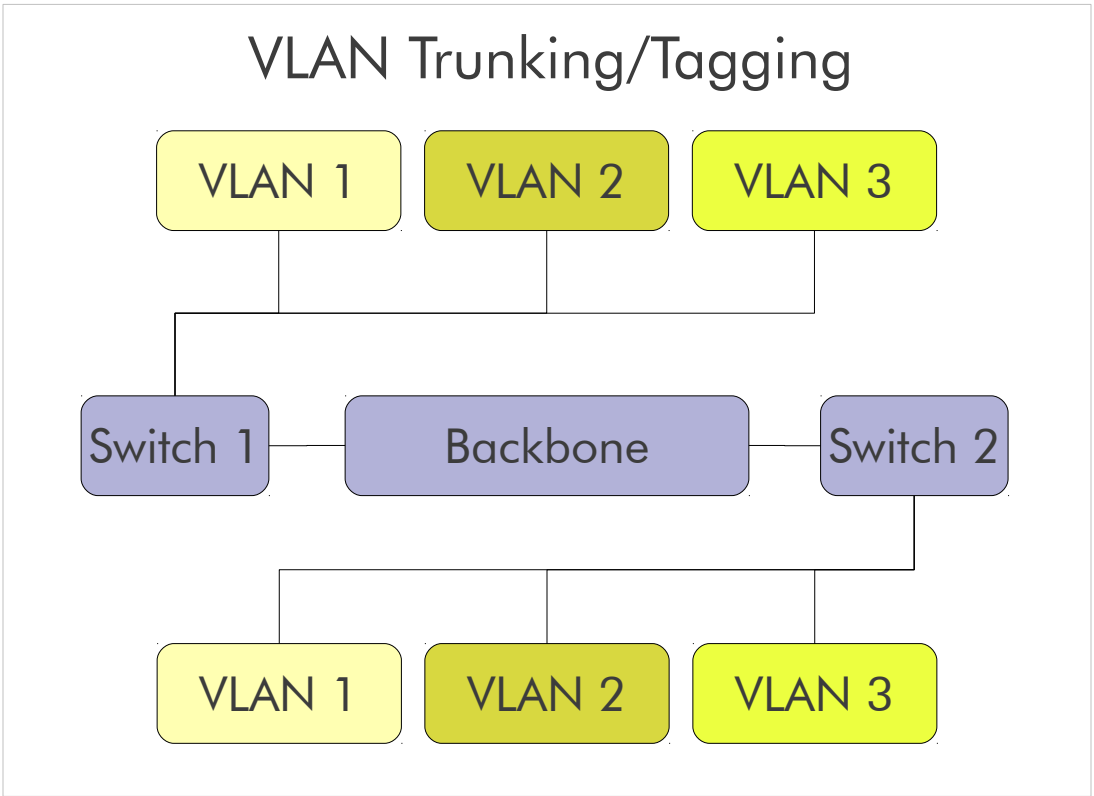
Storage virtauliztion, is separating the physical storage from the operating system. SAN sort of does this, but this is really to show LVM or SVC.


Network



Everyone uses this if they have a WAP/Router at home. NAT is used to reduce the number of IP's needed.

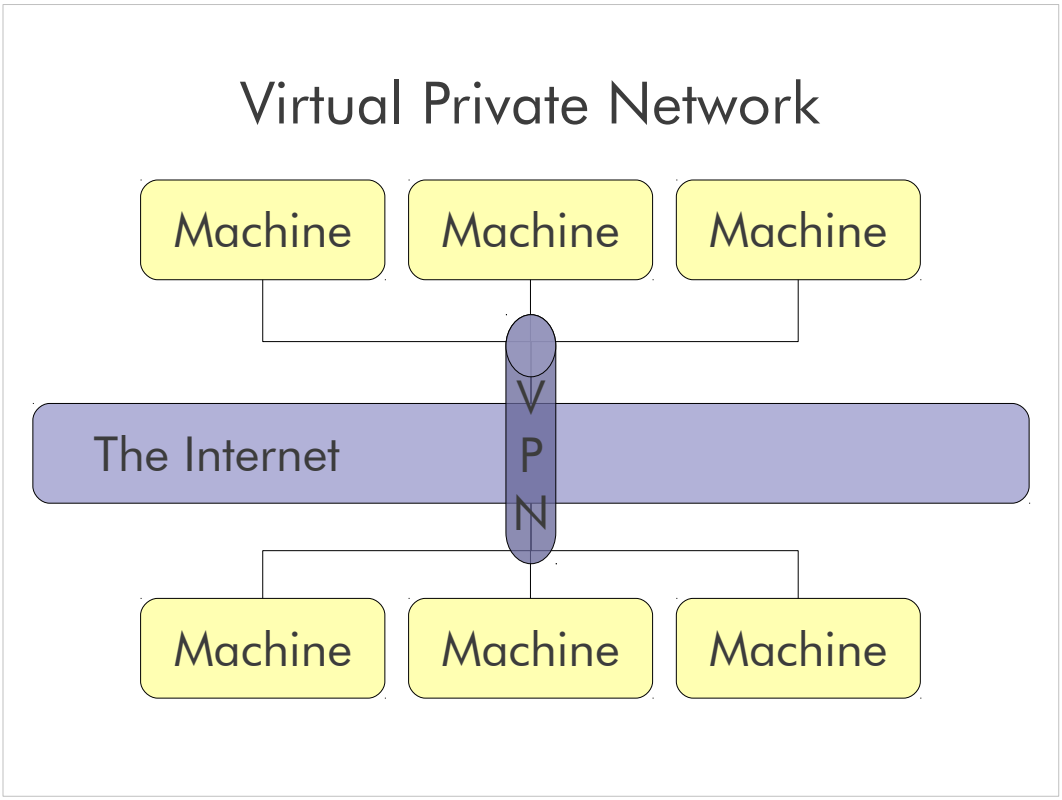
This chart is to start the discussion with something people are familiar with.



VLAN Tagging is used when a link needs to carry traffic for more than one VLAN.

Trunk link: As packets are received by the switch from any attached end-station device, a unique packet identifier is added within each header.

This header information designates the VLAN membership of each packet.

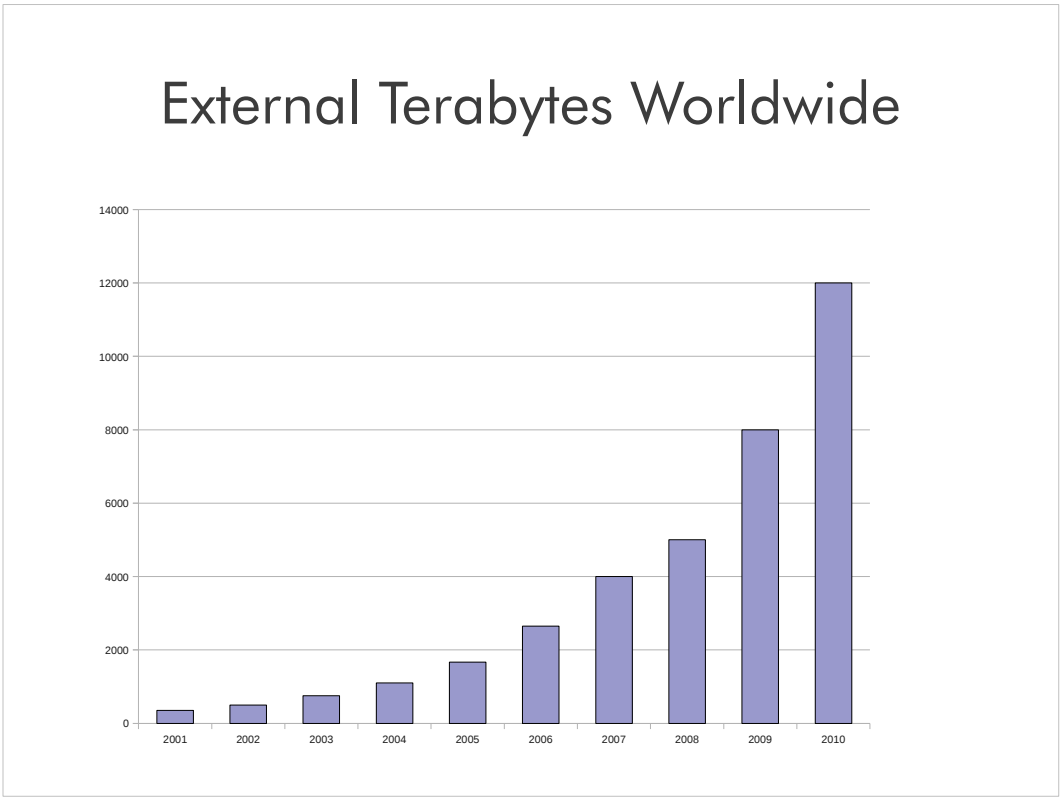


This even has virtual in the name. Allows people to connection from a one or more machines securely over an unsecure connection, such as the Internet.

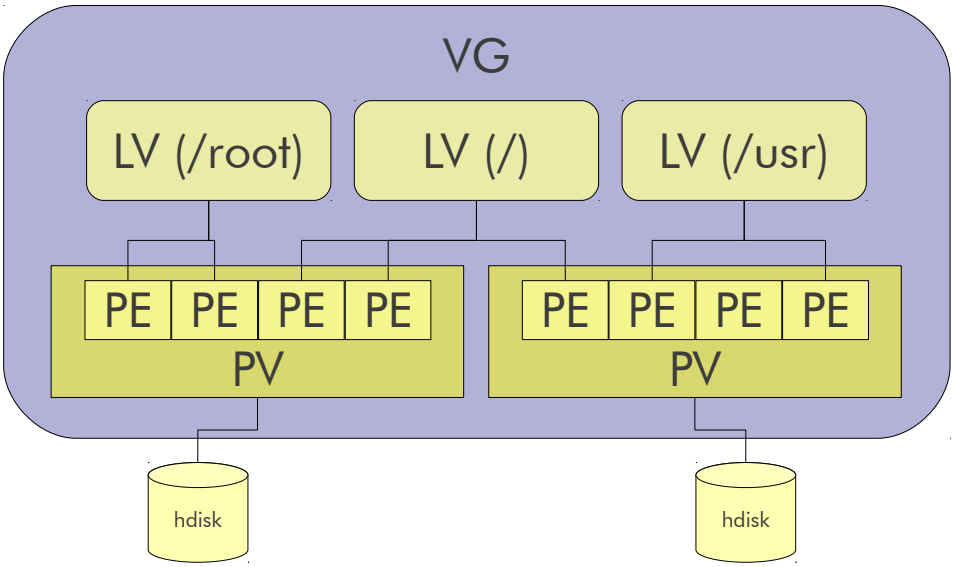
Storage is often forgotten by many of the customers I talk to. They just need a few terabytes here and there and don't care how it gets attached.

This is especially true in the distributed environments. Blades and 1U racks that have internal storage lead customers to believe that that is all they need... and it leads to very bad practices.

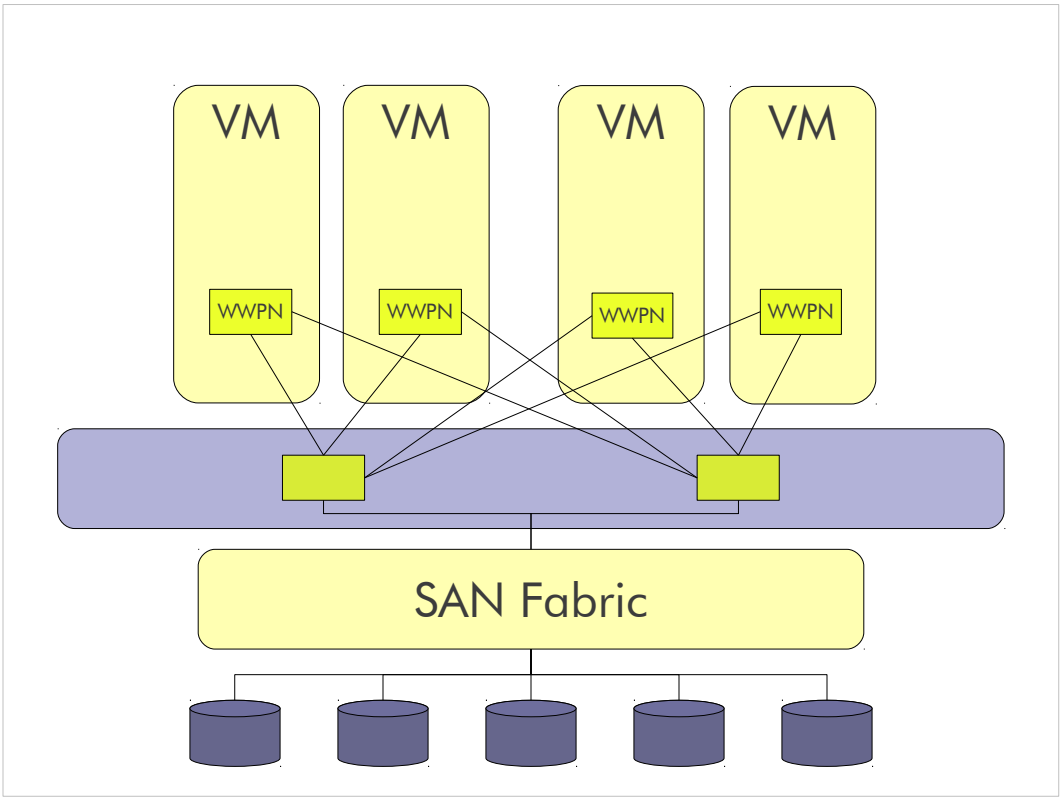
Virtualization of storage helps achieve location independence by abstracting the physical location of the data. The virtualization system presents to the user a logical space for data storage and itself handles the process of mapping it to the actual physical location.



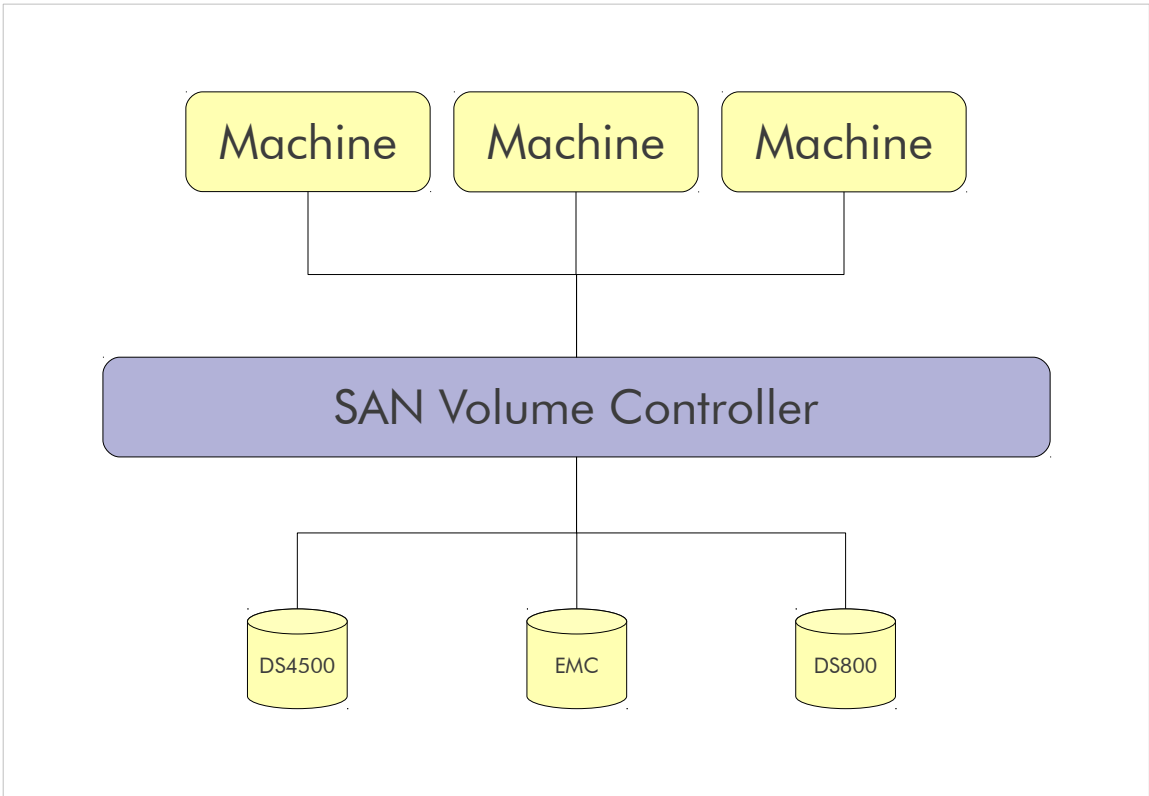
External terabytes are growing at an exponential rate. this char from the IDC in 2007. More and more data is being stored separately from the machines that use said data.



Volume Group
Logical Volume
Physical Volume
Physical extend
Physical disk



NPIV is a facility allowing multiple N_Port IDs to share a single physical N_Port. This allows multiple Fibre Channel initiators to occupy a single physical port, easing hardware requirements in Storage Area Network design, especially where virtual SANs are called for.



By using virtualization each host can be allocated virtual storage space according to its needs. The actual usage of physical disk space occurs independently of the allocation.

The next chart explains the inner workings of the Virtualization Engine shown here.

This chart explains how the Virtualization Engine works.

Transition 1

The SCSI LUNs still have a one to one mapping to what they perceive as the host, but is actually the Managed Disks on the Vitualization Engine.

Transition 2

The Hosts still have a one to one mapping to what they perceive as the storage LUNs, but are actually the Virtual Disks on the Virtualization Engine.

Virtual Disks: Associated with a Managed Disk group. Virtual Disks are created from extents in Managed Disk group.

Virtual disks appear as SCSI LUNs supporting a SCSI command set

User creates Virtual disks and maps to Hosts

Transition 3

The Virtualization Engine controls the mapping of all the groups of Ones and Zeros (these groups are called Extents) between the Virtual Disks and the Managed Disks.

Managed Disk Group: Pool of extents, number of available extents defined by number and size of Managed disks included in the Managed Disk Group

Transition 4

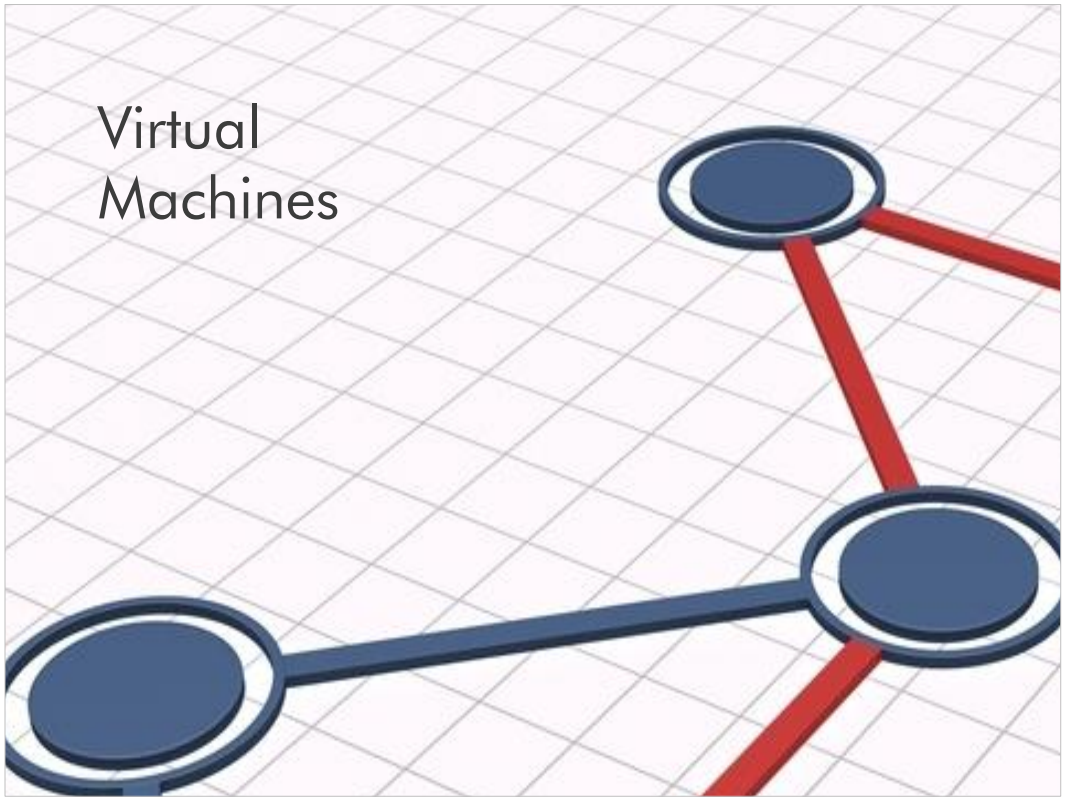
The Managed Disks are collected into Managed Disk Groups to facilitate different of categories of storage devices, in this case High Performance storage relative to Low Cost storage

User defines Groups

Transition 5

A key point is at the bottom in blue - the administrator can choose where the mapping to managed disks is done by striping across multiple disks or disk arrays, or by sequentially grouping the data from a given Virtual Disk onto a given Managed Disk, or in fact having a one to one mapping via Image mode which is very valuable in migrating data from a non-virtualized environment into the virtualized environment without having to migrate the data from one physical device to another.

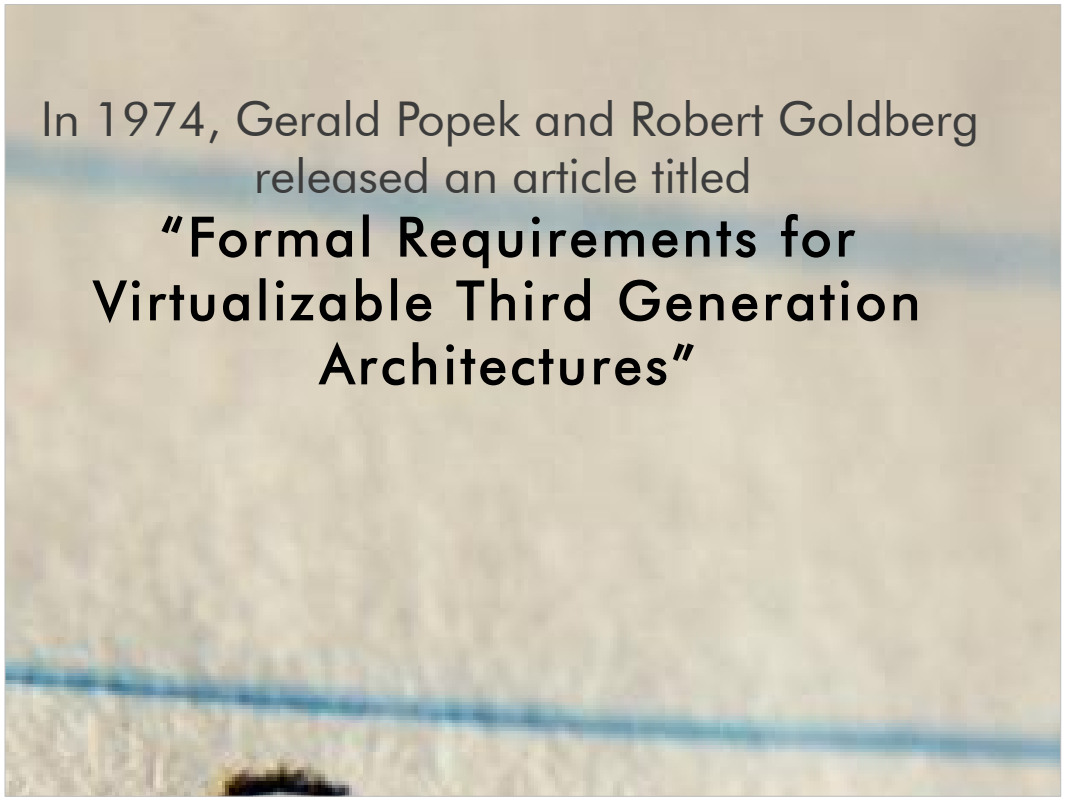
Automatic discovery



The software behind virtualization technology is the virtual machine monitor

- The monitor sits above and abstracts the system hardware
- Conceptually guest operating systems interact with the virtual machine instead of directly with the hardware

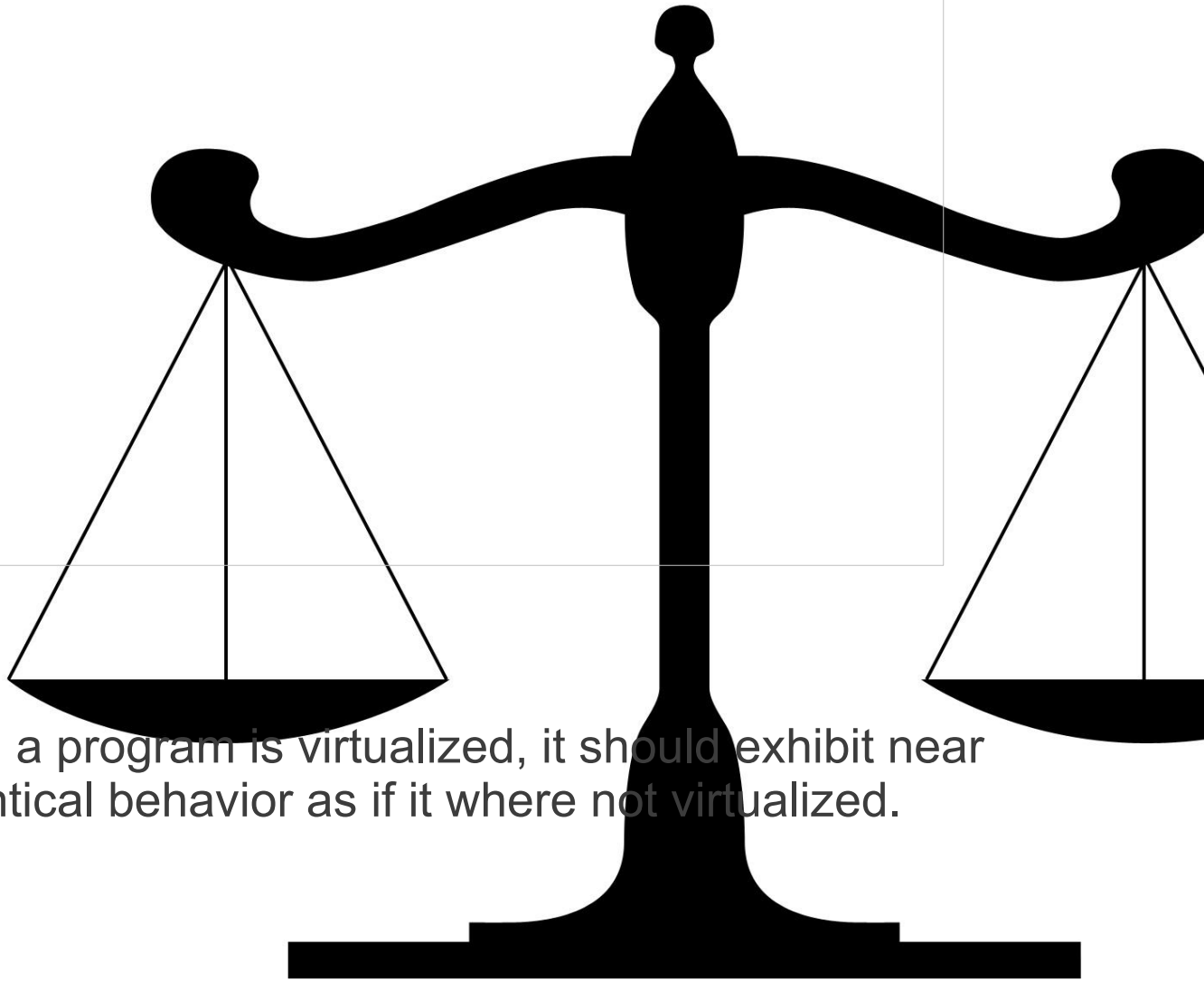
The VMM is the software behind the virtual machine
It hosts multiple guest OS instances
Each instance gets its own virtual cpu, virtual memory, virtual disk, etc.



This paper talks about the requirements to provide virtualization at the hardware level.

The paper explained that for 3rd generation (s360, Dec PDP 10) the instruction sets must allow the following characteristics...

Equivalence

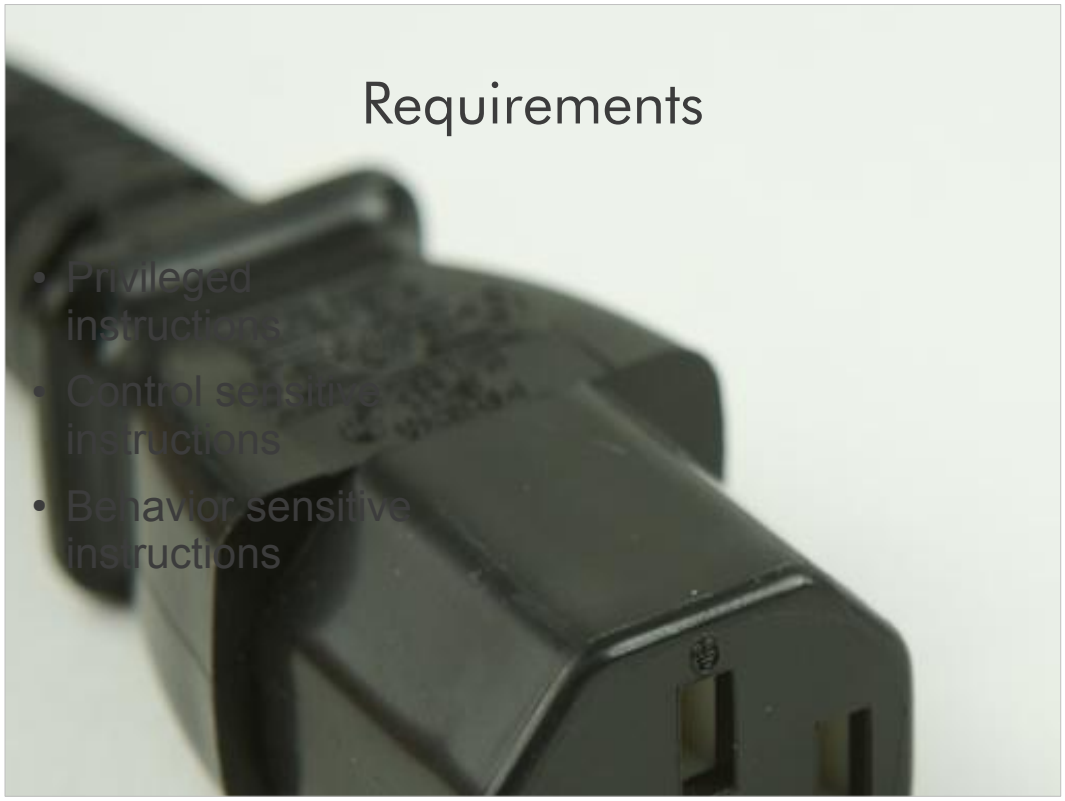


When a program is virtualized, it should exhibit near identical behavior as if it where not virtualized.

Resource Control

They hypervisor, or whatever is doing the virtualization, should have complete control over the the physical resources.

Efficient virtual manager should not used a large portion of the physical resources to allow the virtualized processes to run.



Privileged instructions

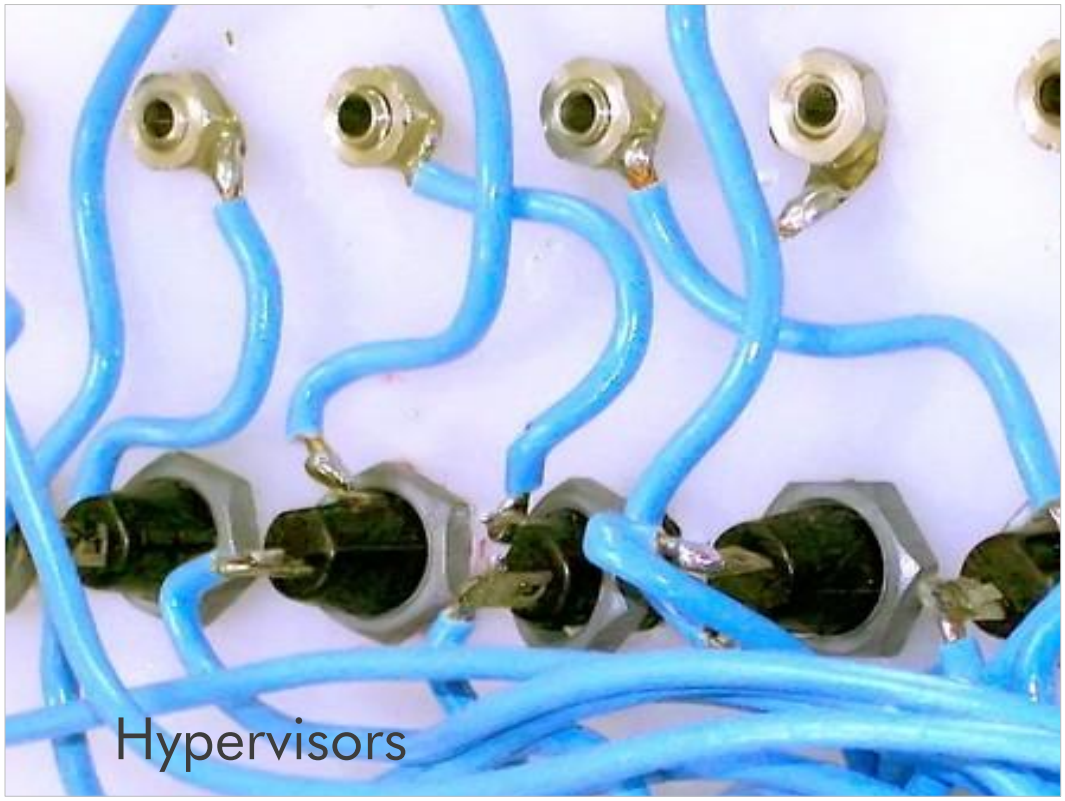
Those that trap if the processor is in user mode and do not trap if it is in system mode.
While in user mode, if there is an exception, such as divide by zero, a context switch returns control to the kernel.

Control sensitive instructions

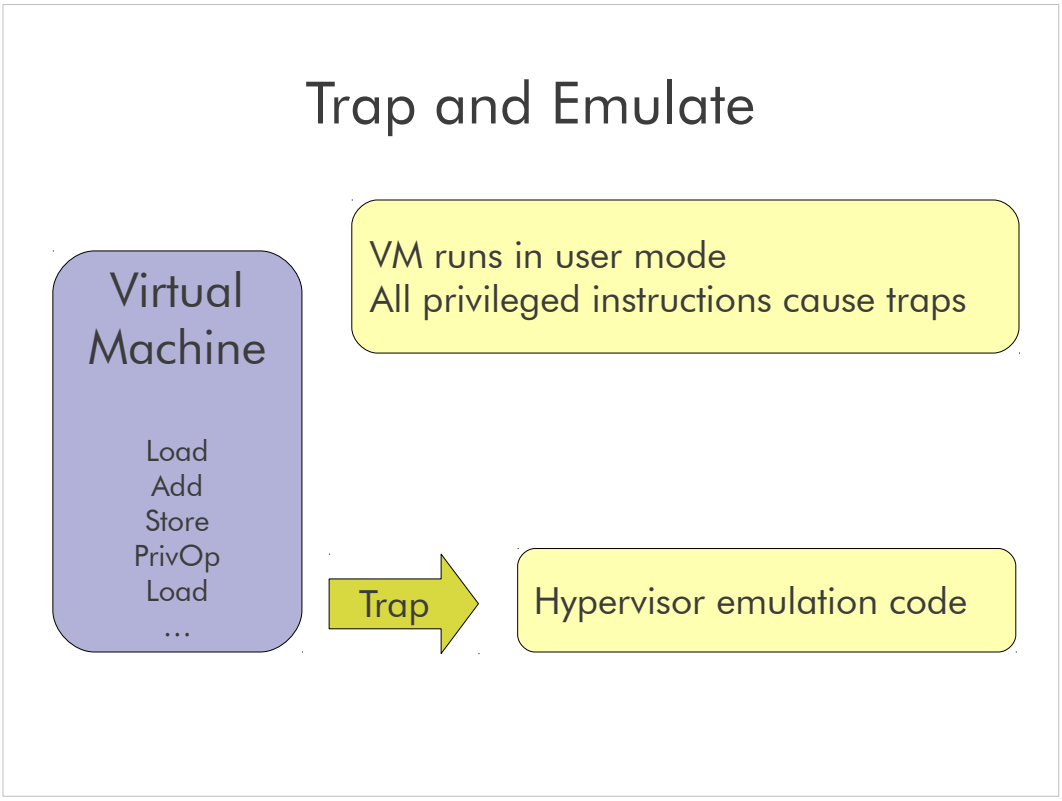
Those that attempt to change the configuration of resources in the system.

Behavior sensitive instructions

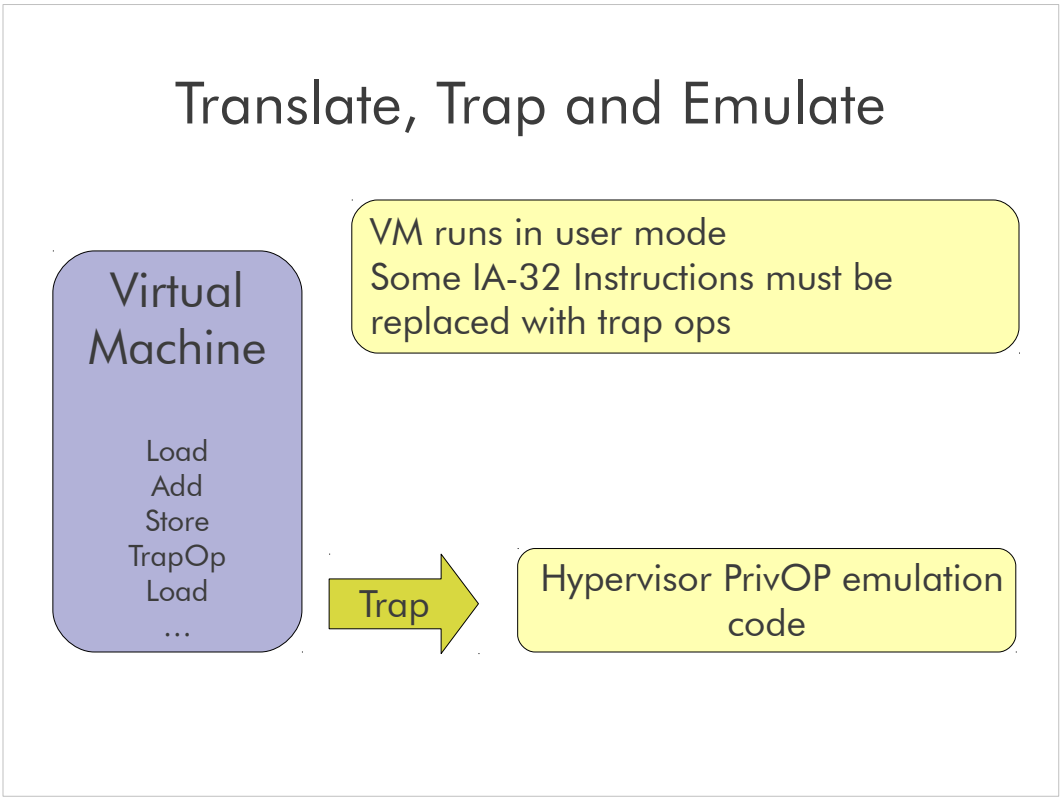
Those whose behavior or result depends on the configuration of resources (the content of the relocation register or the processor's mode)



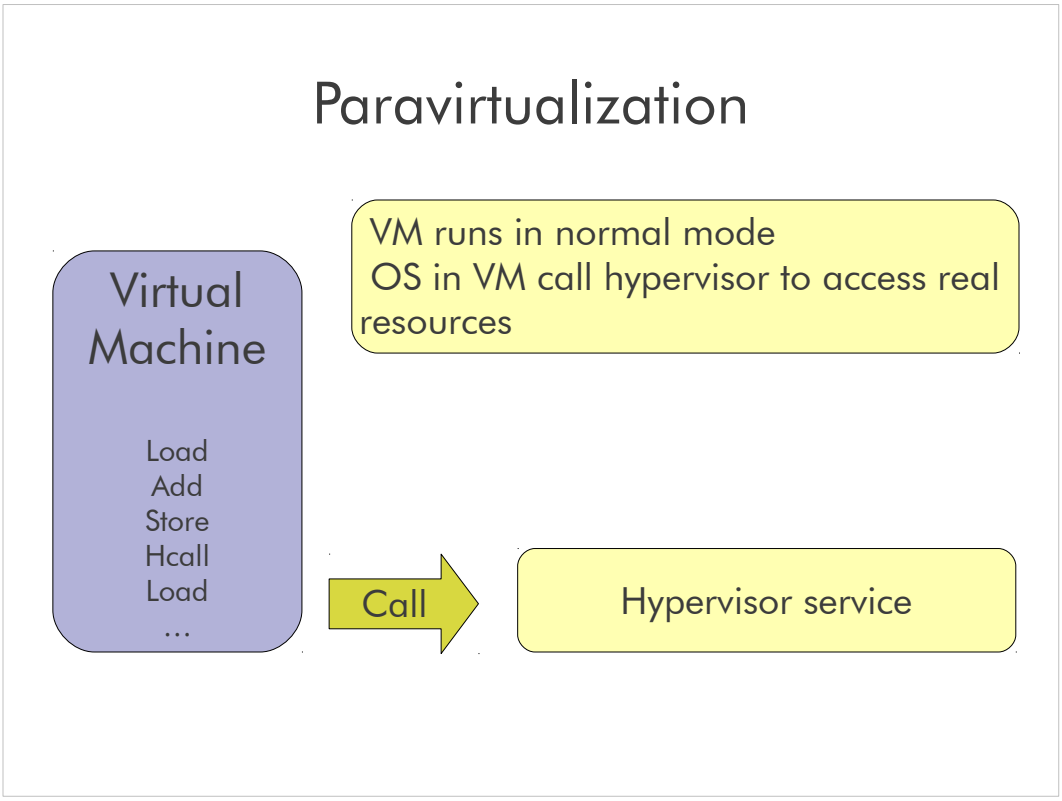
Hypervisors allow multiple operating systems to run on the same physical hardware.



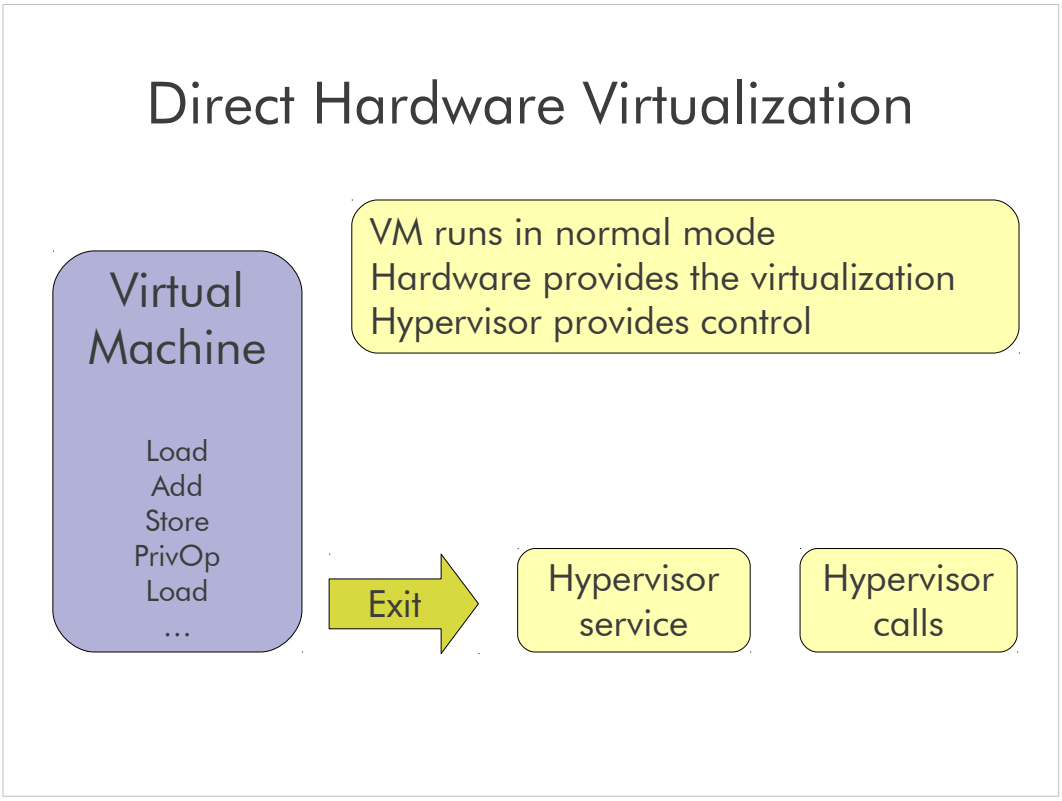
Examples: CP-67, VM/370
Benefits: Runs unmodified OS
Issues: Substantial overhead



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



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



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

System x (Intel)

Virtualizing Intel

- 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

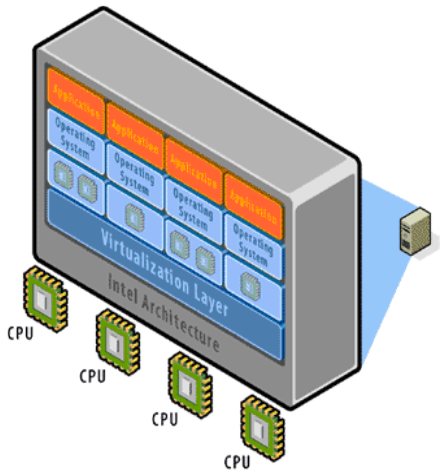
The Intel IA-32 architecture was never designed to be virtualized and this causes complications.
Compared to IBMs well-designed virtualization architecture the IA-32 presents as a poor contender.
However, the IA-32 is by far the most widely available and so virtualization is still demanded.

Example complications are non-protected privileged instructions and enormous I/O requirements.

- VMWare is one of the most popular full system virtualization tools available
- Supports both a hosted environment approach and a hypervisor approach
- For performance enhancements operating system drivers are installed by VMWare

VMware Virtualization

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



The processor is virtualized by using direct execution on the processor

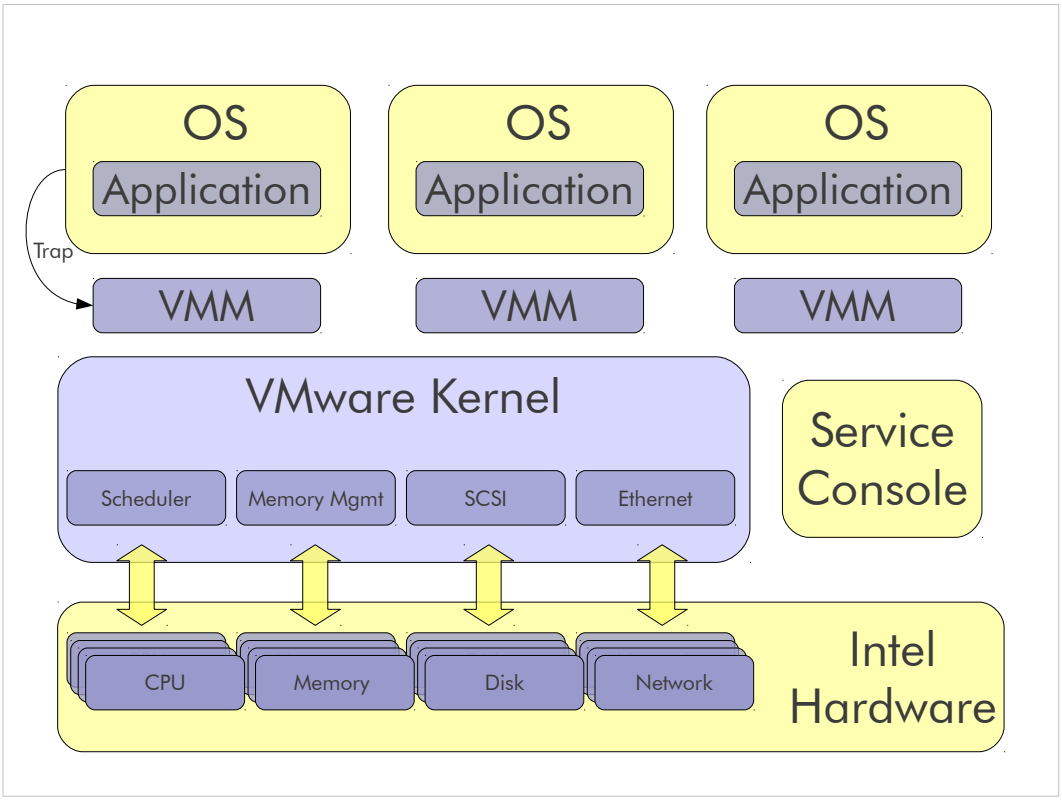
- Combined with binary translation to eliminate problem instructions
- Results in very performance only “slightly” lower than paravirtualized approaches

Memory is virtualized using the very straight forward shadow table approach

- Additionally a special ballooning driver is installed in each guest operating system
- This trick gives the VMM insight into page usable inside of the guest

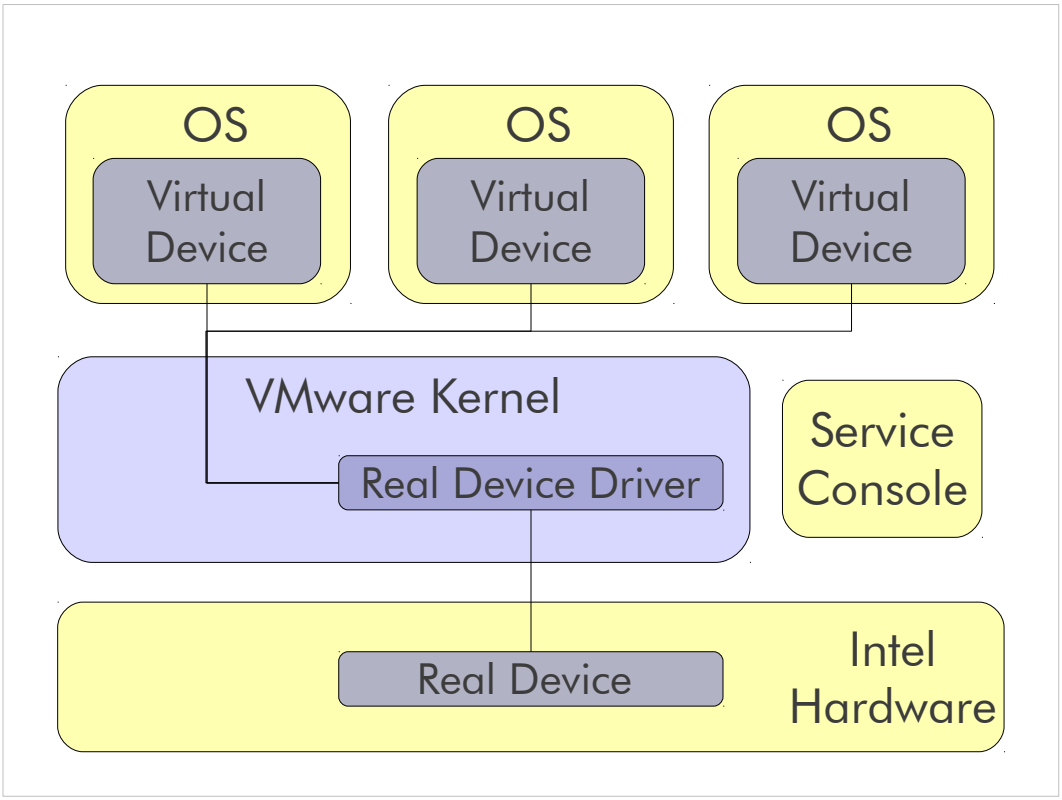
Device I/O is virtualized in one of two different ways

- The hosted architecture relies on the existing host for I/O support
- The hypervisor architecture supports only a limited number of “certified” devices



Main points here. The Service console is Linux like. Really it is a bunch of GNU bin-utils. The kernel is NOT Linux, but a special proprietary kernel.

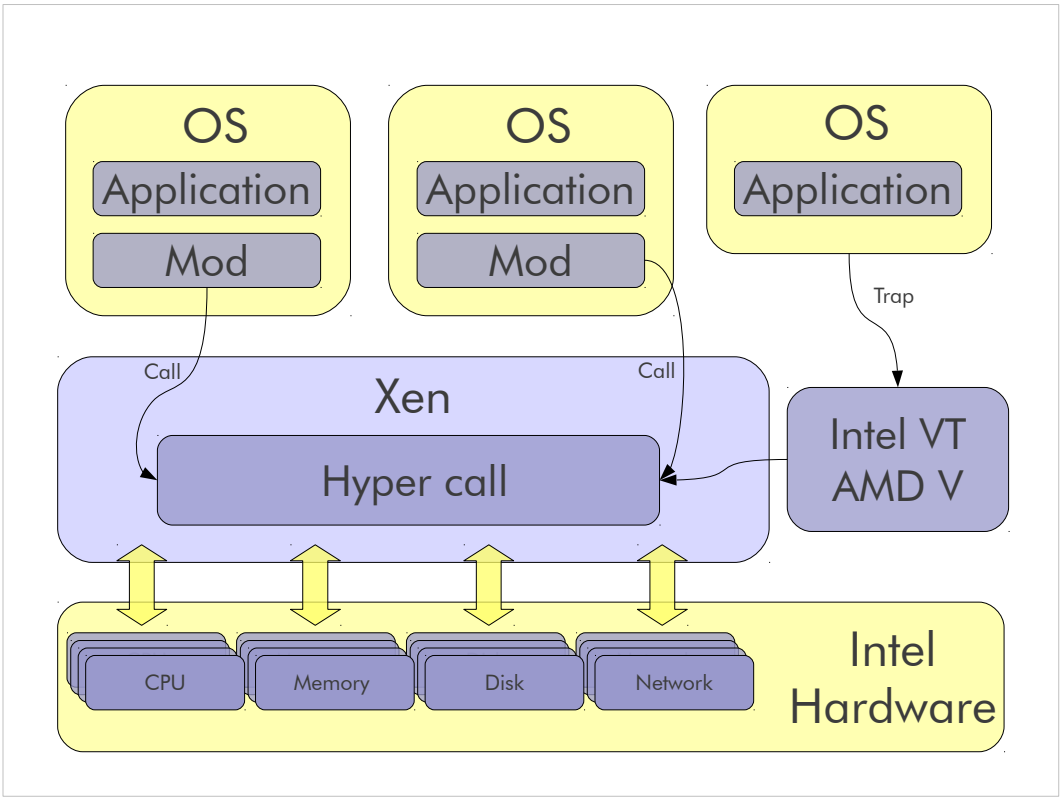
The virtual OS traps the the VMM which is the connection to the Kernel. The kernel then talks to the underlying hardware.



The device must be supported by Vmware, and must be virtualizable. This could cause problems for older legacy devices.

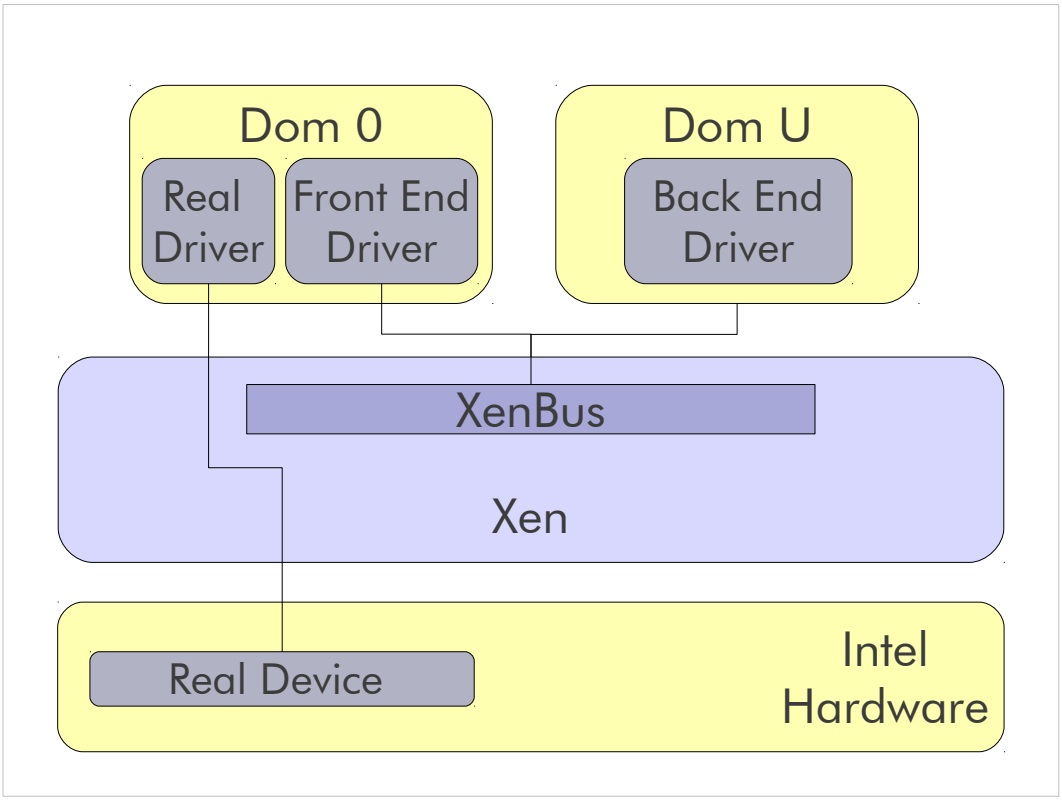
I was working with a client that wanted to use Vmware to run a number Windows 95 images on new Intel machines. The only issued they ran into was the legacy hardware the Vmware did not support.

- Open Source virtualization software solution based on Linux
- Uses paravirtualization to abstract CPU, memory, and I/O resources
- Guest operating systems are responsible for allocating and managing page tables
- Management and control software runs in Domain 0
- IVT and AMD-V enables hosting of unmodified guest operating systems
- For Windows support on Xen, users need IVT-capable hardware
- IBM is actively contributing to the Xen open source project
- Xen source is now owned by Citrix



You can either have a dom u that has a xen modification, or fully virtulalized dom u on hardware that supports it.

This chart does not show dom 0, but it is implied by the box that contains the hyper call



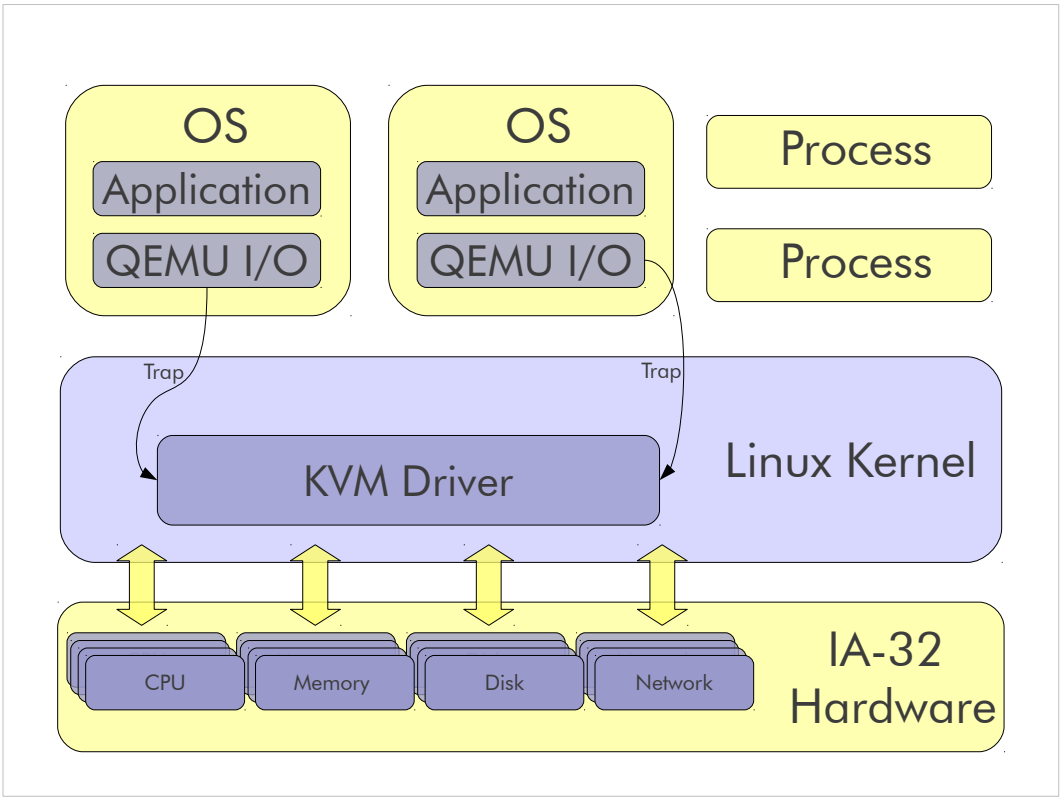
Xen has a split driver model.
Real device drivers are loaded into Dom0 as well as a generic front end device.
Generic back end drivers are loaded into each DomU

KVM is maintained by Avi Kivity and is funded primarily by Qumranet, a technology start up,now owned by Red Hat.

Supports AMD-V, Intel VT, z9 and above, ppc64

Unlike Xen, which is not mainline kernel, KVM is.

IE Ubuntu doesn't ship a xen enabled dom0



Normal Linux processes have two modes of execution: "kernel" and "user"

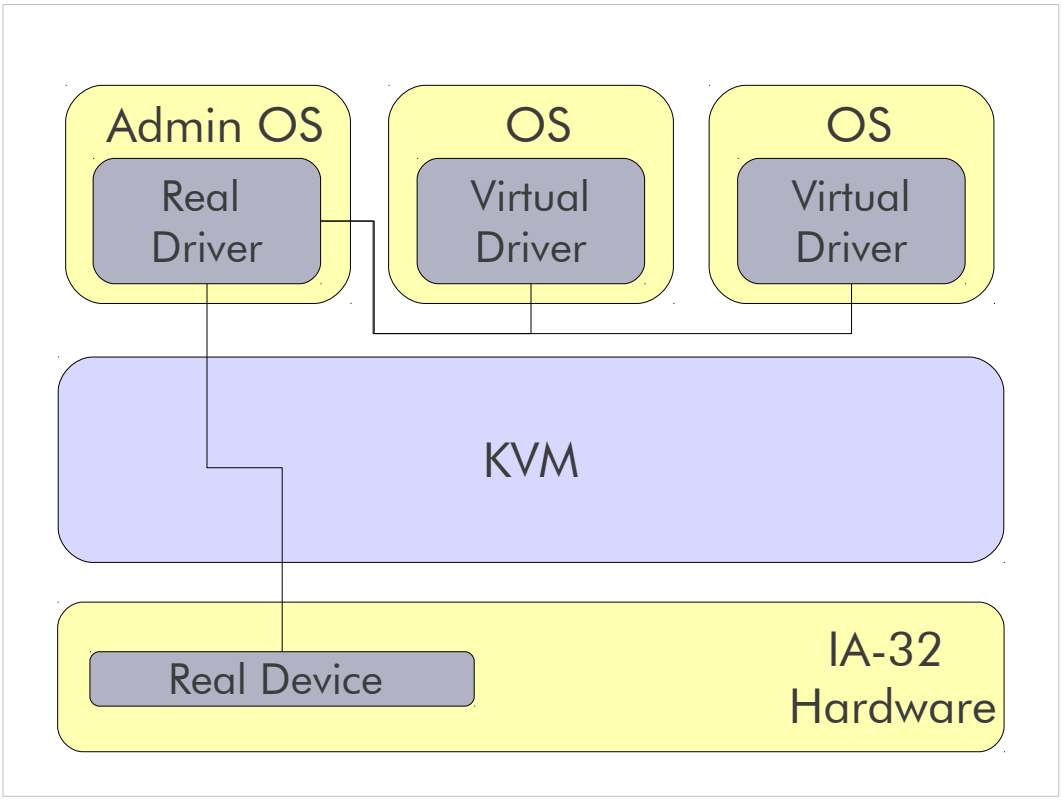
KVM adds a third mode: "guest"

KVM virtualizes CPU, I/O Advanced Programmable Interrupt Controller (IOAPIC), and Memory Management Unit (MMU); requires IVT or AMD-V

QEMU is a user space component that emulates PC hardware

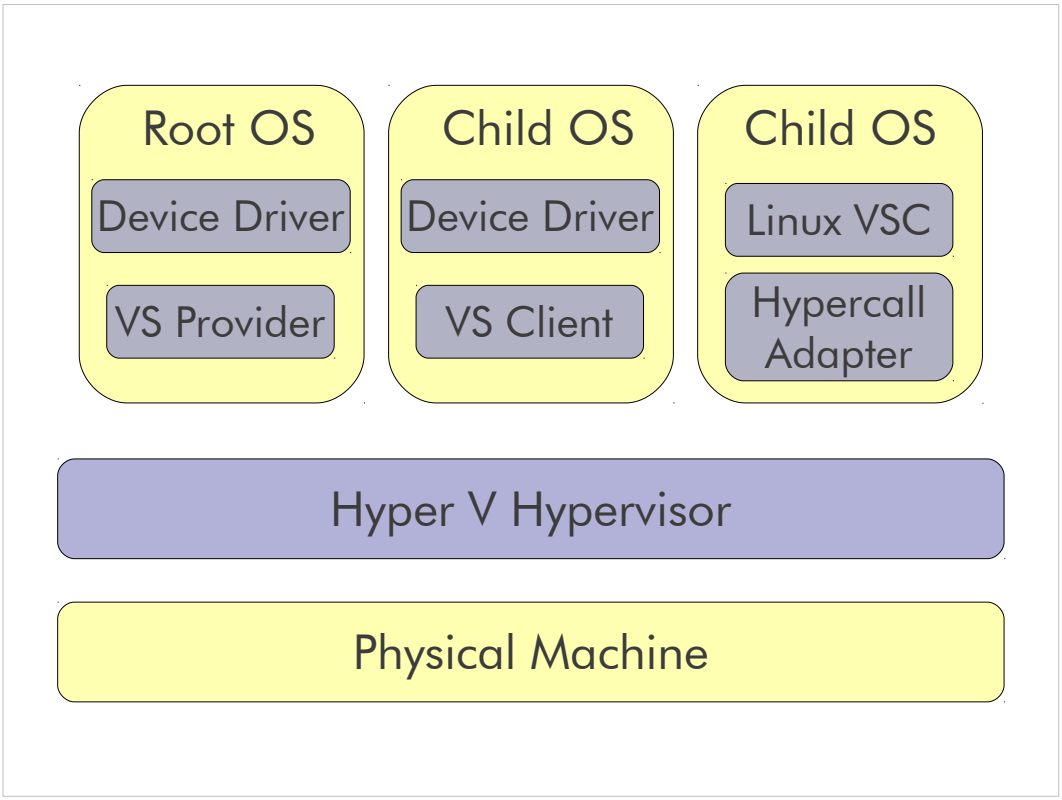
KVM gives QEMU near-native CPU virtualization

Each virtual machine is a normal Linux process

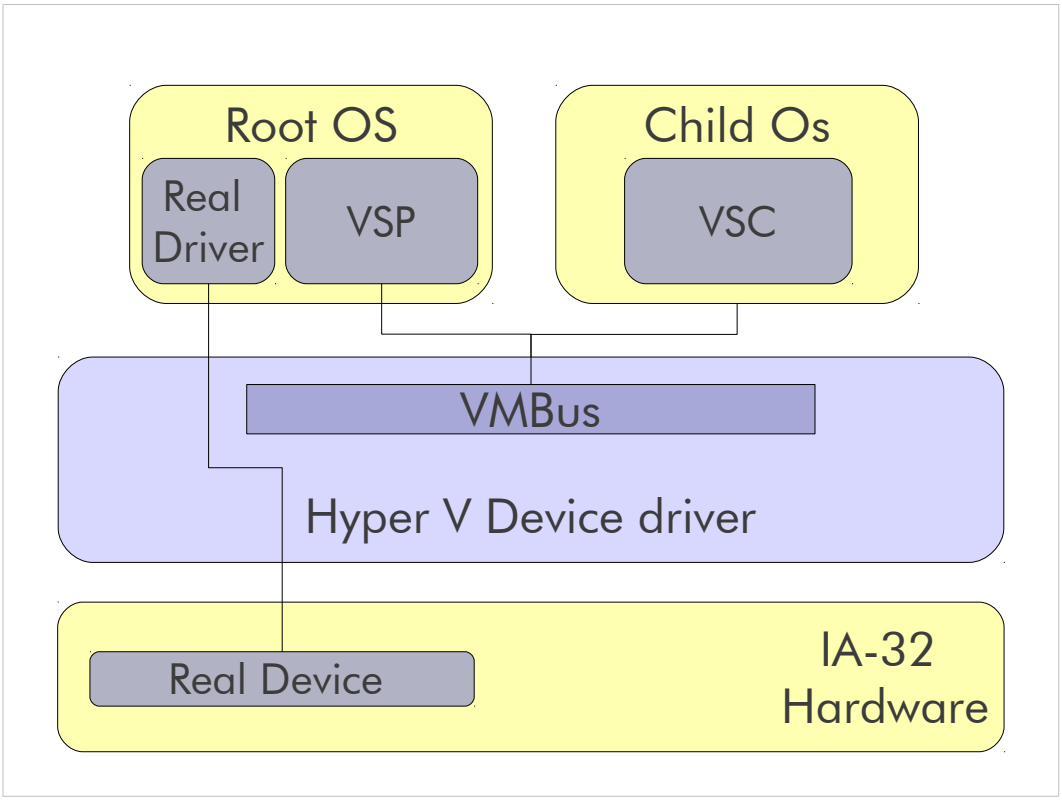


KVM has some para-virtualized device drivers for windows, this allows near native speed.
The rest are emulated

Formerly known as Windows Server Virtualization.



Uses a hyper call approach

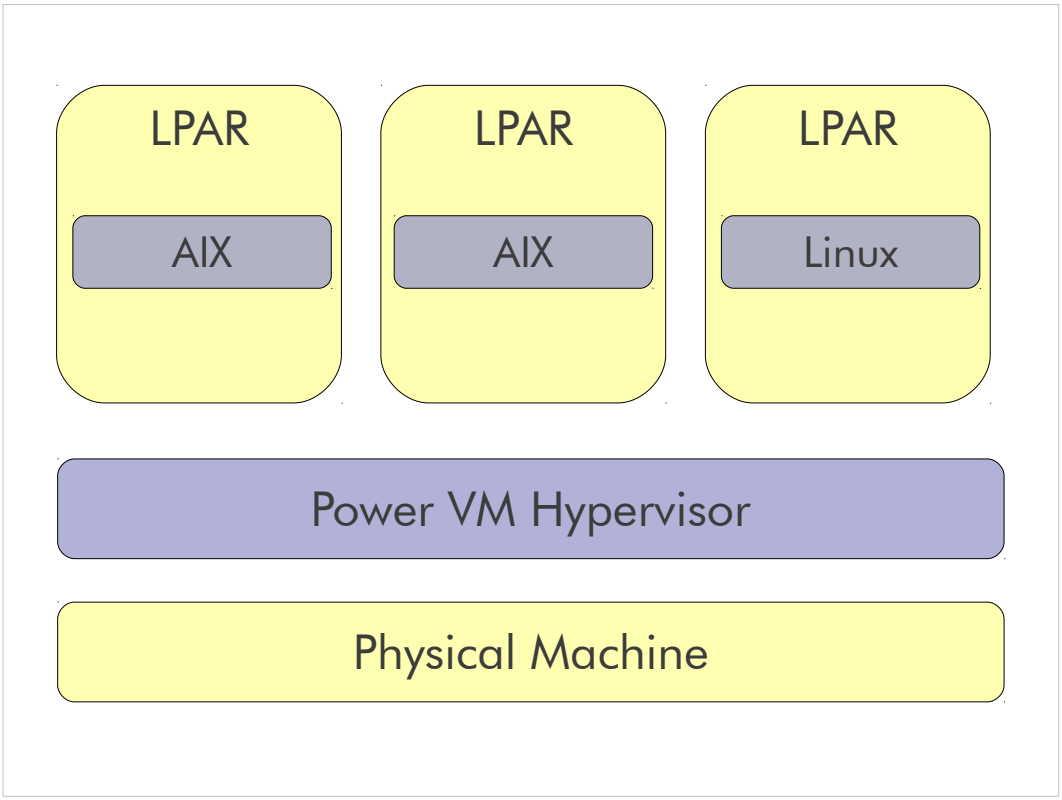


Any request to the virtual devices is redirected via the VMBus to the devices in the parent partition.The VMBus is a logical channel which enables inter-partition communication. The response is also redirected via the VMBus. Parent partitions run a Virtualization Service Provider (VSP), which connects to the VMBus and handles device access requests from child partitions. Child partition virtual devices internally run a Virtualization Service Client (VSC). This entire process is transparent to the guest OS.

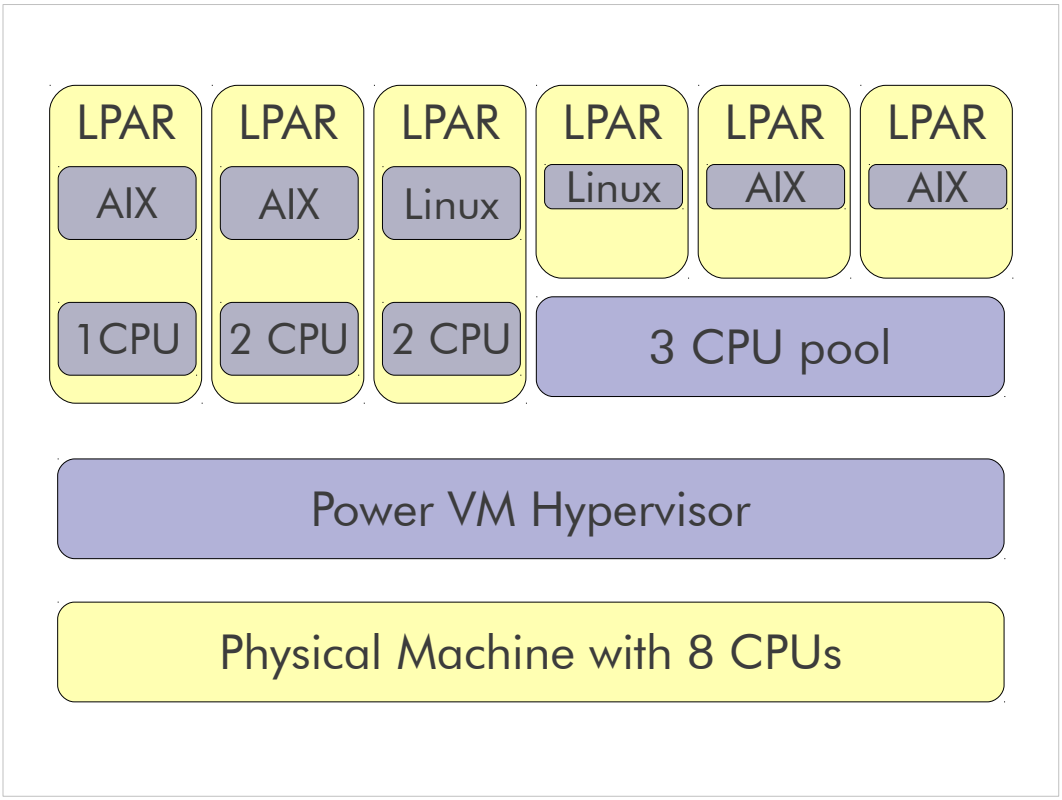
Virtual Devices can also take advantage of a Windows Server Virtualization feature, named Enlightened I/O, for storage, networking and graphics subsystems, among others. Enlightened I/O is specialized virtualization-aware implementation of high level communication protocols like SCSI to take advantage of VMBus directly, that allows bypassing any device emulation layer. This makes the

System p

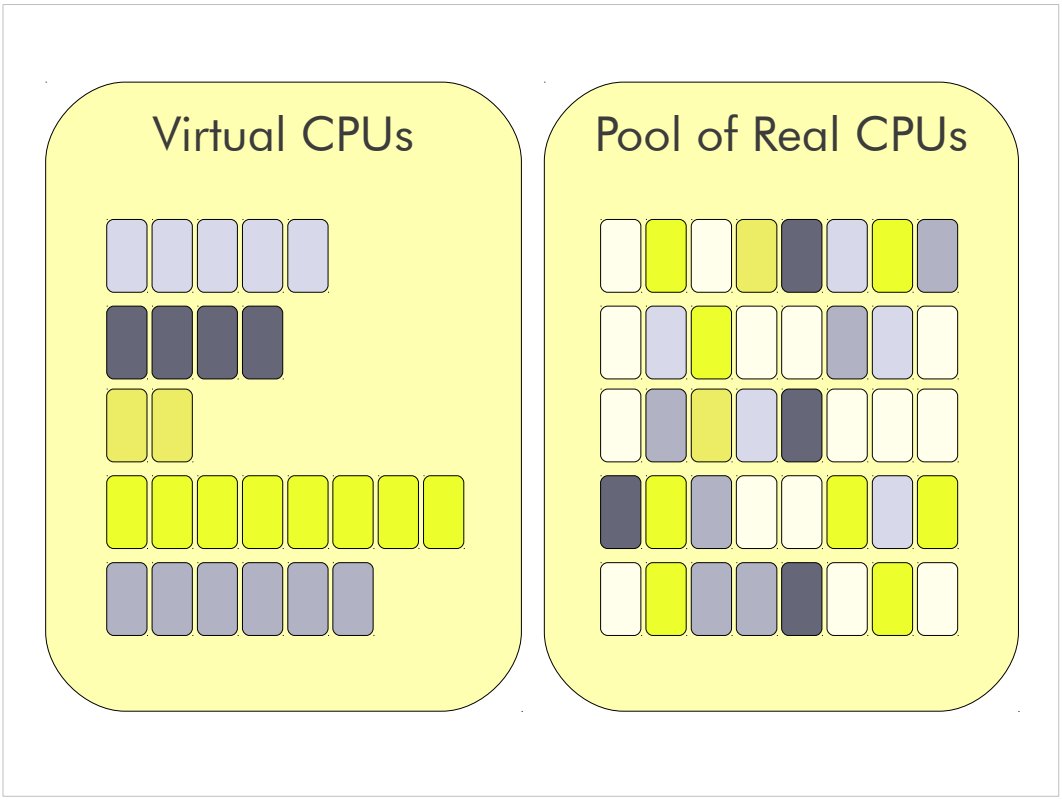
SMP guests
Up to 64 way
Driver IO is not emulated
Power4, 5, 6,7, 7+
Hypervisor
Micro lpars
Hypercalls to fireware based hypervisor



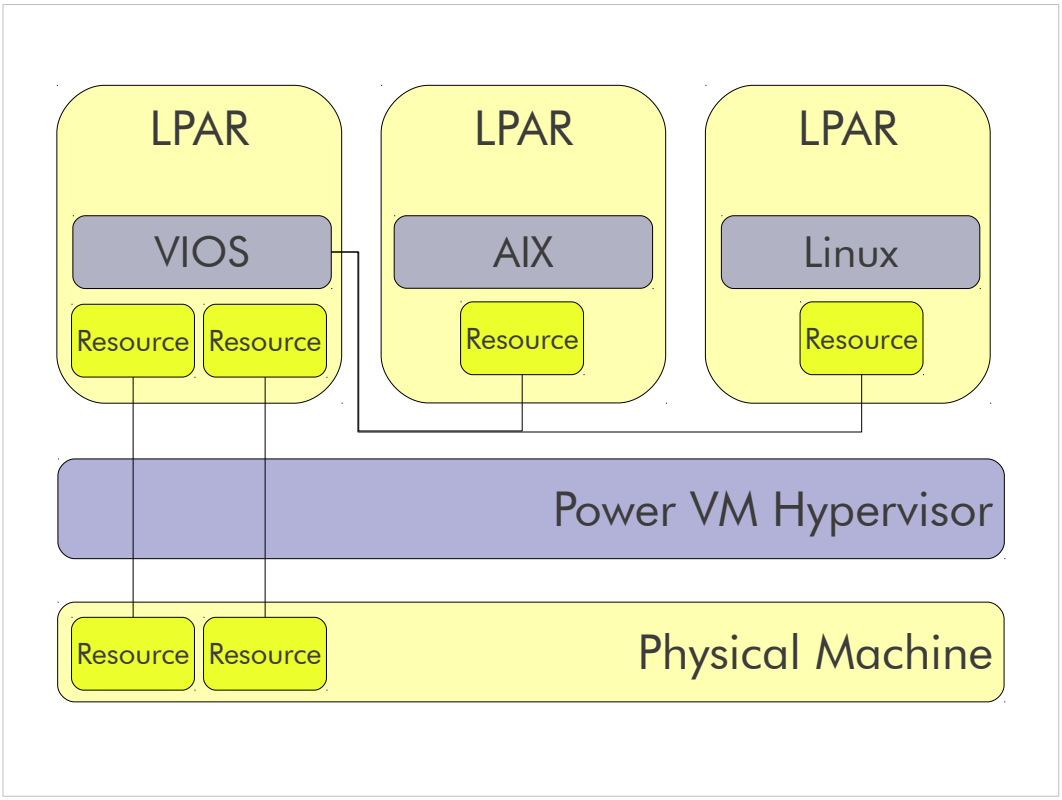
Power VM inherits many of it qualities from System z. Here we are showing physical lpar separation.



PowerVM allows sharing a pool of CPUs. There can only be a single pool, but each user of the pool can dynamically be provided more or less CPU.

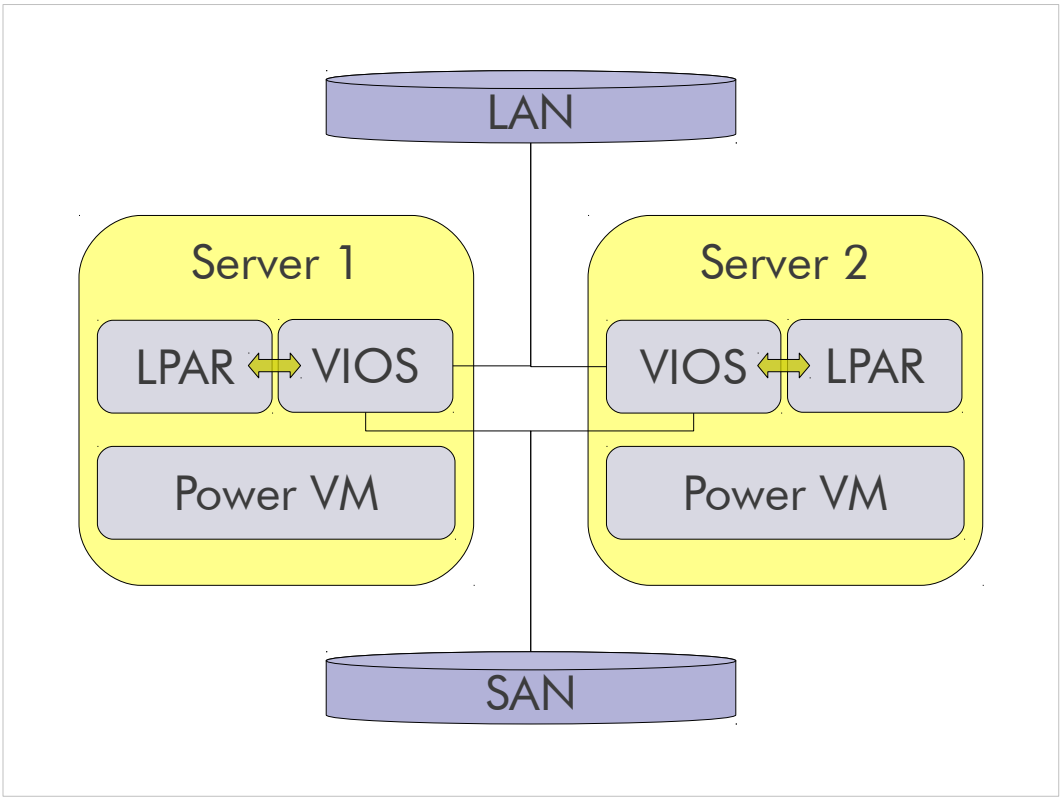


Each unit is (1/100) * Real CPU



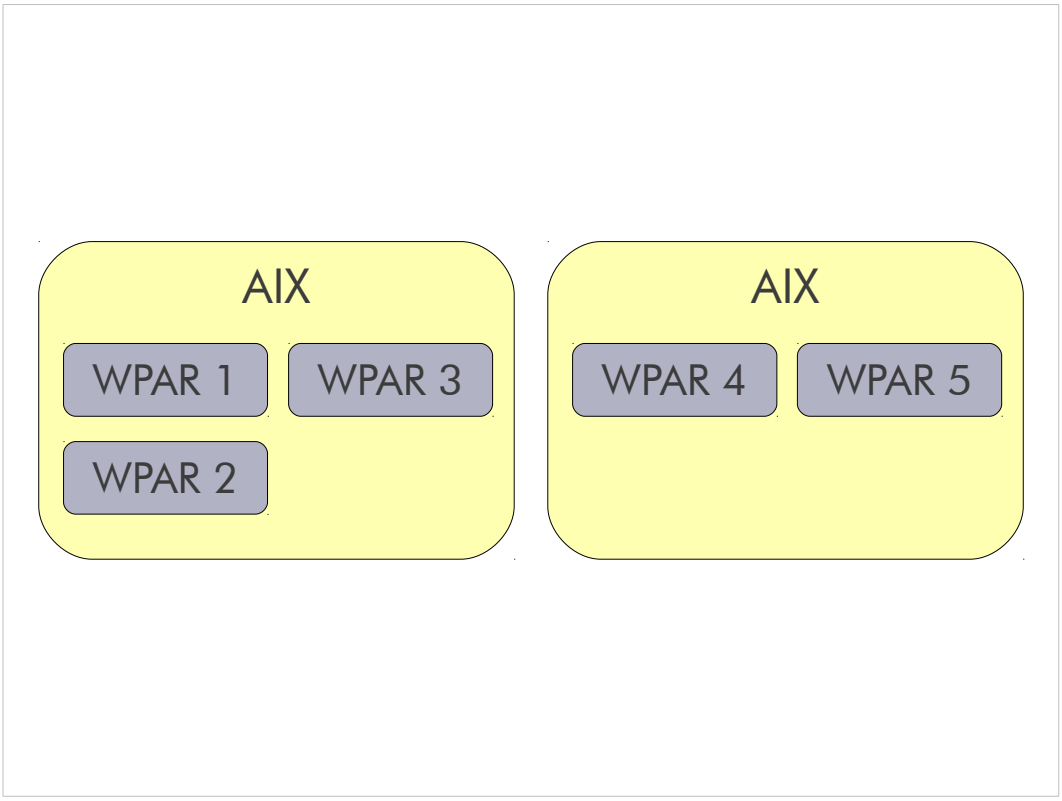
VIOS is a the Virtual I/O server for System p. The VIOS server owns all for resources and shares them with the other LPARS.

VIOS provides link aggregation for network HA



Live partition mobility.
If the AIX images live on the SAN, and are controlled by the VIOS

Uses NPIV



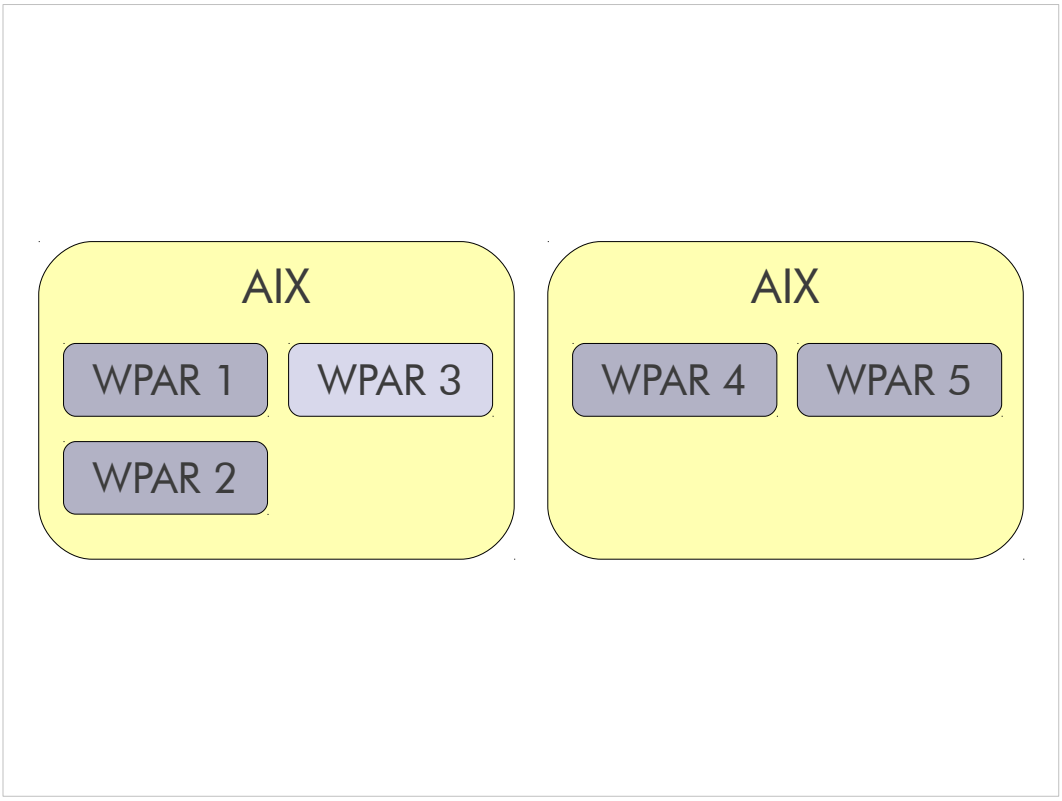
This chart is the first of the migration charts.

The only down side to WPARs is that the software needs to be certified to run in the WPAR and may require application changes.

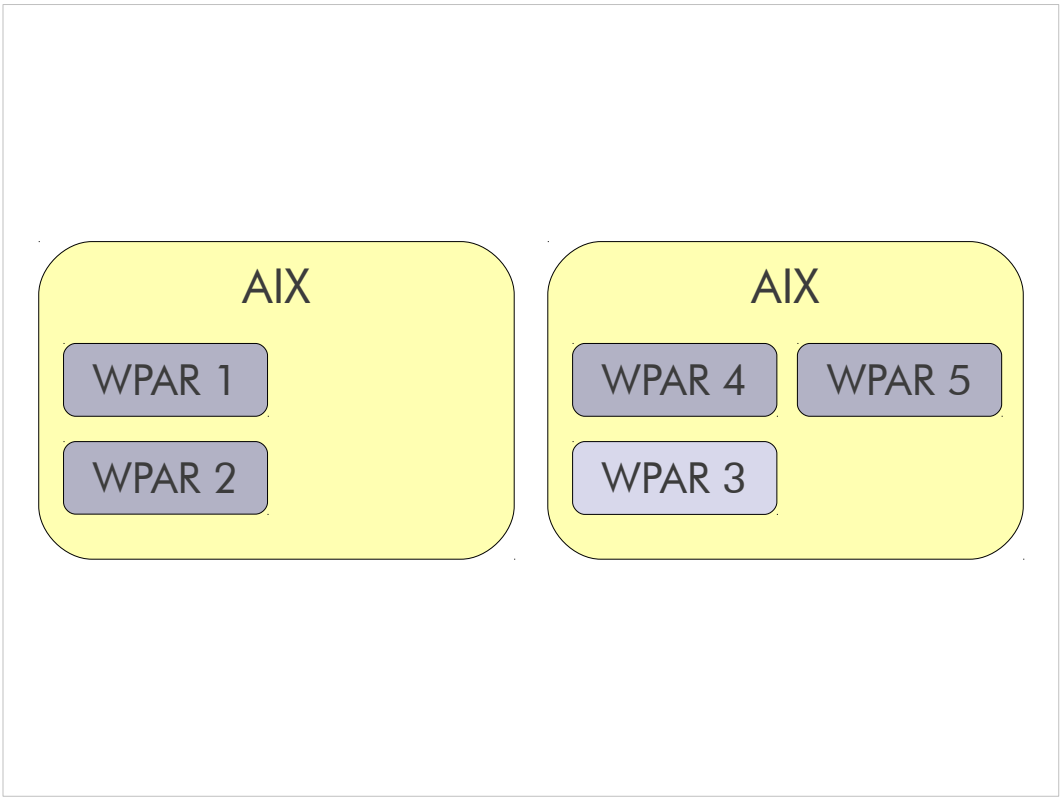
AIX 6.1

Here we have two AIX images. They may or may not be on the same CEC

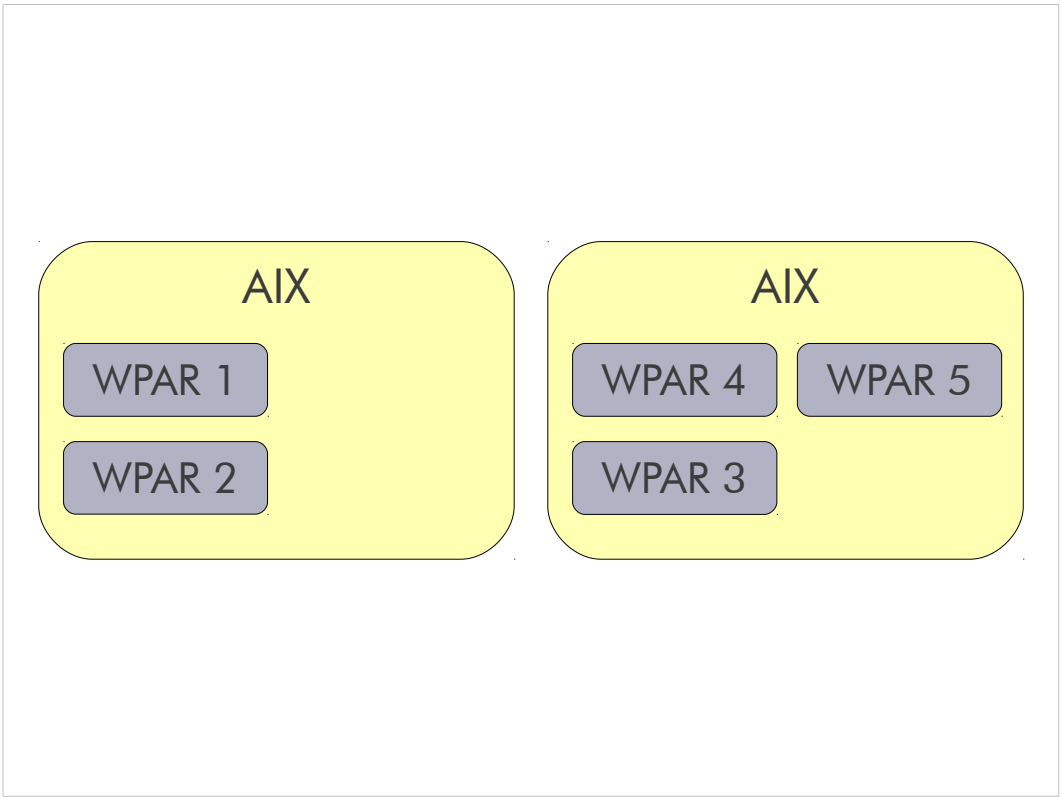
WAS 6.1 and 7 and parts of SAP are certifed



We select wpar 3 for migration.
A check point is created.
The Application is notified.
The application can the prepare to be notified.



Once the application is moved, a restart is triggered. The application can then do post migration routines.



The work can now continue as normal.

IBM developed the first VMM with the CP-67 but its performance was not good enough
A decision was made to create a new architecture with the goal of virtualization
The result was the VM/370 (Virtual Machine Facility 370)

First VMM was CP-67 for System/360
Its performance was less than desirable
IBM decides to tailor the architecture for running virtual machines
Result is VM/370, a VMM for
System/370 Extended Architecture

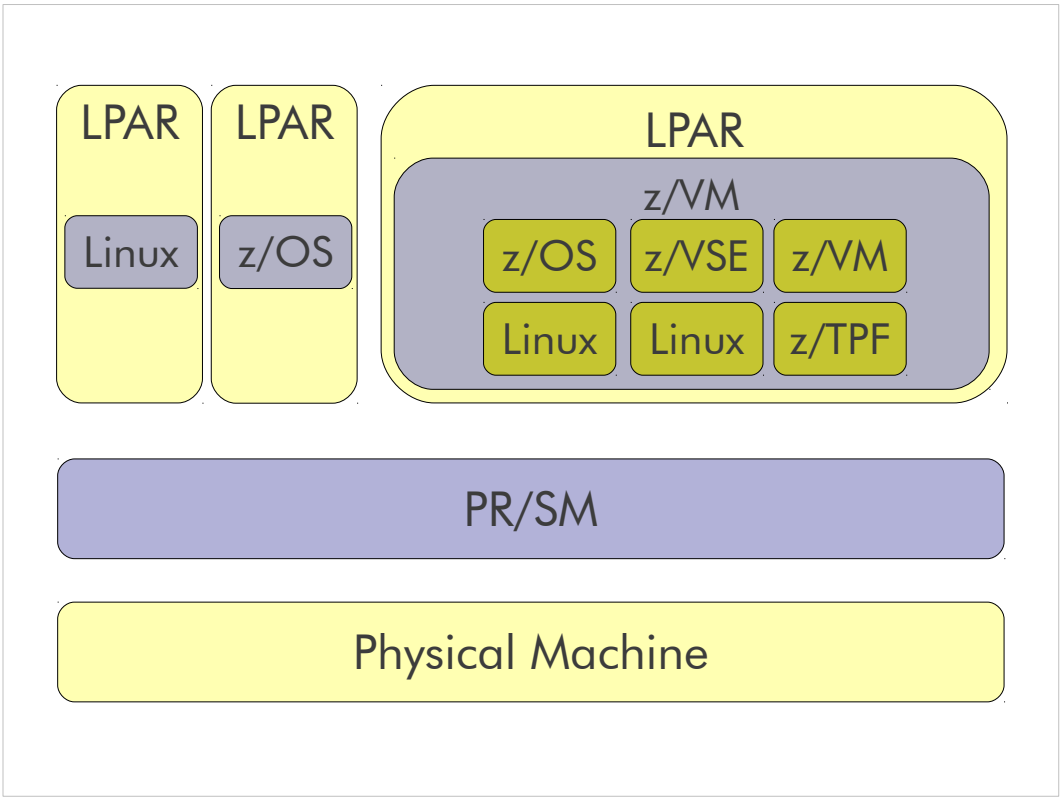
Virtualizing System/360

The VM/370 added several hardware assists to enable high performance VMM designs.

1. Virtual Machine Assist
13 instructions which replaced guest virtual machine instructions that would have been emulated in software otherwise. 35% performance increase.

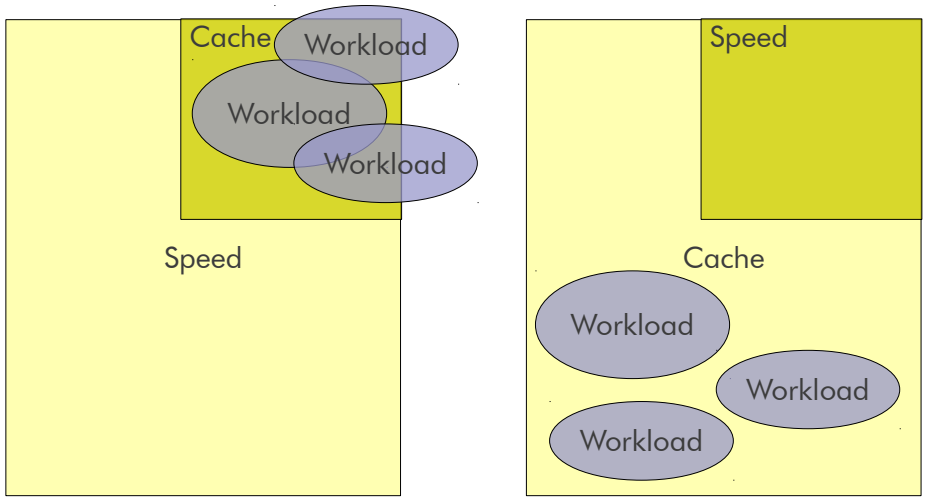
2. Extended Control Program Support
A set of 35 instructions which were targeted at specific applications. These instructions replaced some functions which were previously supplied by the vmm.

3. Shadow Table Bypass
Assists placed in hardware which allowed trusted guests to access the virtual memory system directly. A security risk but most machines were “well behaved” because they were designed by IBM.



Very high level architecture of both PR/SM and z/VM.
Ont thing that this chart highlights is the fact that System z has multiple levels of virtualization going on. PRSM and z/VM are both hardware supported. z/VM on z/VM is also good but the high level guest is software emulated.

Chip Design Affects Virtualization



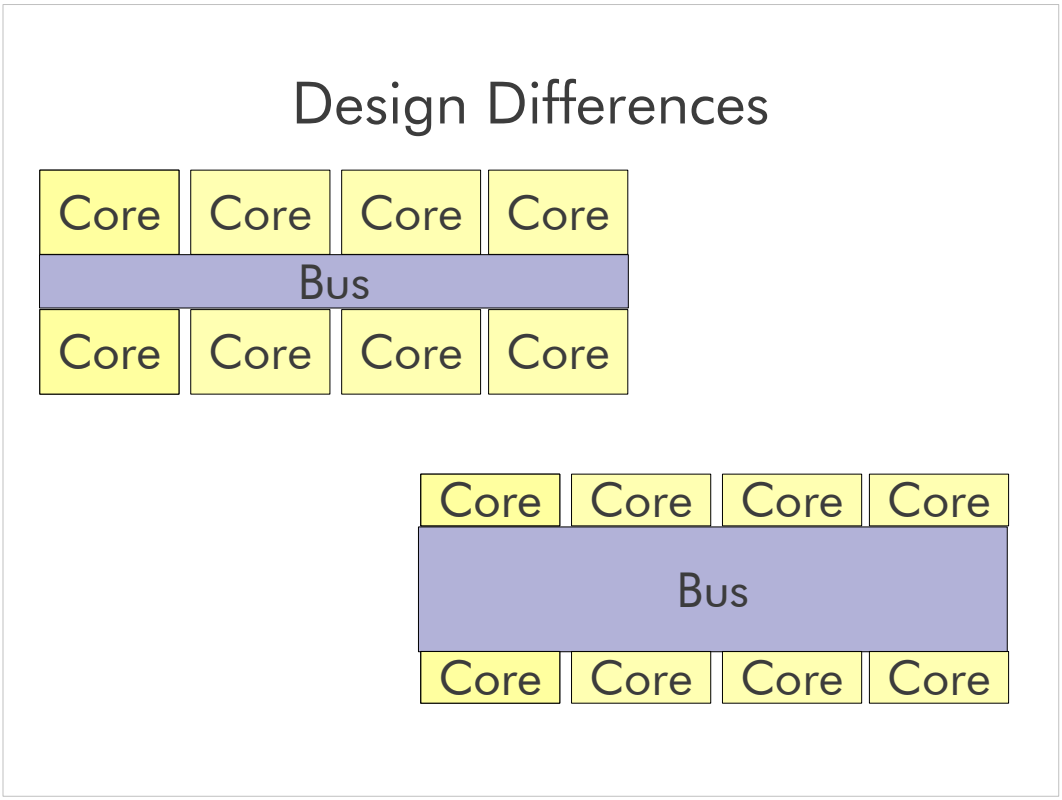
Mixed workloads stress cache usage, requiring more context switches

Working sets may be too large to fit in cache

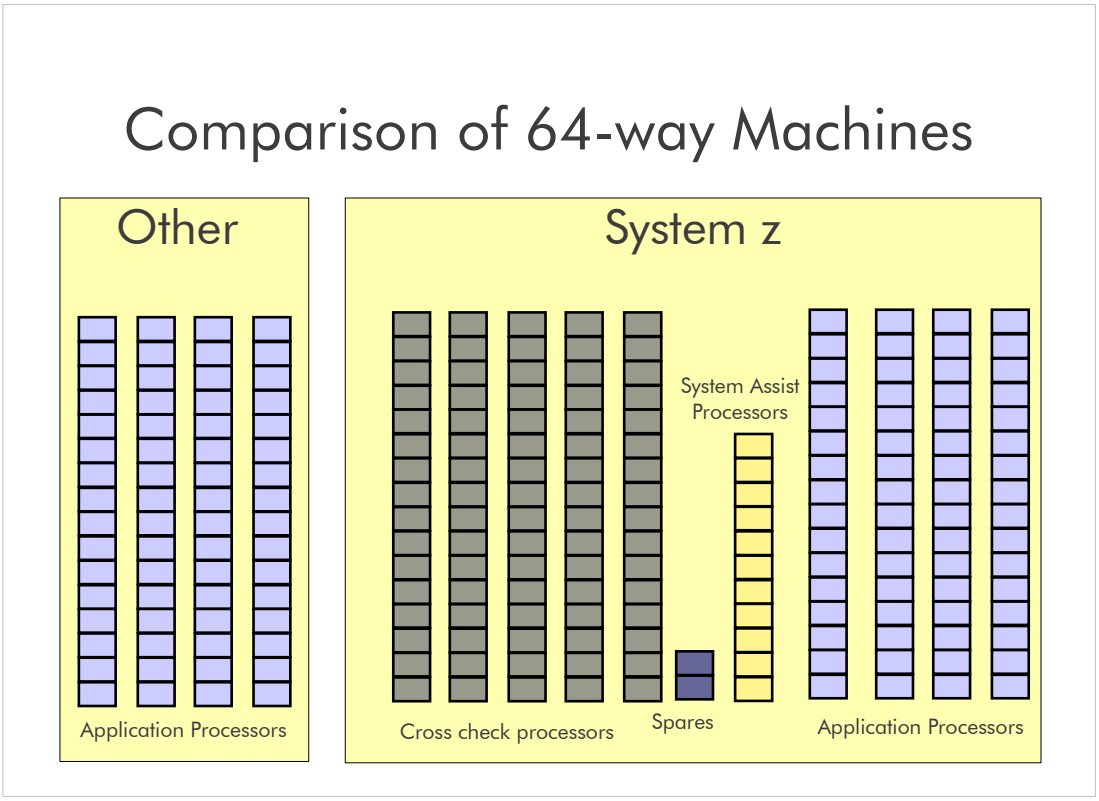
“Fast” processor speed is not fully realized due to cache misses

System z cache is able to contain more working sets

Processor speed is optimized by increased cache usage



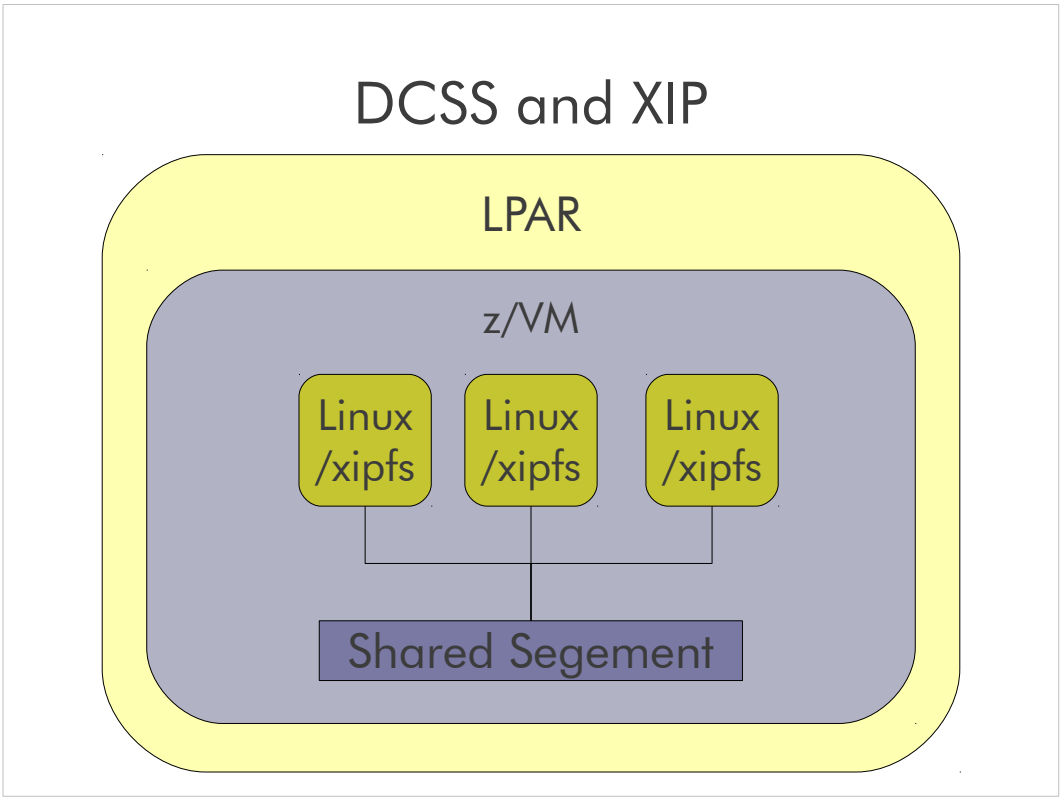
The top box is a distributed arch. The bottom is z.
The top has more speed, but small bus where as the z has smaller core but larger bus.
For workloads that are shipping lots of data, a larger bus prevents the core for data starving.
This is great for virtualization due to the fact that we are sending large parts of the machine between CPs



Generally speaking, there is way more processors in a system z machine. Each of the processors may have different abilites.

Another thing to state with this chart is you can over by processors. Say you get a full book and only use one. You have an upgrade path.

CPS can also be used for zIIP zAAP and IFLs.



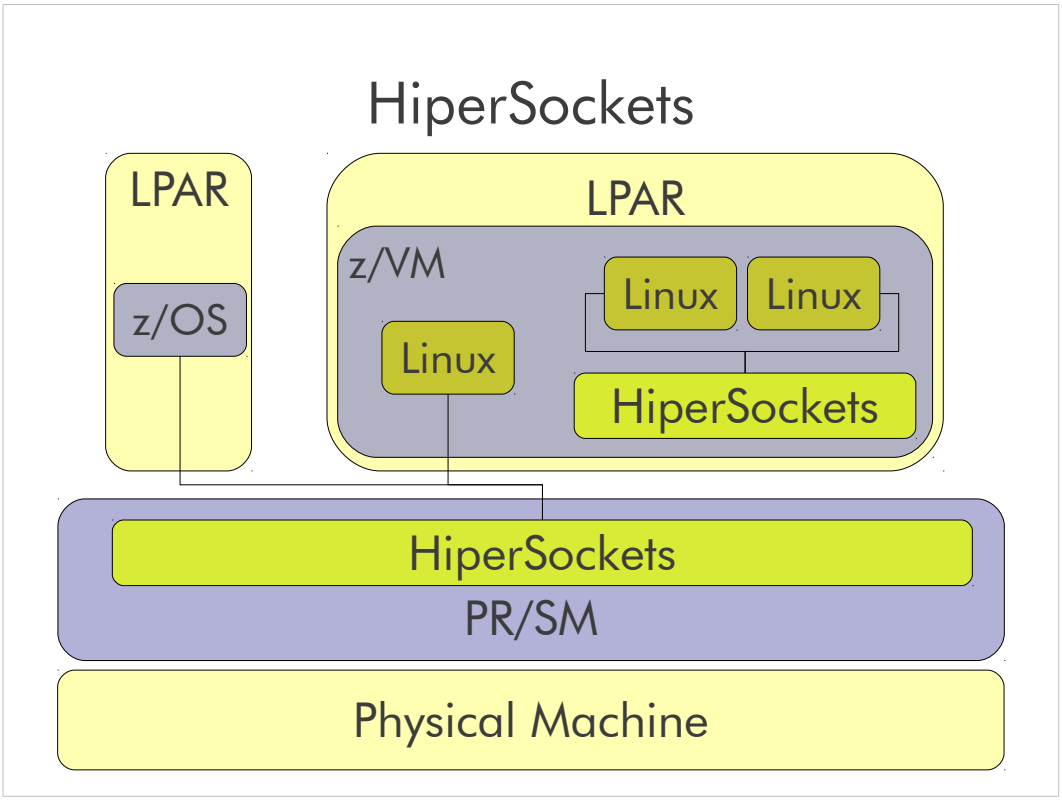
Discontiguous Shared Segment (DCSS)

Memory can be discontiguous to the virtual machine's address space, and a shared copy is loaded at the same address in all virtual machines that load the same DCSS. Can be a saved s above the virtual machine's defined storage size. In a virtual server farm with similar Linux instances there is often a considerable amount of data that is required by all instances.

eXecute In Place (XIP)

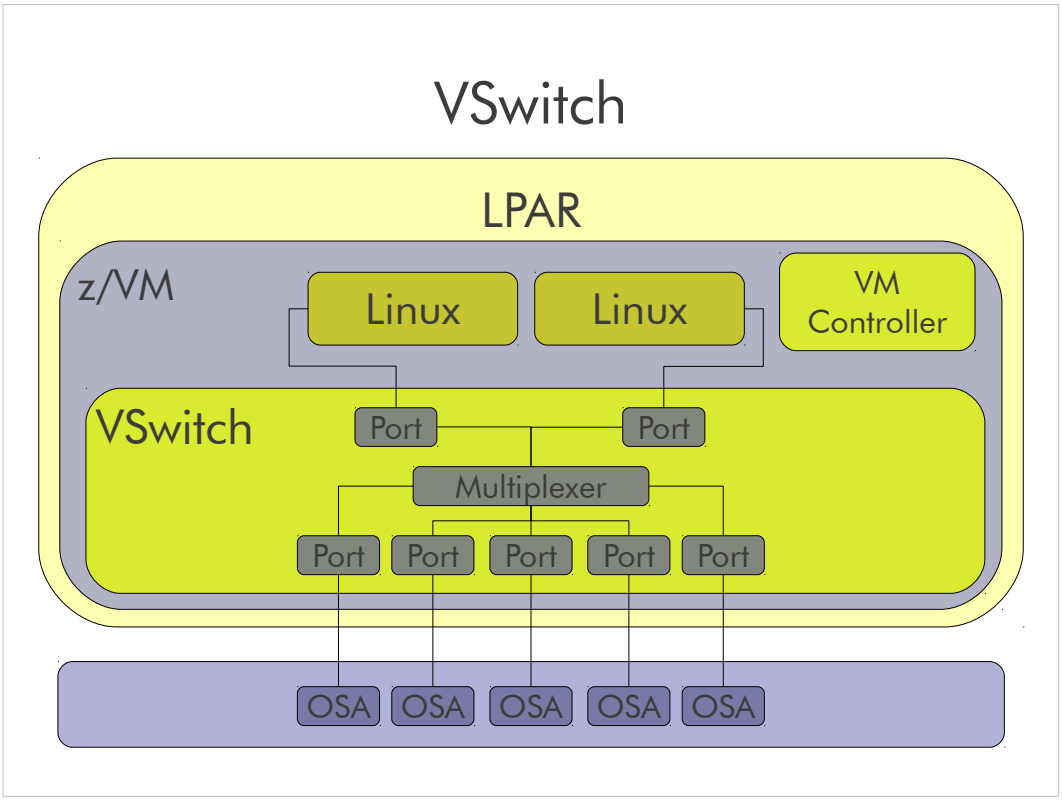
Filesystem the allows Linux to execute a file directly without need to load it into memory.

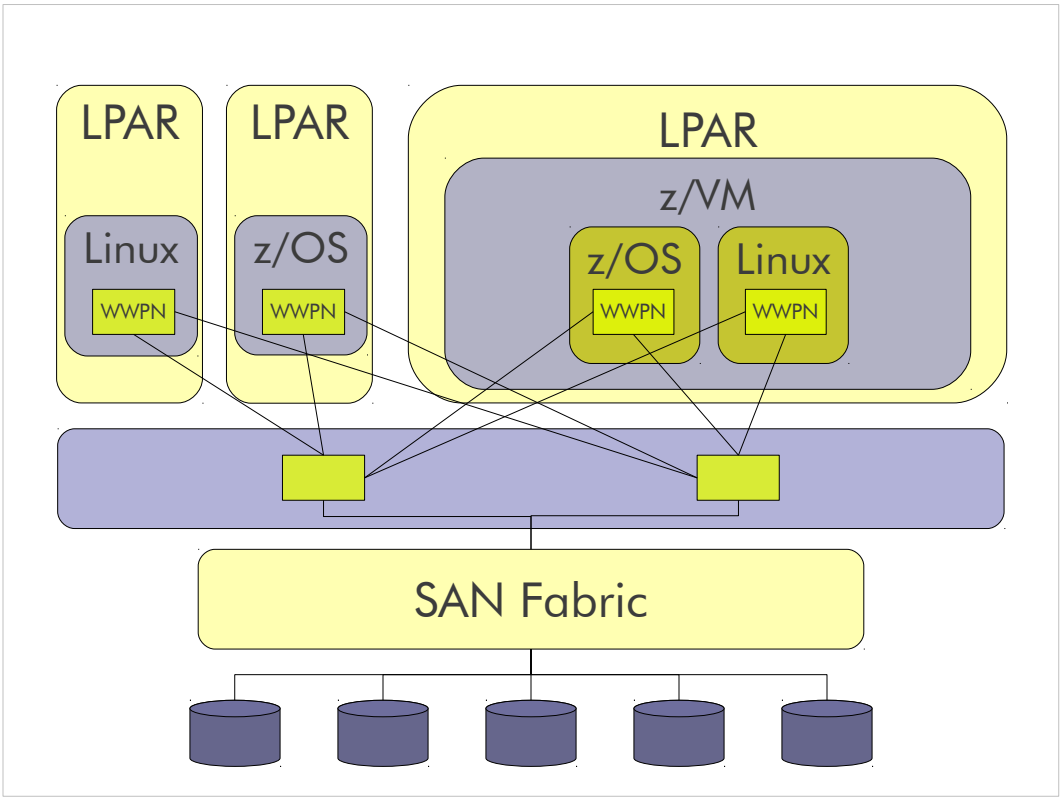
This is perfect for Linux Server Farms.



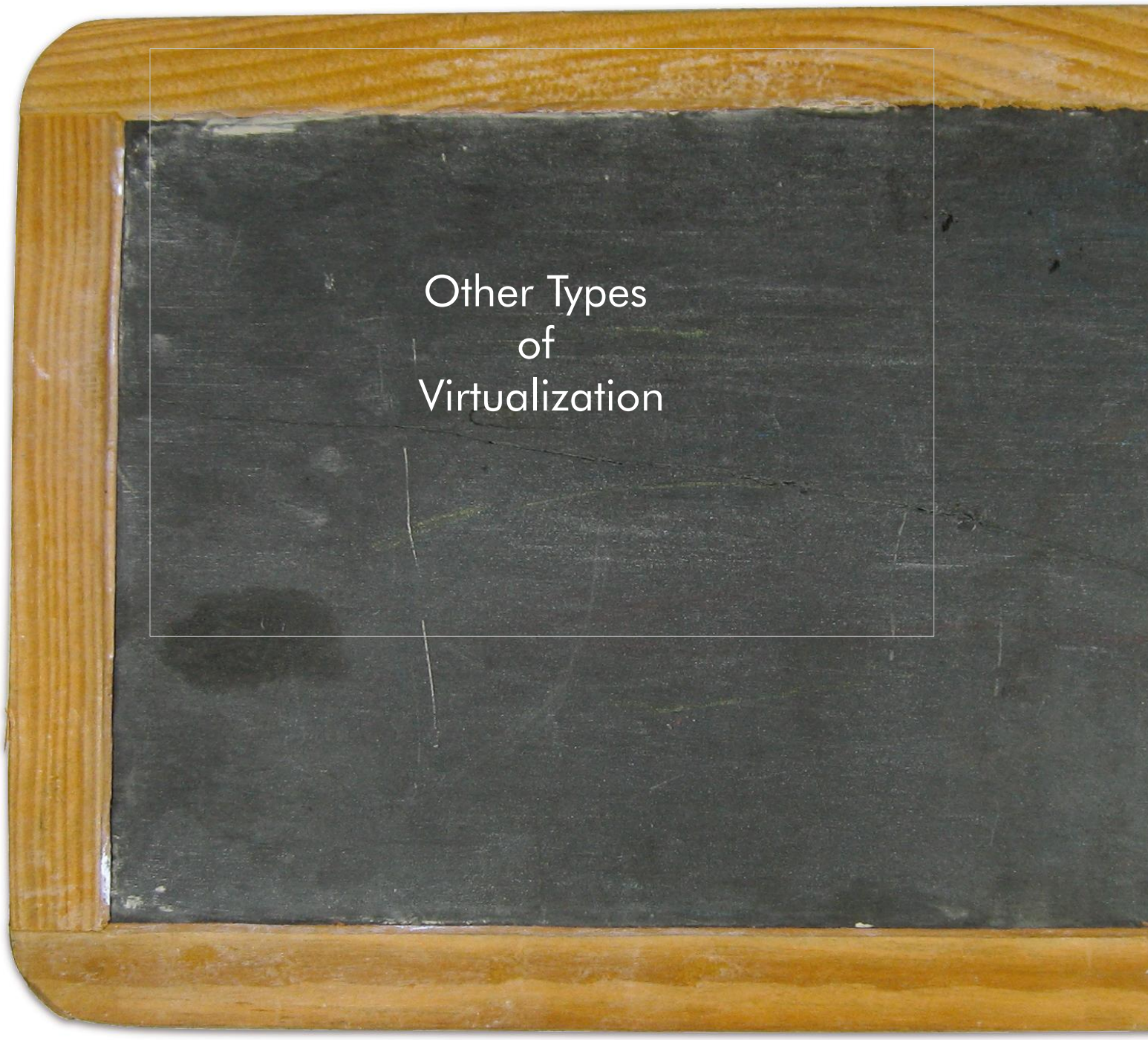
Internal only TCPIP stack.
We can mess with the frames so we can send giant datasets and not have the overhead of TCPIP

Down side is that is will not cross CECs

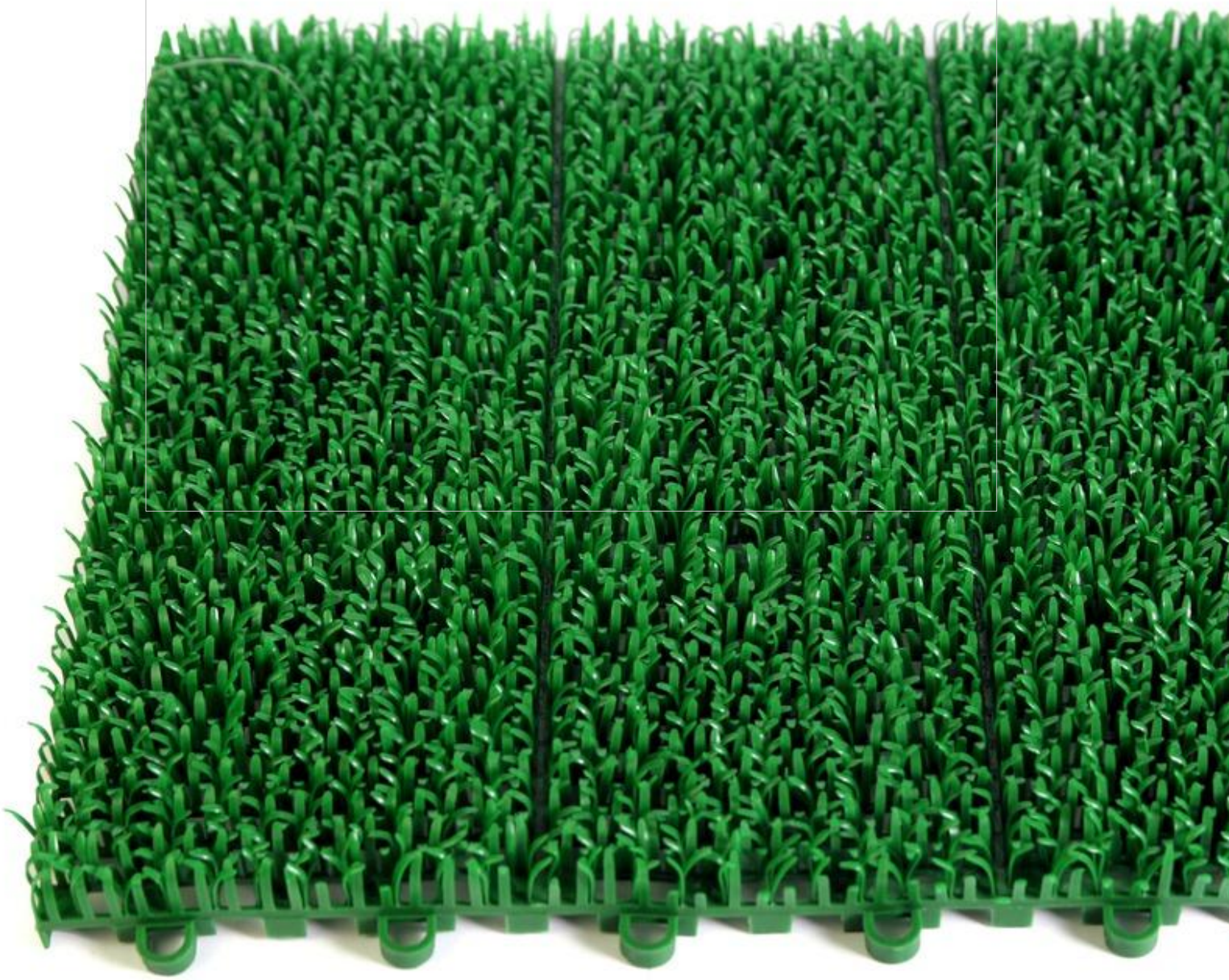

VSwitch
LPAR
z/VM
Linux
Linux
VM
Controller
VSwitch
Port
Port
Multiplexer
Port
Port
Port
Port
Port
OSA
OSA
OSA
OSA
OSA



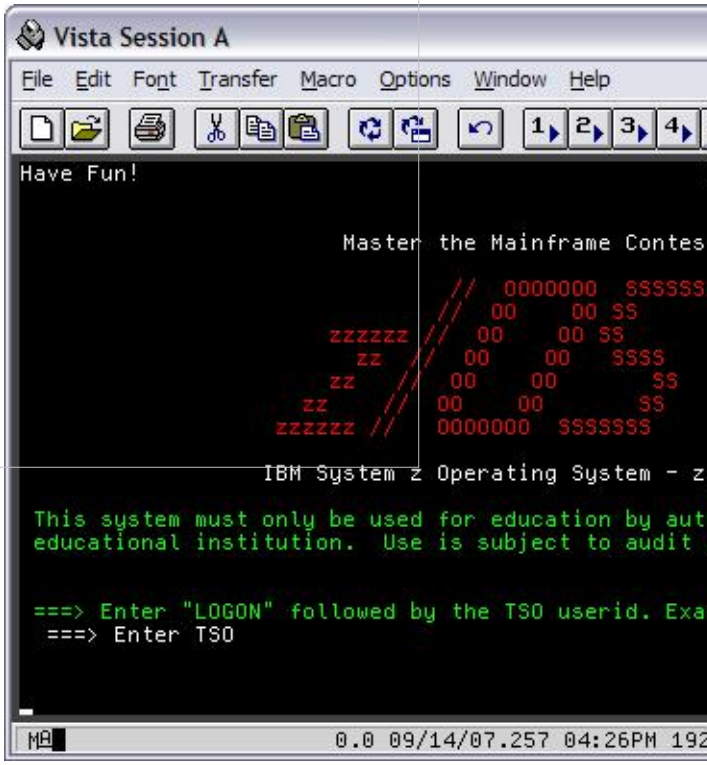
NPIV is a facility allowing multiple N_Port IDs to share a single physical N_Port. This allows multiple Fibre Channel initiators to occupy a single physical port, easing hardware requirements in Storage Area Network design, especially where virtual SANs are called for.


Other Types
of
Virtualization

Emulation

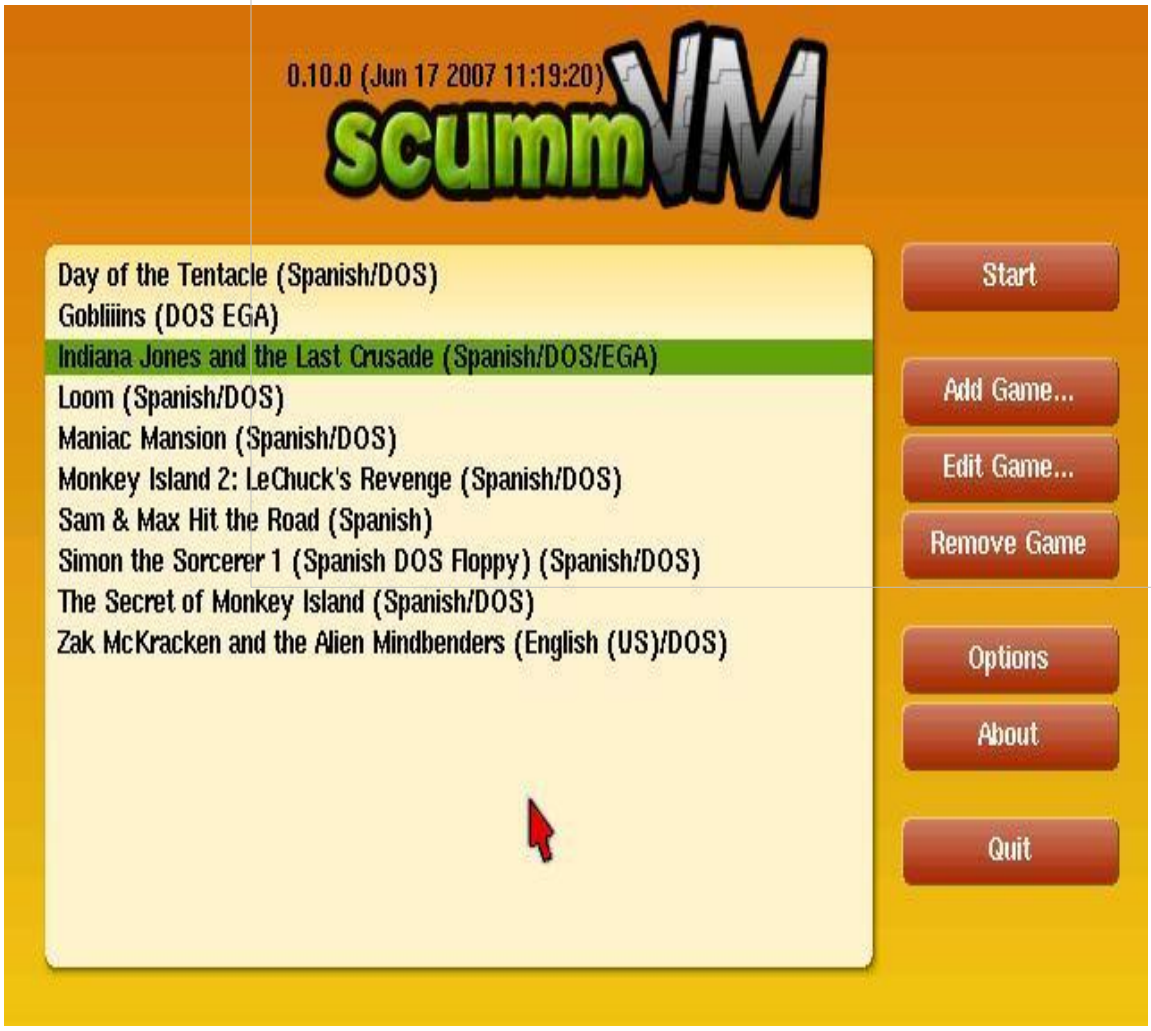


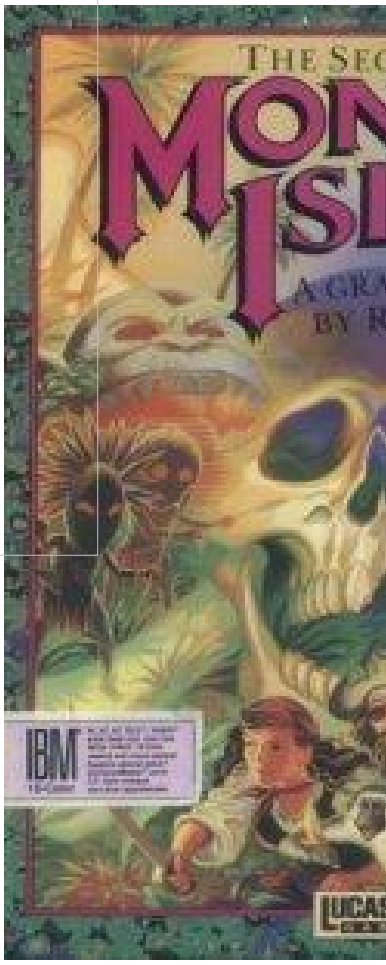
3270


Vista Session A
File Edit Font Transfer Macro Options Window Help
Have Fun!
Master the Mainframe Contes
IBM System z Operating System - z
This system must only be used for education by aut
educational institution. Use is subject to audit
===> Enter "LOGON" followed by the TSO userid. Exa
===> Enter TSO
MA
0.0 09/14/07.257 04:26PM 192

ScummVM

Script Creation Utility for Maniac Mansion





Wii

Wine

Chroot (Jail)

A chroot environment can be used to create and host a separate virtualized copy of the operating system. This can be useful for:

Testing and development
: A test environment can be set up in the chroot for software that would otherwise be too risky to deploy on a production system.

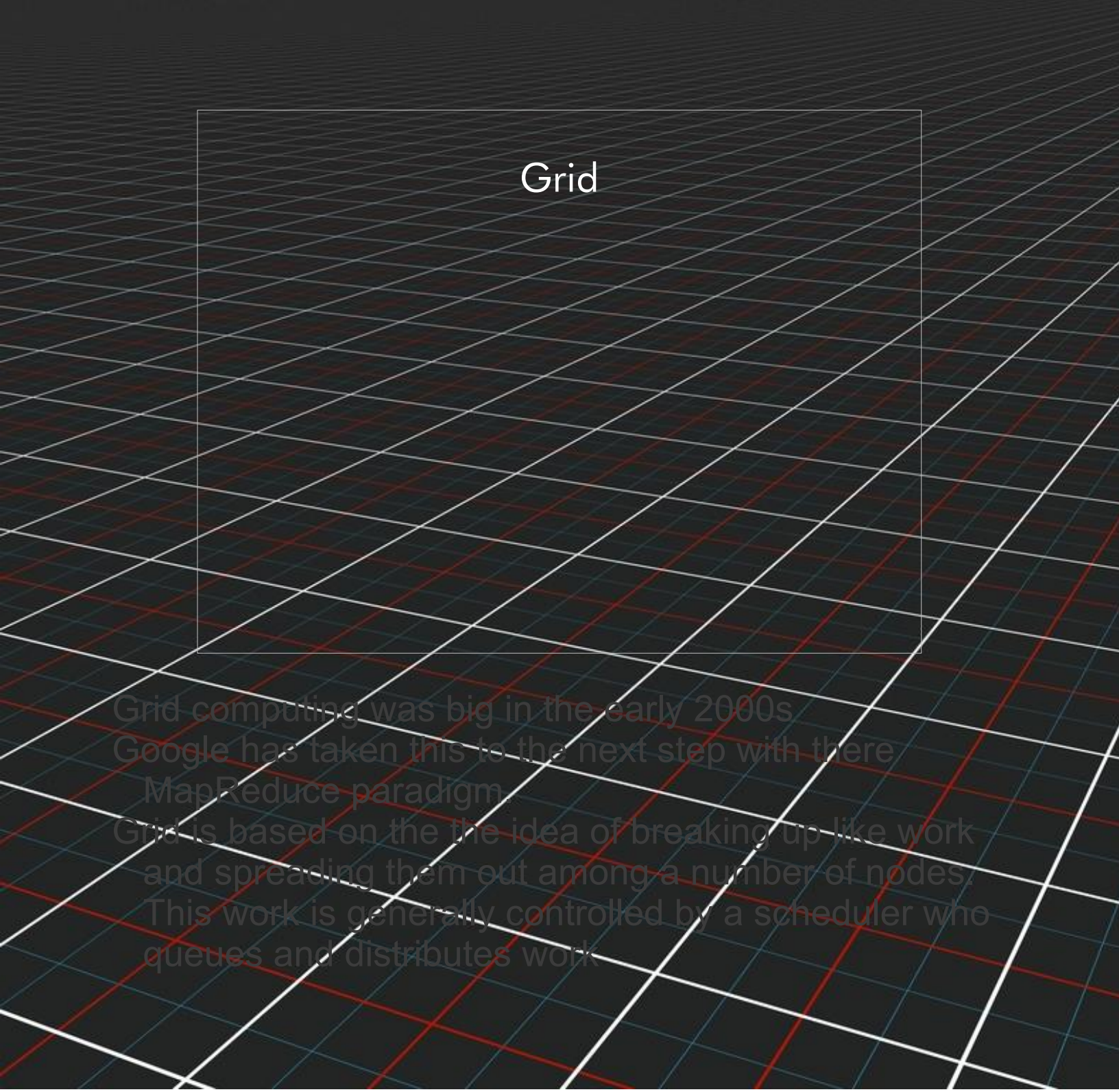
Dependency control
: Software can be developed, built and tested in a chroot populated only with its expected dependencies. This can prevent some kinds of linkage skew that can result from developers building projects with different sets of program libraries installed.

Compatibility
: Legacy software or software using a different ABI must sometimes be run in a chroot because their supporting libraries or data files may otherwise clash in name or linkage with those of the host system.

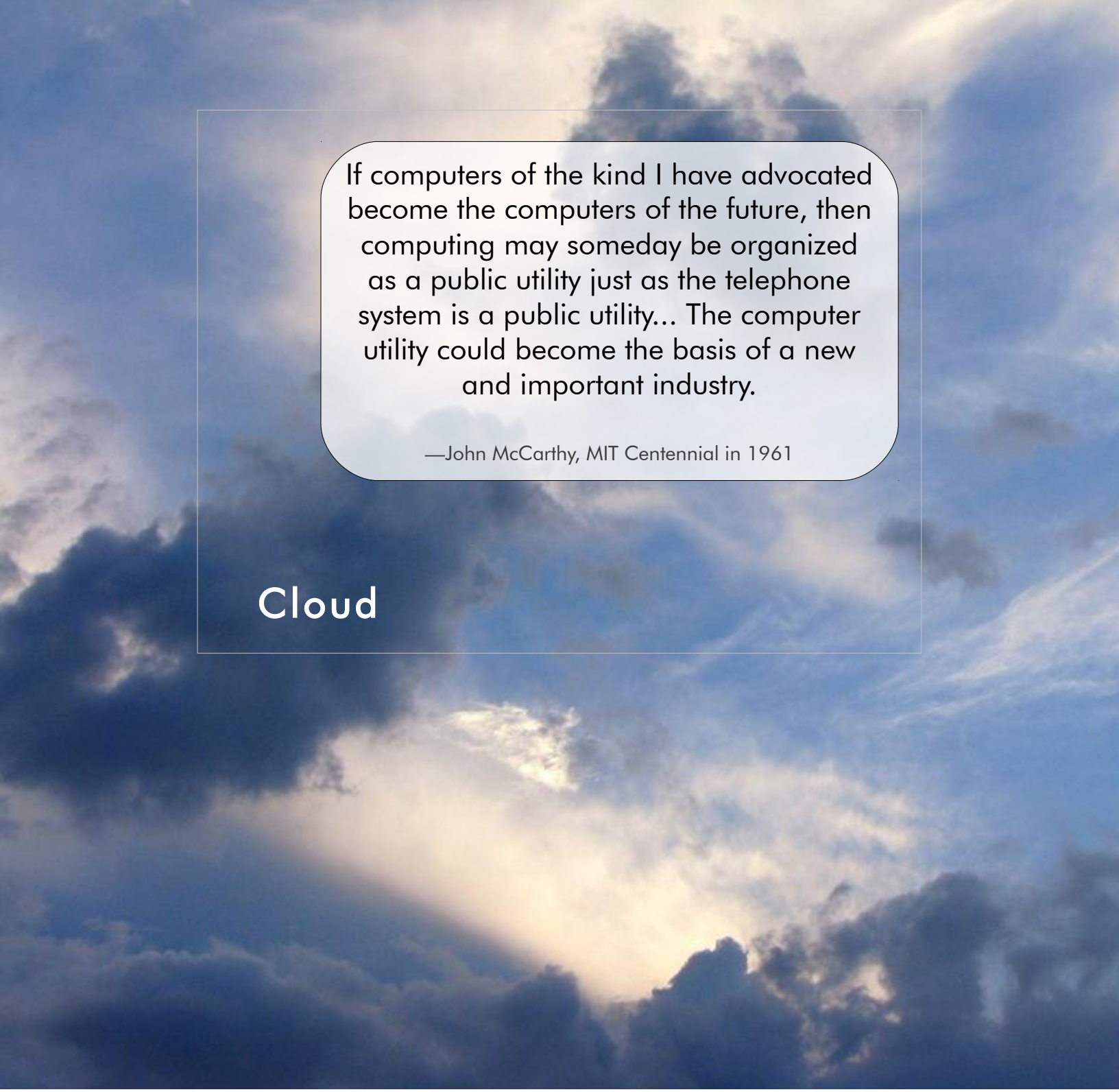
Recovery
: Should a system be rendered unbootable, a chroot can be used to move back into the damaged environment after bootstrapping from an alternate root file system (such as from installation media, or a Live CD).

Privilege separation
: Programs are allowed to carry open file descriptors (for files, pipelines and network connections) into the chroot, which can simplify jail design by making it unnecessary to leave working files inside the chroot directory. This also simplifies the common arrangement of running the potentially-vulnerable parts of a privileged program in a sandbox, in order to pre-emptively contain a security breach. An attacker with root privileges, however, may trivially defeat this separation because the chroot does not bar system calls, shield processes outside the chroot from tracing or disallow access to block devices.

Honeypotting
: A chroot can be populated so as to simulate a real system running network services. However, as chroot does not virtualize system calls, access to block devices or virtual file systems (such as /proc and /sys on Linux), it may still be possible for an attacker in the honeypot to detect the presence of the honeypot and the system outside the chroot.



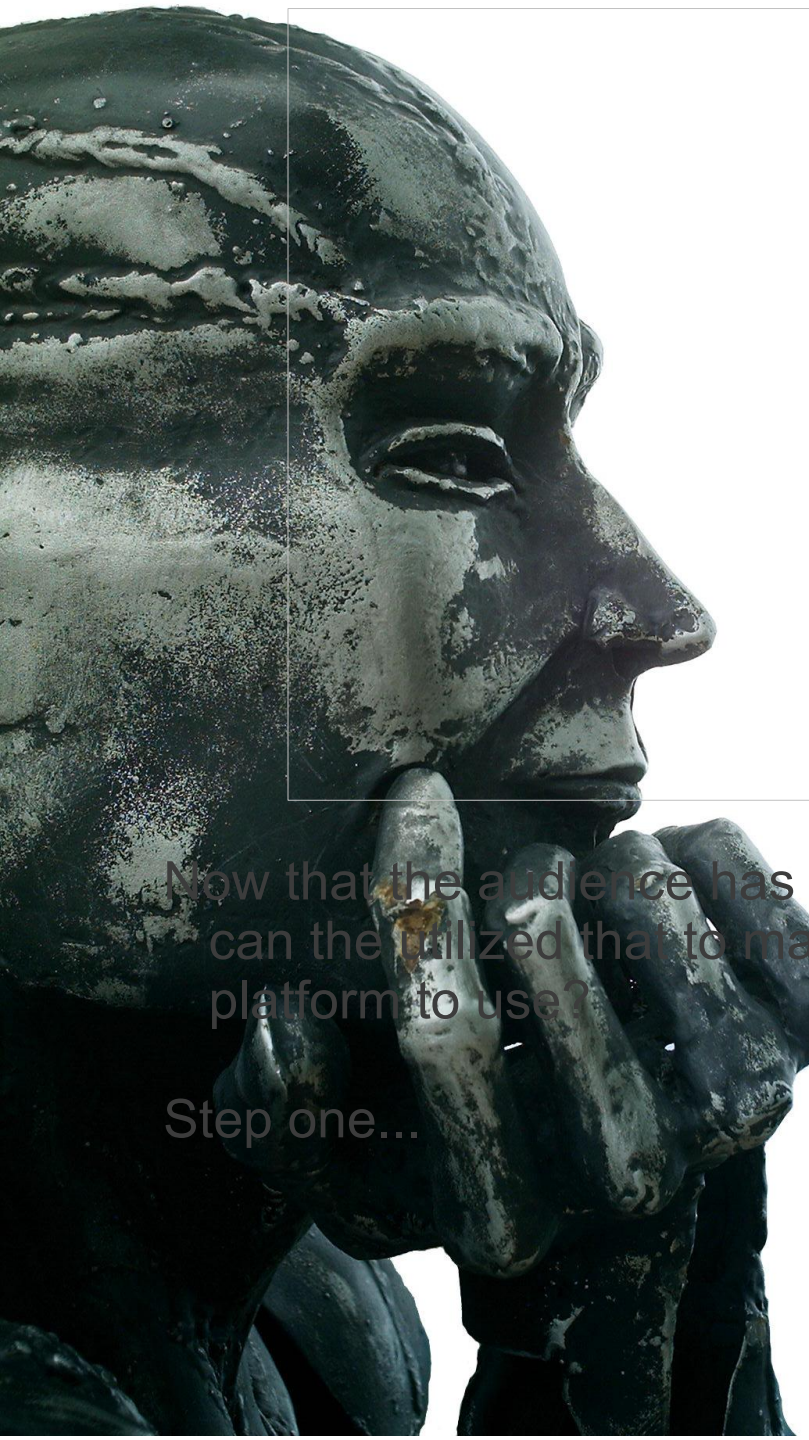
Grid computing was big in the early 2000s
Google has taken this to the next step with there MapReduce paradigm.
Grid is based on the the idea of breaking up like work and spreading them out among a number of nodes. This work is generally controlled by a scheduler who queues and distributes work



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



Now that the audience has a bit of knowledge, how can the utilized that to make the choice of which platform to use?

Step one...

What are the pain points

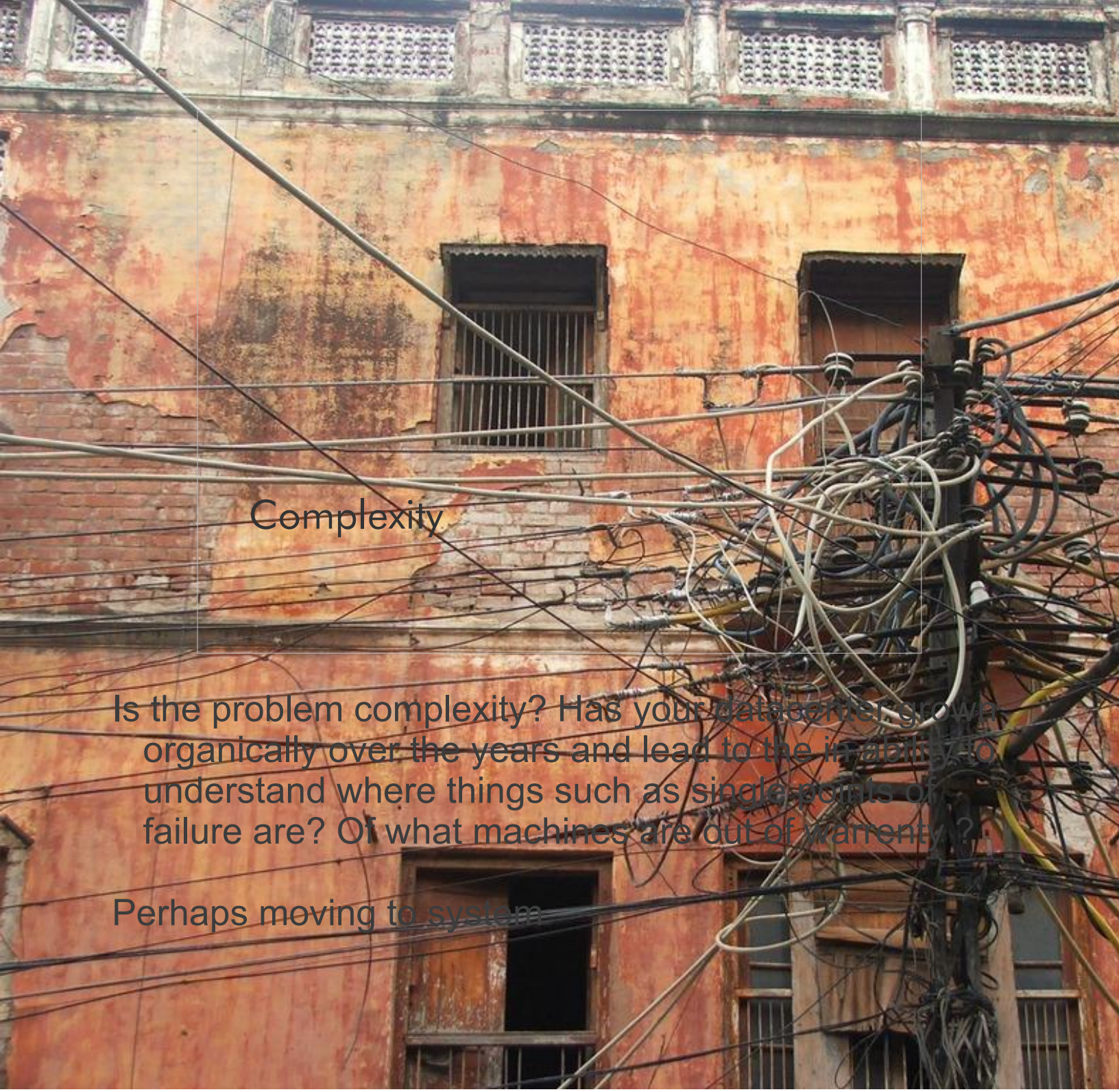
- You are virtualizing for a reason. What is that reason? Once that is understood, one can narrow down the choices.
- In addition to the pain points, what is the business context you have to operate in? What are the functional and non functional requirements?
- This is the process that the Design Center uses to understand a customers environment and help them address those pain points.

Cost

Is the problem cost? Of course it is.
But where? Software? Hardware?

Knowing where you are spending that shrinking IT budget is key into understanding which virutalization technology suites your environment.

Having a total cost of ownerships will help change mindsets that state cheaper hardware is the cheapest solution.

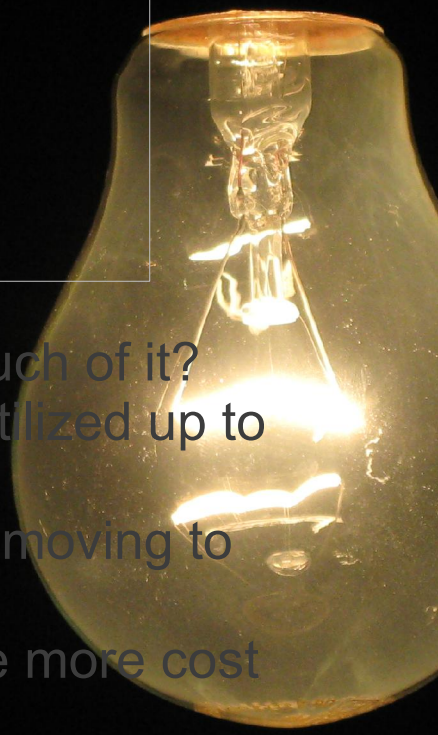


Complexity

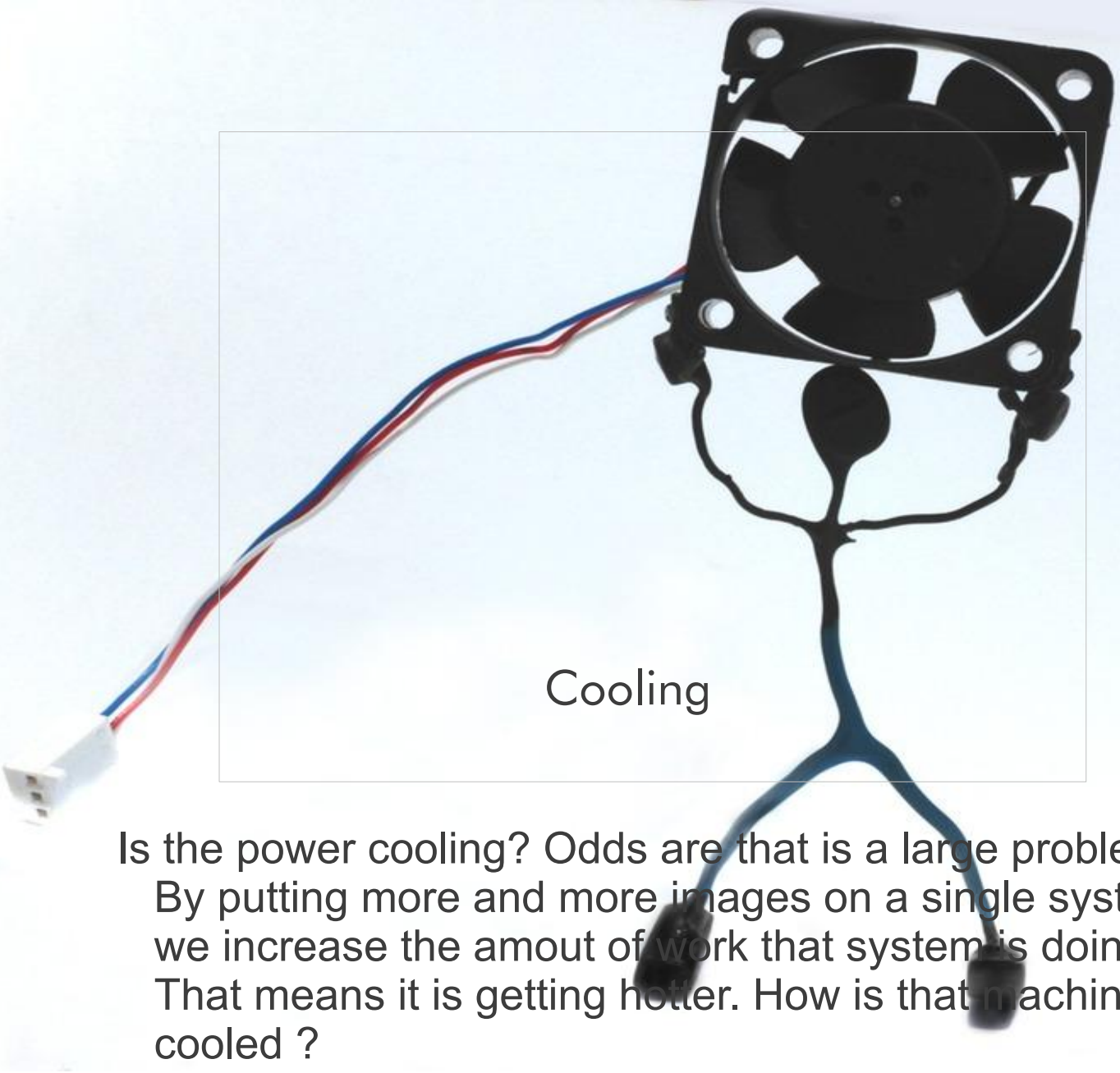
Is the problem complexity? Has your datacenter grown organically over the years and lead to the inability to understand where things such as single points of failure are? Of what machines are out of warrenty?

Perhaps moving to system

Power

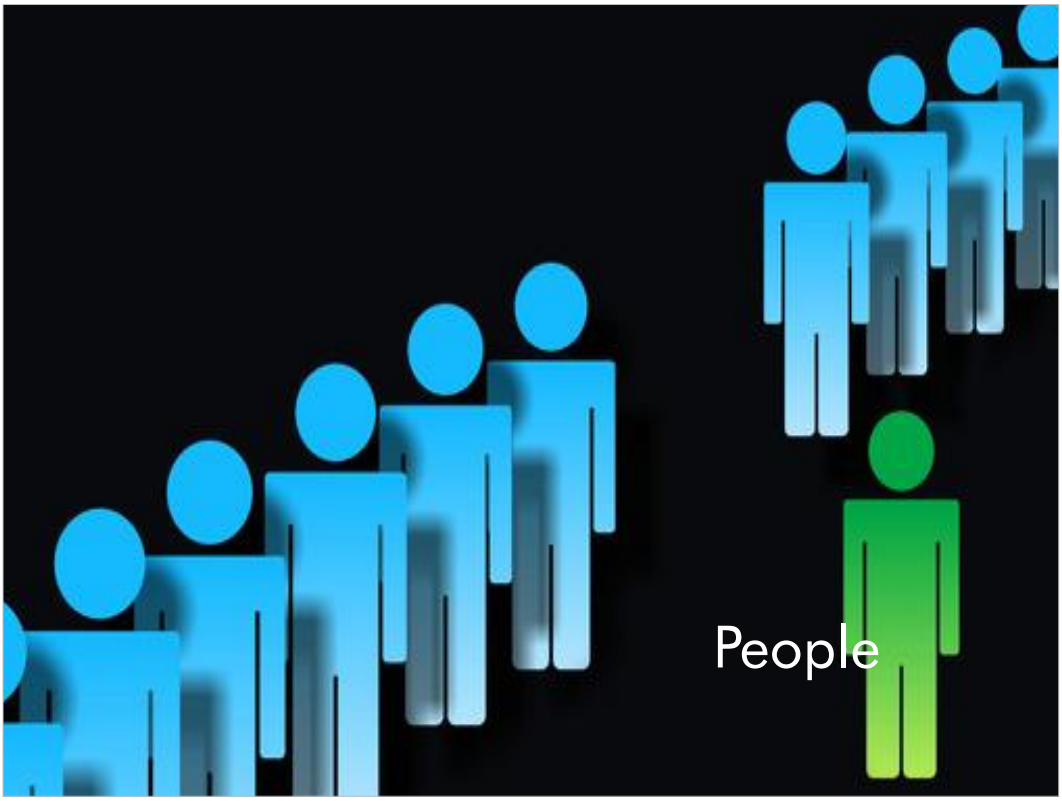


- Is the problem power? Are you using too much of it? Maybe moving to a platform that can be utilized up to 100% instead of 10% is better.
- If the problem is cost of power, then maybe moving to a virtualized platform that can shift power consumption based on workload would be more cost cutting.
- Some clients have power availability issues. What do you do when you can't get another 480 line dropped into your datacenter? Do more with less and virtualize.



Cooling

Is the power cooling? Odds are that is a large problem. By putting more and more images on a single system we increase the amout of work that system is doing. That means it is getting hotter. How is that machine cooled ?



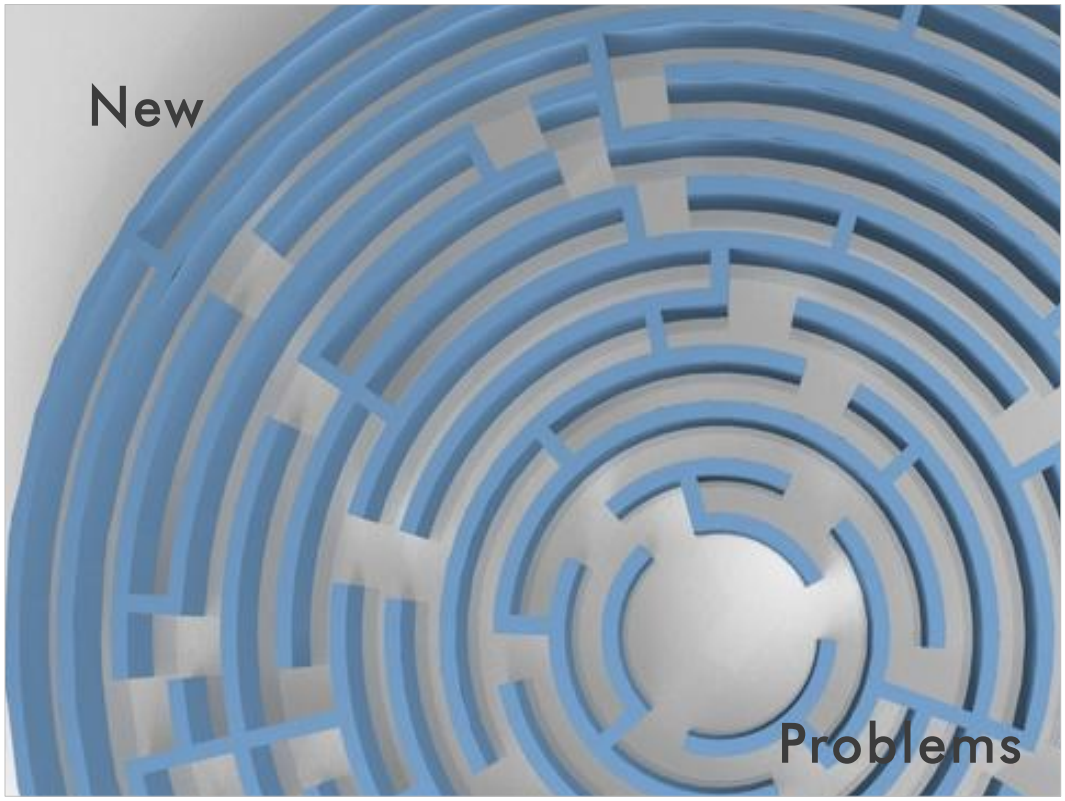
Is the problem that you just don't have the people to keep up with the systems administration that is required to maintain a non virtual environment ? Maybe VMware's suite of system management will help. Or maybe use libvirt to manage more than on hypervisor.


DEAD
END

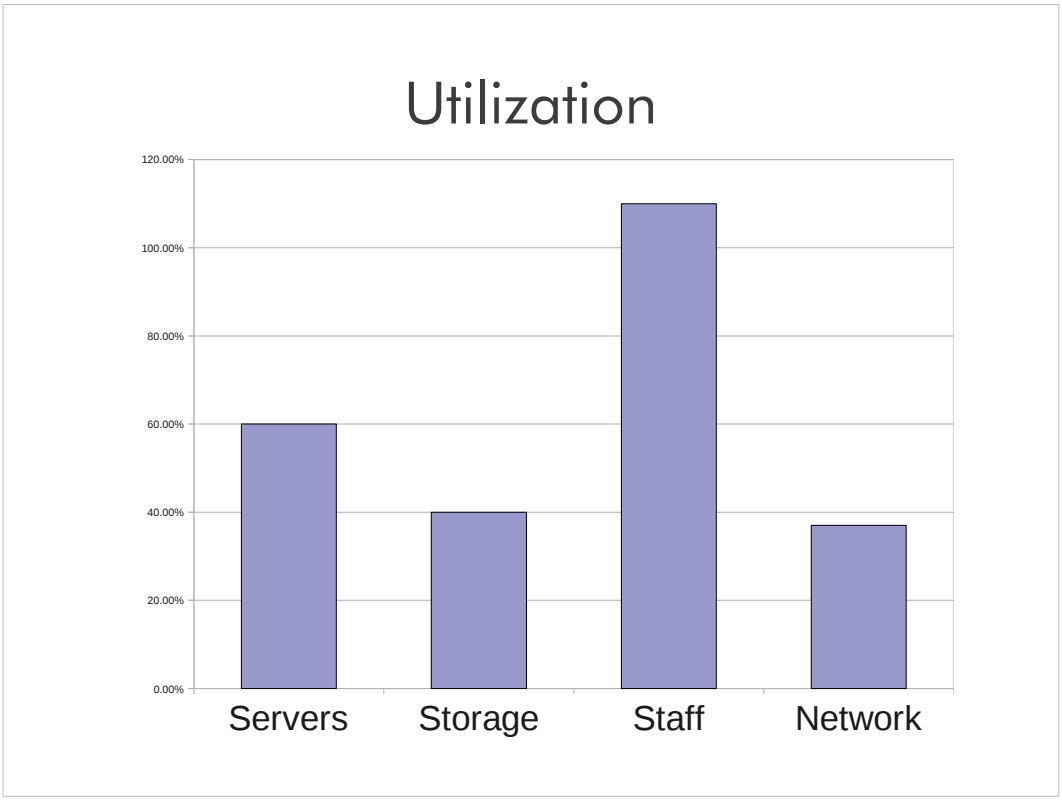
What kinds of workload do you have? OLTP with lots of small packets? Large packets? Lots of computation?

Does it need to connect to a remote datastore ?

Knowing what the workload looks like can really help chose the best place for it to live. For example a CP heavy WebSphere application might be better on a zLinux environment with a HyperSockets connection to a z/OS DB2. But if the packets are really really little, HyperSockets are not providing much.

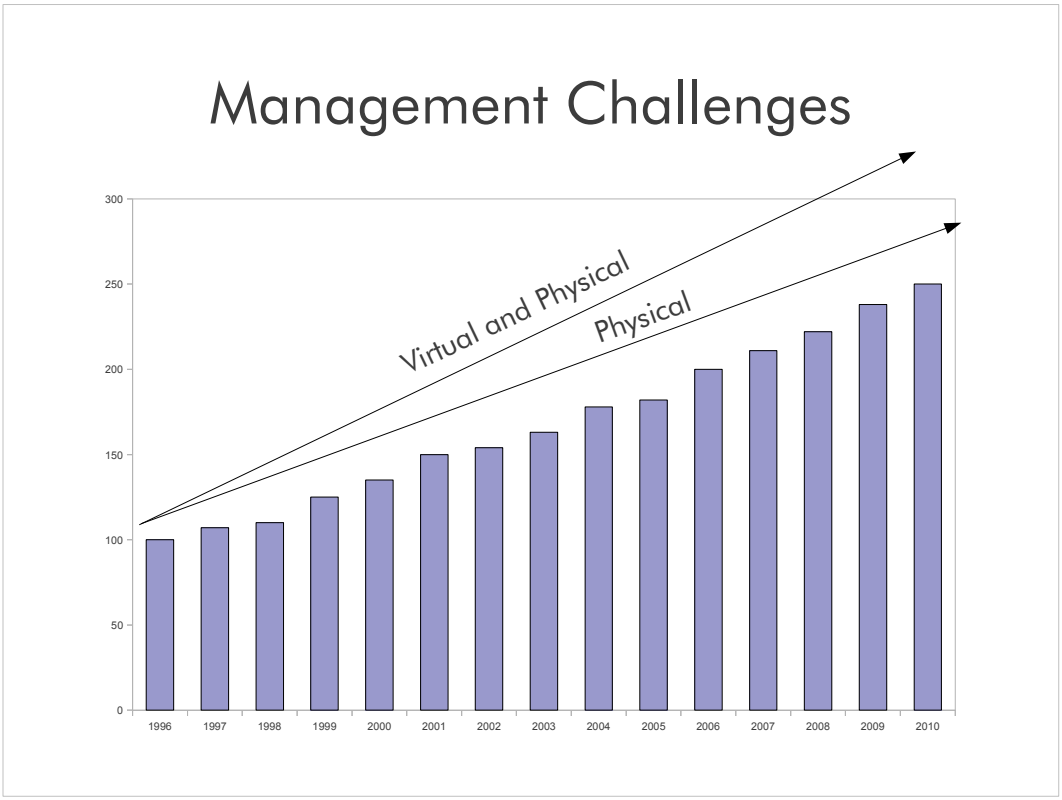


They say that great power comes great responisiblity. This is also true for the adoption of new technologies. The next few slides deal with the new problems one may encounter when adopting virtualization.



Declining costs of storage make it easy to grow but the result is overutilized staff and underutilized IT resources.

This glues the people cost towards in the previous slide and the management costs that are in the next slide. Think about how much it costs to maintain your storage environment



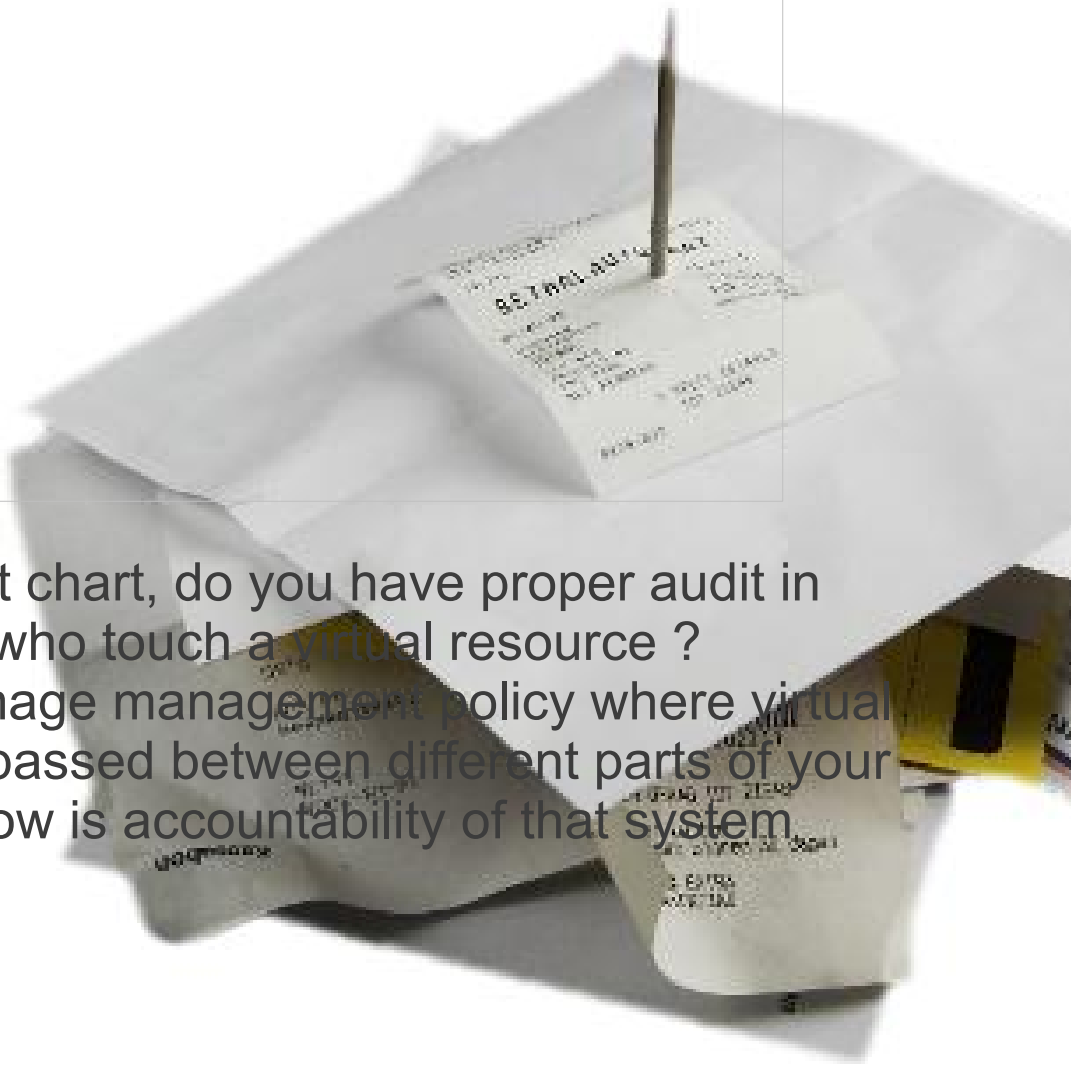
The amount of physical sever purchases continues to grow nearly linearly The amount of servers though is increasing at an almost exponential rate.

- Agent proliferation deals with the fact that if an agent, such as IBM's director agent, needs to exist at the OS level, there are going to be a lot more on the physical hardware.
- To get around this one can use DCSS on System z or WPARs on System p

We discussed live migration in a number. How them does one know where a virtual machine or workload application partition at a given time? Do you have compliance issues that may be affected by not knowing where the applications and or data are? How actuate are your audit logs?

Another issue is with knowing when hardware faults occur. If you are using really advanced fail over mechanisms, how do you get notified of failures.

Accountability



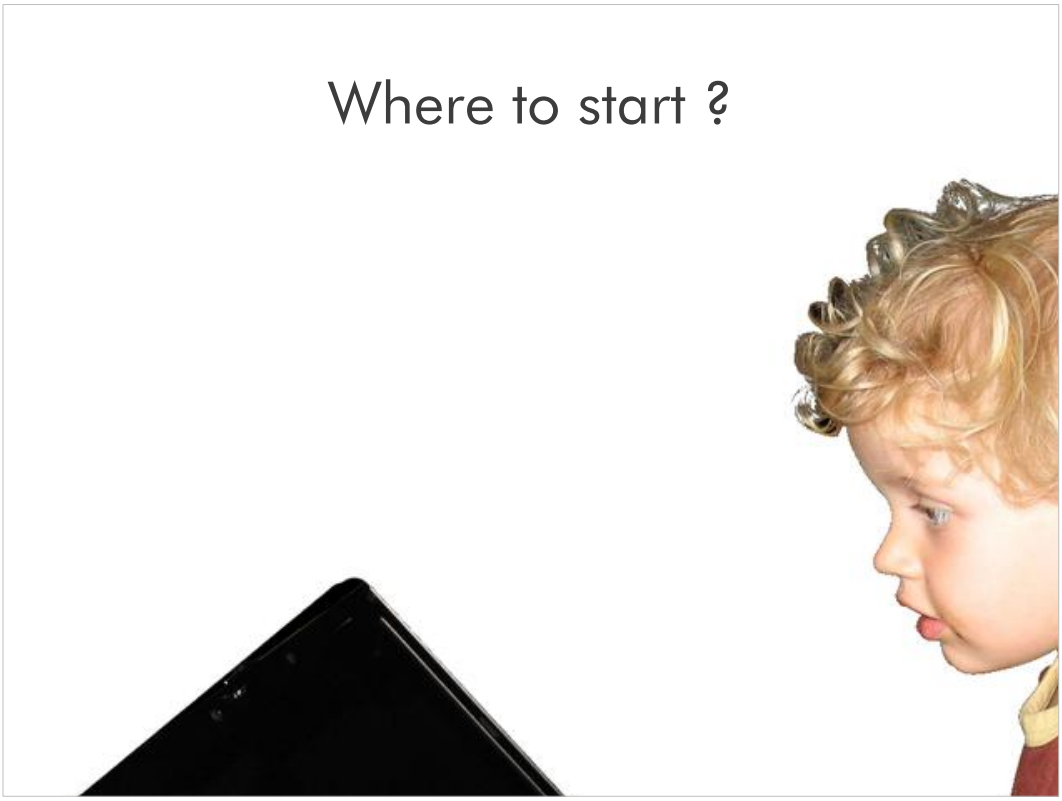
Similar to the last chart, do you have proper audit in place to know who touch a virtual resource ?

If you have an image management policy where virtual machines are passed between different parts of your environment how is accountability of that system recorded?

- Different virtulization technologies have been designed from the ground up with security in mind. Others have not.
- As one moves from dedicated to virtualized and higher, having an important security road map to match is very important.
- Security regulations might also impact your choice.

Developers and testers will use everything they can get there hands on. A corporate strategy on image management will help reduce this cost.

For instance, I am testing a product that is being hosts on Vmware. Some of the other people take daily snapshots of the virtual server. We now have 500 gigs or so of snapshots with no process to remove or clean them up. VMWare makes it too easy to do this.

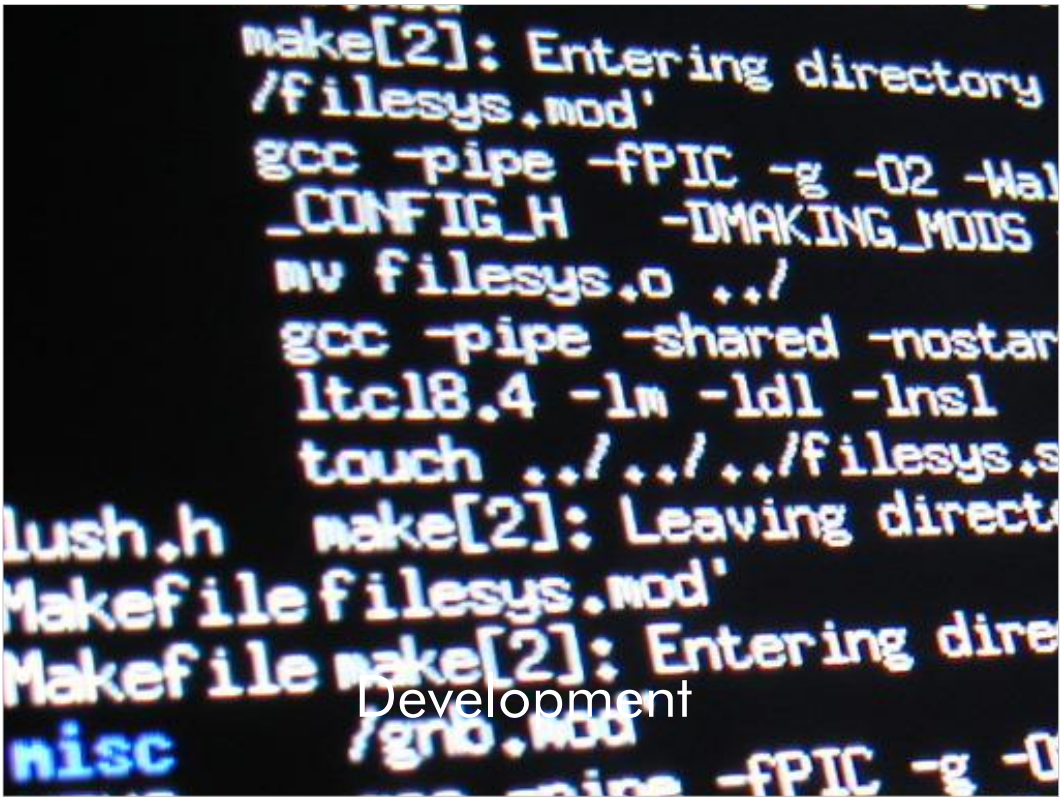


Once a technology is chosen, where does one start?. Management is going to be looking for a very good return on investment.

The first thing I tell customers to look at is on the next slide.

Seems straight forward. Go for the lowest impact/highest return.

Each company is different with there cost structure for providing a service.



Developers tend to need a lot of resources for a little time. Whether it is storage, machines, or network, virtualizing these resources can greatly reduce the cost of developing a service.

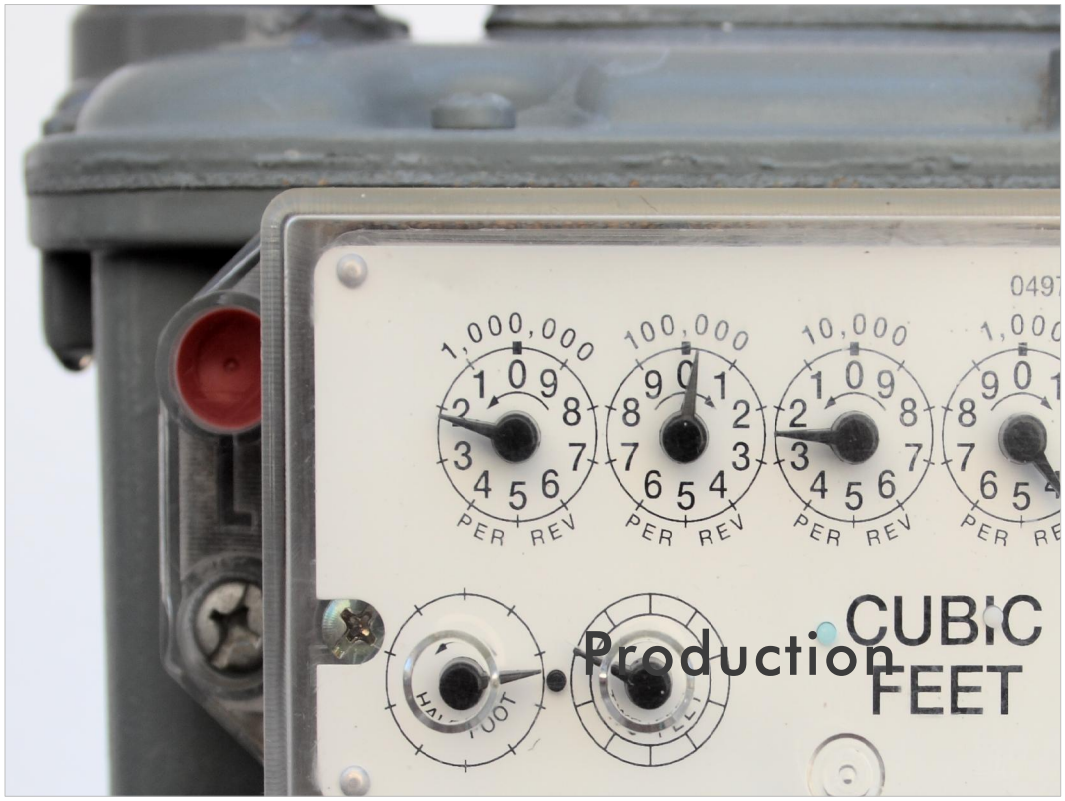
This is generally the area that has the least impact if the chosen virtulaization platform does not meet your needs.

After production comes test. Function test, System test, and integration test can all benefit by using virtualized resources.

For example, I was in system test for many years. Since we were almost the last line of defense before the product went to the customer, we where always stressed for time. By using VMware, and being able to save images off to the network, we could system test around the clock with our co tests in Shanghai

Quality Assurance

- Quality assurance is an ever increasing part of service delivery. As we move to cloud and service oriented architectures, it is imperative to ensure that our products continue to meant the service levels.
- Virtulizing here can be difficult. We do not want to impact our production like quality, but this could be the real world test of a virtualized environment in production.



And finally one can move our virtualized environment into production. One can drive up utilization of equipment, reduce single points of failure and provide a more dynamic service offering to ones' clients.

The End