FUNDAMENTALS OF VIRTUALIZATION

Lesson 3

Virtualization

Goals

- □ Introduction of what virtualization is.
- □ Understanding of where virtualization can be applied.
- □ Knowledge of virtualization could help for BG.

Contents

- What virtualization is.
- Where virtualization could be used.
- Why virtualization is used.
- Which kind of virtualization options could be used.
- Main aspects to remember when deploying a virtualization system.

Contents

- □ **What virtualization is.**
- □ Where virtualization could be used.
- □ Why virtualization is used.
- □ Which kind of virtualization options could be used.
- □ Main aspects to remember when deploying a virtualization system.

"Googling" for virtualization

- It is possible to search information from several sites:
  - Wikipedia: http://en.wikipedia.org/wiki/Virtualization
  - Forums: http://virt.kernelnewbies.org/
  - On-line courses: http://www.govirtual.org/docs/DOC-1024
  - Etc.

“Googling” for virtualization

□ Main aspects that we are going to find:

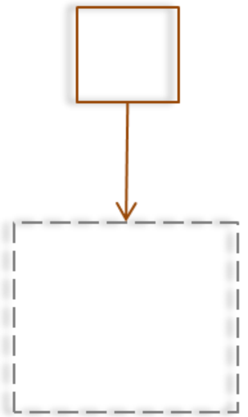
1. Virtualization term is not new:
   - It has been used since 60’s

Vintage !
☺

2. It has been applied to different aspects and areas of computing:
   - Components, servers, personal computers, etc.

Virtualization in computing

- Virtualization is a broad term that refers to the abstraction of computer resources.
- A technique for hiding the physical characteristics of computing resources from the way in which other systems, applications or end users interact with those resources.



Virtual resource

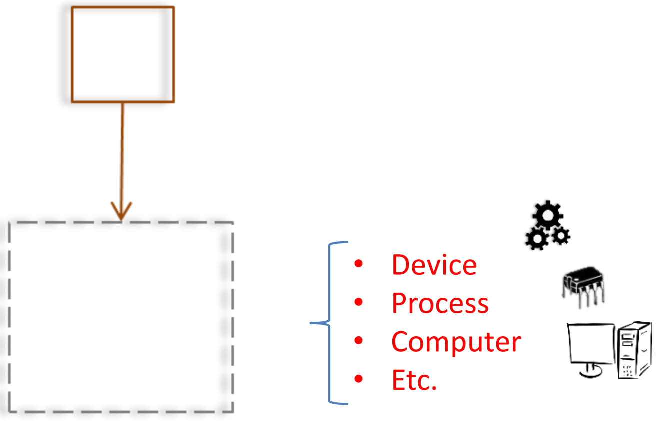
Contents

- What virtualization is.
- **Where virtualization could be used.**
- Why virtualization is used.
- Which kind of virtualization options could be used.
- Main aspects to remember when deploying a virtualization system.

Virtualization in computing

- □ Virtualization is a broad term that refers to the abstraction of computer resources.

- □ A technique for hiding the physical characteristics of computing resources from the way in which other systems, applications or end users interact with those resources.

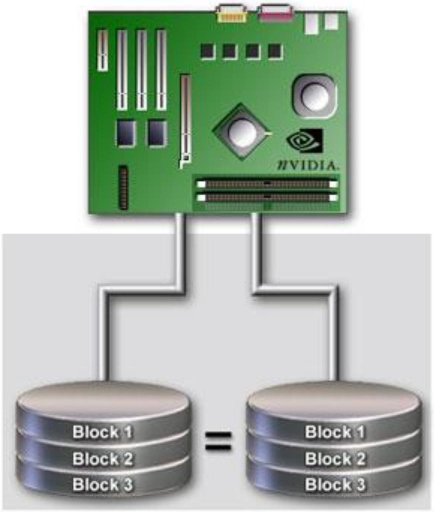


- Device
- Process
- Computer
- Etc.

Virtual resource

Virtualization examples

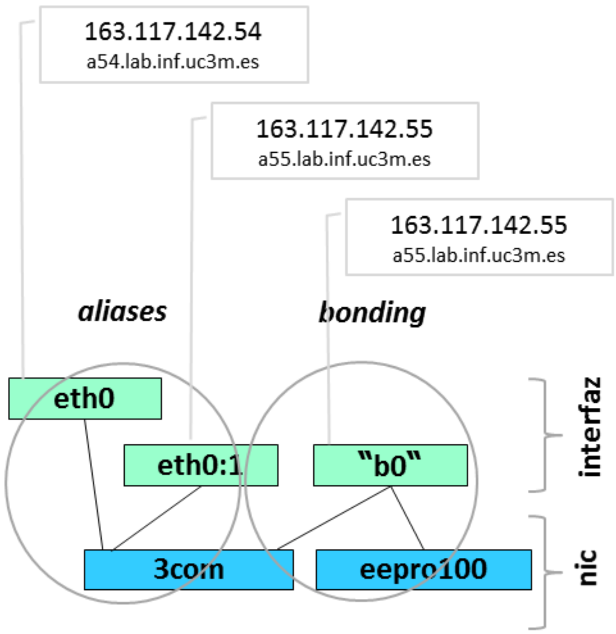
device level



- Storage device:
  - **E.g.: RAID**
- Networking:
  - E.g.: IP bonding, IP aliasing, NAT
- Processor:
  - E.g.: emulate a different instruction set

Virtualization examples

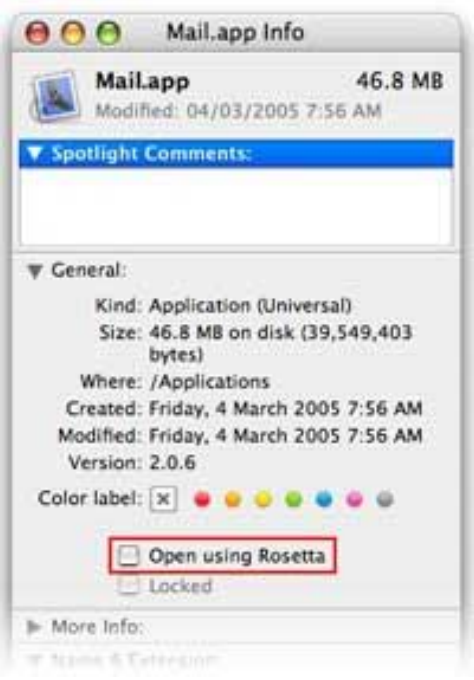
device level



- Storage device:
  - E.g.: RAID
- Networking:
  - **E.g.: IP bonding, IP aliasing, NAT**
- Processor:
  - E.g.: emulate a different instruction set

Virtualization examples

device level



Mail.app Info

Mail.app 46.8 MB
Modified: 04/03/2005 7:56 AM

▼ Spotlight Comments:

▼ General:

Kind: Application (Universal)
Size: 46.8 MB on disk (39,549,403 bytes)
Where: /Applications
Created: Friday, 4 March 2005 7:56 AM
Modified: Friday, 4 March 2005 7:56 AM
Version: 2.0.6

Color label: ☒ ● ● ● ● ● ● ●

☐ Open using Rosetta
☐ Locked

▶ More Info:

▼ Name & Extension:

Rosetta was used by Apple in the PowerPC to Intel transition.

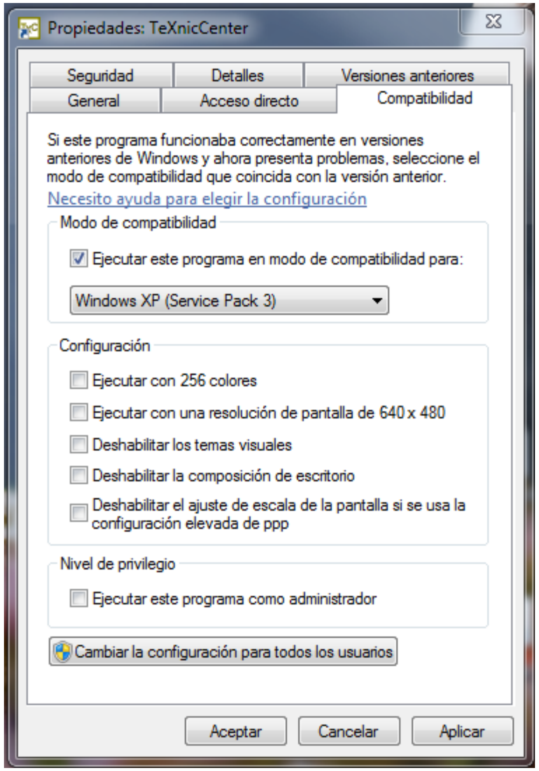
– Storage device:
  - E.g.: RAID

– Networking:
  - E.g.: IP bonding, IP aliasing, NAT

– Processor:
  - **E.g.: emulate a different instruction set**

Virtualization examples

process level



Si este programa funcionaba correctamente en versiones anteriores de Windows y ahora presenta problemas, seleccione el modo de compatibilidad que coincida con la versión anterior.

Necesito ayuda para elegir la configuración

Modo de compatibilidad

☑ Ejecutar este programa en modo de compatibilidad para:

Windows XP (Service Pack 3)

Configuración

□ Ejecutar con 256 colores

□ Ejecutar con una resolución de pantalla de 640 x 480

□ Deshabilitar los temas visuales

□ Deshabilitar la composición de escritorio

□ Deshabilitar el ajuste de escala de la pantalla si se usa la configuración elevada de ppp

Nivel de privilegio

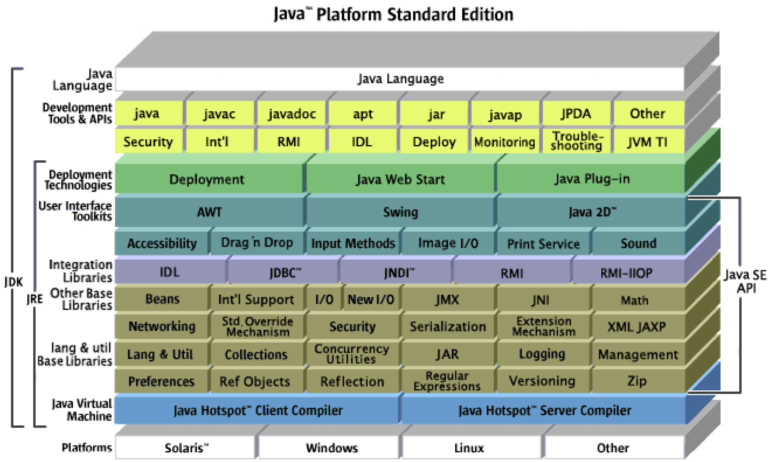
□ Ejecutar este programa como administrador

Cambiar la configuración para todos los usuarios

- Emulation of the application private environment:
  - **E.g.: Windows Vista/7 compatibility mode**
- Language level virtualization:
  - E.g.: Java and .NET

Virtualization examples

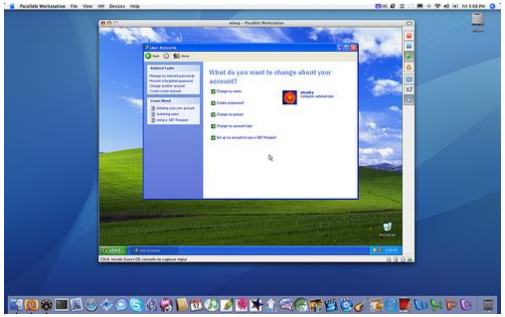
process level



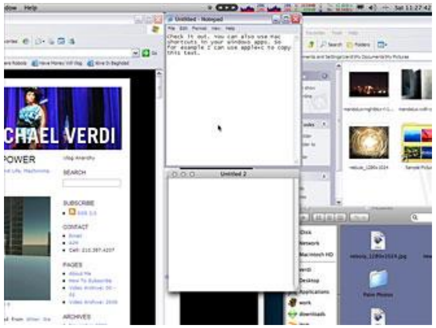
- Emulation of the application private environment:
  - E.g.: Windows Vista/7 compatibility mode
- Language level virtualization:
  - **E.g.: Java and .NET**

Virtualization examples

computer level

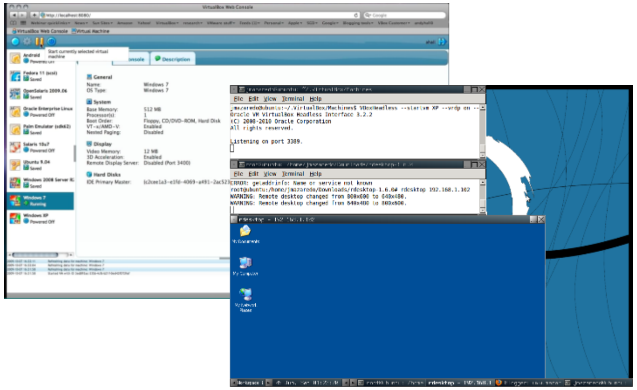


- Guest desktop as a window:
  - E.g.: VMWare, VirtualBox, etc.



```
Check it out. You can also use mac
shortcuts in your windows apps. So
for example I can use apple+c to copy
this text.
```

- A window for each guest application:
  - E.g.: Coherence mode, fluid view, etc.



```
jmazaredo@ubuntu:~/.VirtualBox/Machines$ VBoxHeadless --startvm XP --vrdp on --
Oracle VM VirtualBox Headless Interface 3.2.2
(C) 2008-2010 Oracle Corporation
All rights reserved.

Listening on port 3389.
```

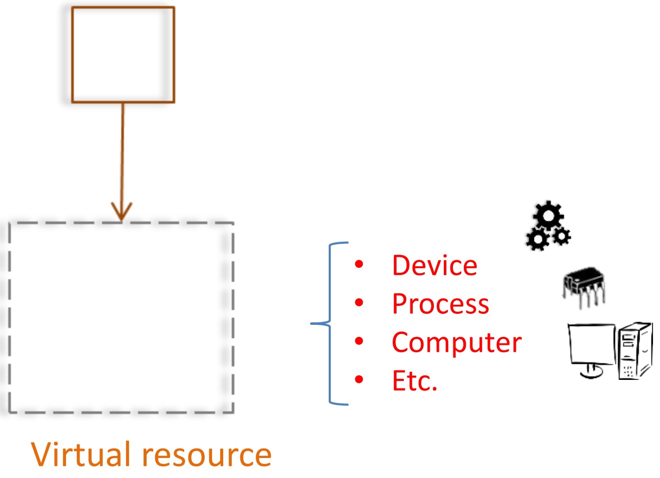
```
ERROR: getaddrinfo: Name or service not known
root@ubuntu:/home/jmazaredo/Downloads/rdesktop-1.6.0# rdesktop 192.168.1.102
WARNING: Remote desktop changed from 800x600 to 640x480.
WARNING: Remote desktop changed from 640x480 to 800x600.
```

- Remote desktop window:
  - E.g.: XEN, VMWare ESX,
    VirtualBox Headless, etc.

Virtualization in computing

summary

- □ Virtualization is a broad term that refers to the abstraction of computer resources.
- □ A technique for hiding the physical characteristics of computing resources from the way in which other systems, applications or end users interact with those resources.



Virtual resource

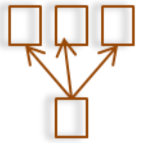
- Device
- Process
- Computer
- Etc.

Virtualization in computing

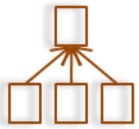
summary

- It includes:

  - To make a single physical resource (such as a server, an operating system, an application, etc.) be exposed as a different logical resource.

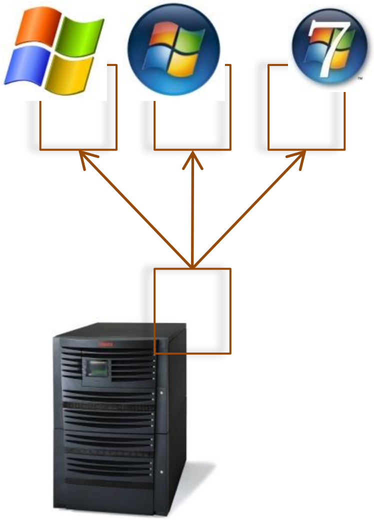


  - To make a single physical resource (such as storage devices, servers, etc.) be exposed as multiple logical resources.



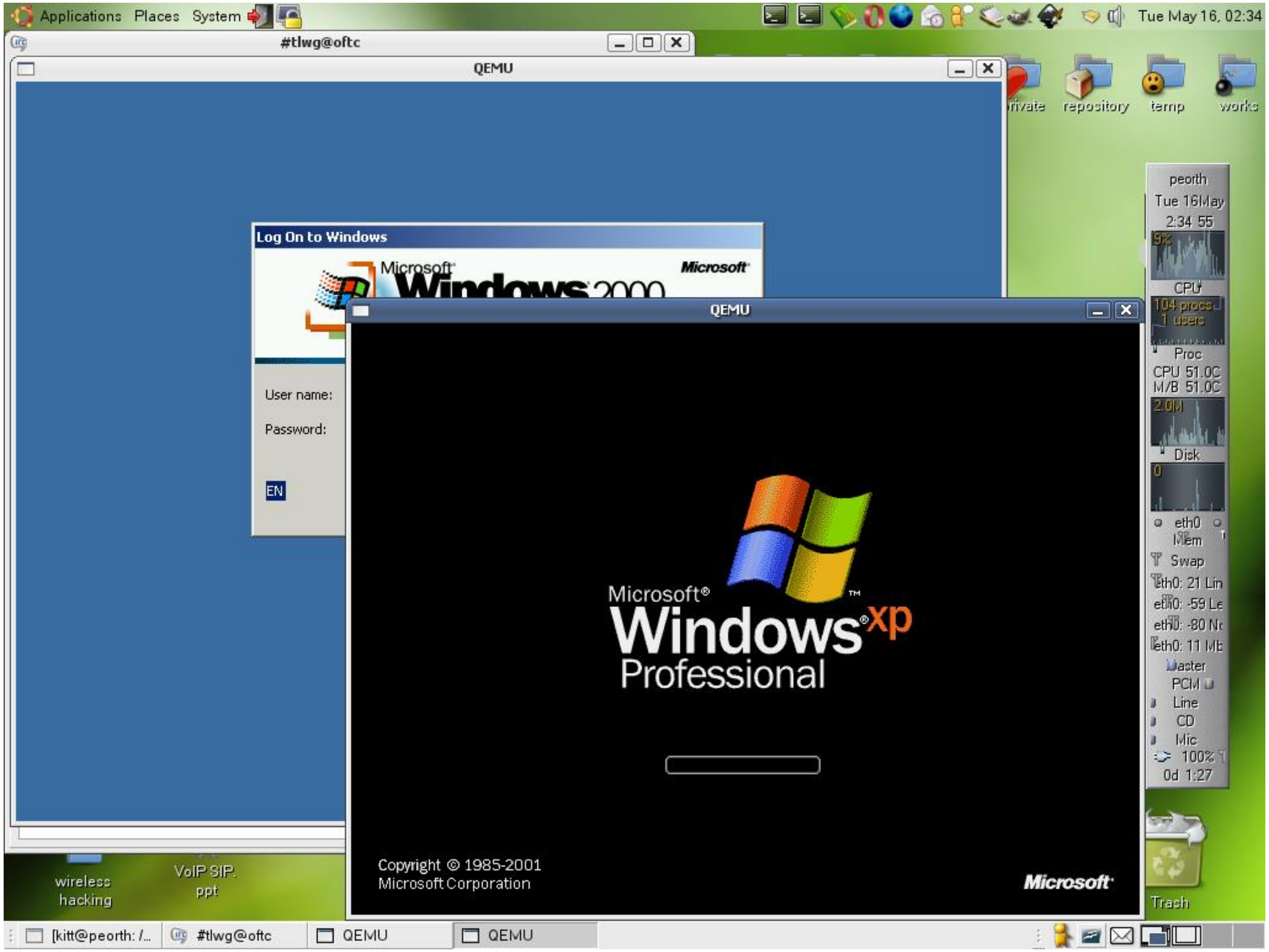
  - To make multiple physical resources (such as storage devices, servers, etc.) be exposed as a single logical resource.

Platform Virtualization



- □ We will focus on the platform virtualization in terms of virtual machines.
- □ The real system will be named as host system, the virtualization system will be named guest.

Example: Windows on Linux

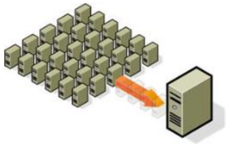


http://www.kitty.in.th/files/376/qemu.png

Contents

- □ What virtualization is.
- □ Where virtualization could be used.
- □ **Why virtualization is used.**
- □ Which kind of virtualization options could be used.
- □ Main aspects to remember when deploying a virtualization system.

Reasons for Virtualization



□ Server consolidation

□ Service isolation

□ Disaster recovery

□ Testing or training

□ Application portability

Consolidation

- Server consolidation:
  - To reduce costs (by multiplexing resources).
  - Simplifying the administration and management.

1000 €

1000 €

1000 €

1000 €

Consolidation

- Server consolidation:
  - To reduce costs (by multiplexing resources).
  - Simplifying the administration and management

3000 €

Insolation

- Improve security:
  - Insolate services in different computers.
  - Different security policy for each computer.

.COM
.ORG
.NET
Monitor não incluso

Insolation

- Improve security:
  - Insolate services in different computers.
  - Different security policy for each computer.

.COM
.ORG
.NET
Monitor não incluso

Disaster recovery

- Improve disaster recovery:
  - Hot-spare machine(s).
  - Automatic work re-routing while rebooting/fixing.

Universidad
Carlos III de Madrid
www.uc3m.es

Disaster recovery

ARCOS

- Improve disaster recovery:
  - Hot-spare machine(s).
  - Automatic work re-routing while rebooting/fixing.

Disaster recovery

- Improve disaster recovery:
  - Hot-spare machine(s).
  - Automatic work re-routing while rebooting/fixing.

Monitor não incluso
Monitor não incluso

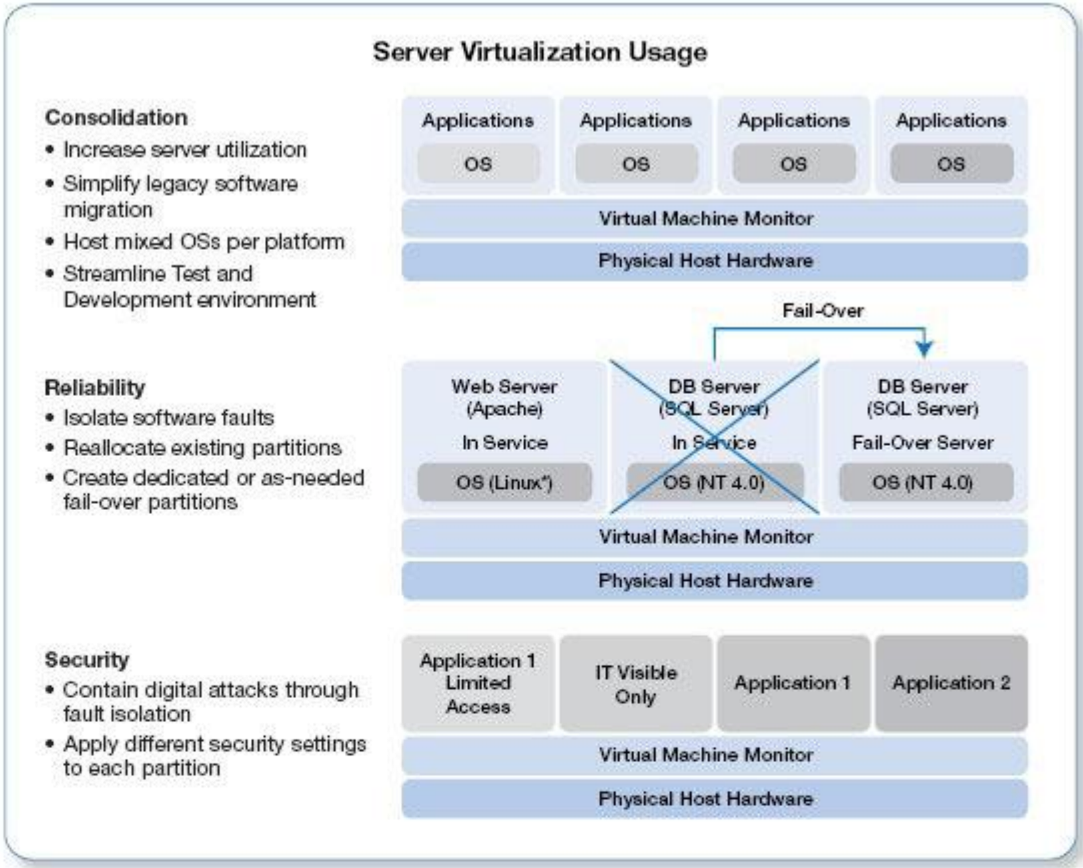
Testing or training

- Better testing environment:
  - It enables the execution in other work environment.
  - It improves the restoring process (easier/bit-faster).

Testing or training

- Better testing environment:
  - It enables the execution in other work environment.
  - It improves the restoring process (easier/bit-faster).

Application portability



Server Virtualization Usage

Consolidation

- Increase server utilization
- Simplify legacy software migration
- Host mixed OSs per platform
- Streamline Test and Development environment

Reliability

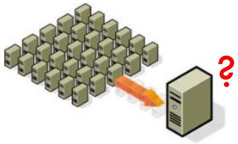
- Isolate software faults
- Reallocate existing partitions
- Create dedicated or as-needed fail-over partitions

Security

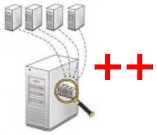
- Contain digital attacks through fault isolation
- Apply different security settings to each partition

http://www.intel.com/technology/magazine/computing/intel-virtualization-0405.htm

Reasons for **NOT** using Virtualization



- □ Complex dimensioning



- □ More resources per node are needed

- □ Double administration level

- □ Some performance loss

Complex dimensioning

- An appropriated sizing is required:
  - The (virtual) servers might change it requirements (memory, cpu, …)
  - An unappropriated sizing has impact in all (virtual) servers.

Monitor não incluso
Monitor não incluso
Monitor não incluso

Big resources

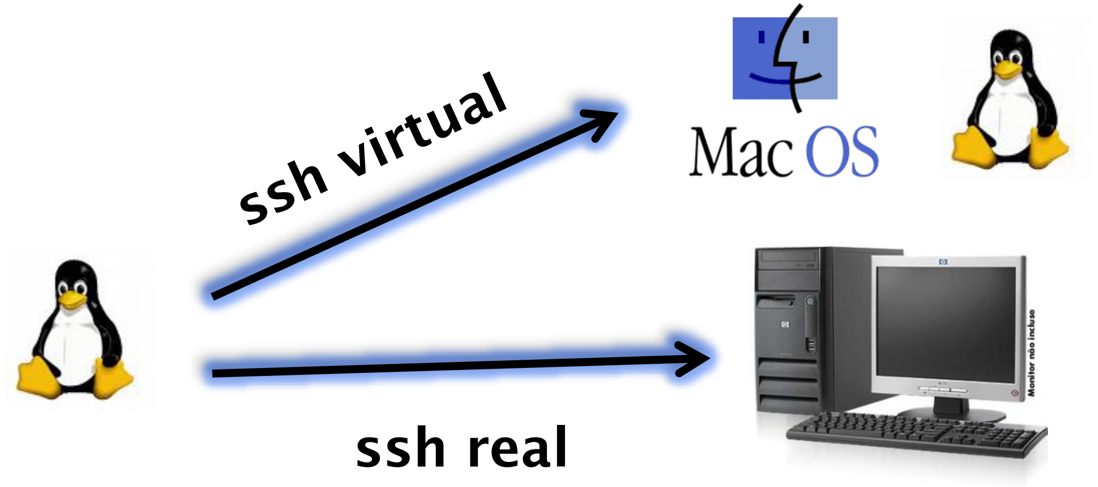
- More resources per host are needed:
  - 8 1GiB servers consolidated on 1 8GiB server
  - Not always is easy (and cheaper) to buy one server with n network cards, n GB of RAM, n TB of disk, etc.

Big resources

- More resources per host are needed:
  - 8 1GiB servers consolidated on 1 8GiB server
  - Not always is easy (and cheaper) to buy one server with n network cards, n GB of RAM, n TB of disk, etc.

Two administration levels

- Double administration level:
  - (Real) host computers
  - (Virtual) guest computers
  - Management of host/guest relationship
    - If a host computer has problems, all guest computer has to be migrated.



Loss of performance

- A "little" loss of performance:
  - In CPU could be low: between 3% and 12%
  - Graphic card and buses bandwidth?
  - Hard disk shared among several guest computers?

Contents

- What virtualization is.
- Where virtualization could be used.
- Why virtualization is used.
- **Which kind of virtualization options could be used.**
- Main aspects to remember when deploying a virtualization system.

Many options…

- □ Good news: many options…

User Mode Linux

User Mode Linux

QEMU
open source processor emulator

IBM
System z™

OpenVZ

Many options…

- Bad news: many options…

User Mode Linux

QEMU open source processor emulator

IBM System z™

bochs 2.3

OpenVZ

vmware®

Xen™

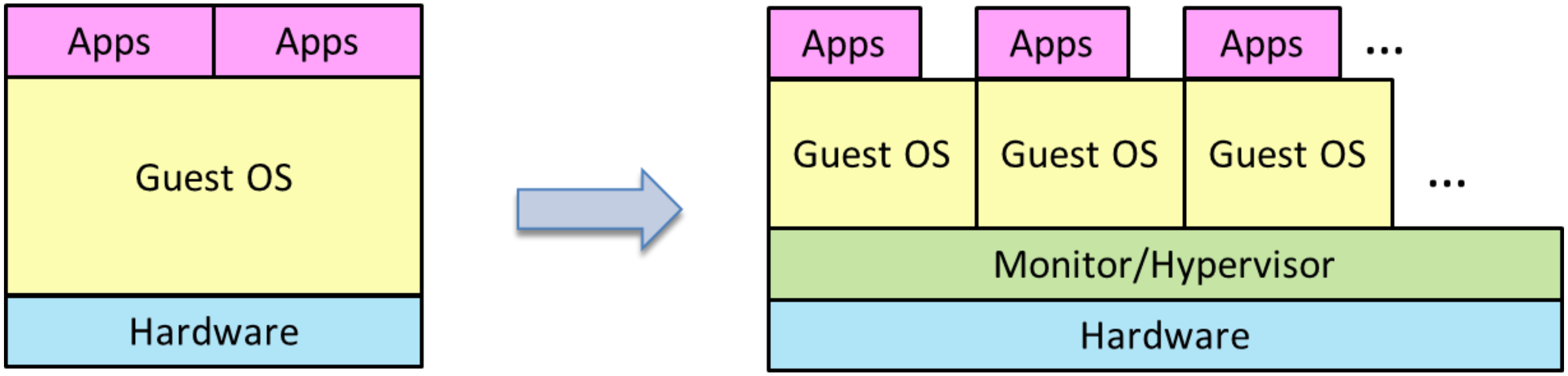
KVM

Many options…

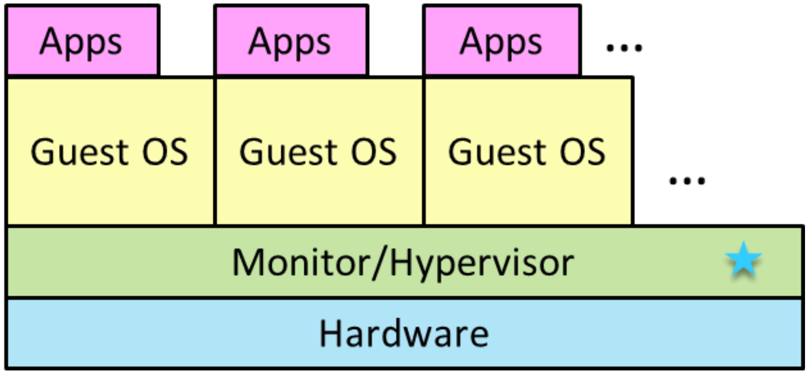
- To know how internally works.
  - Dependencies, restrictions, etc.

- To know the important details about virtualization system architectures.
  - To group solutions by common characteristics.

How internally is…

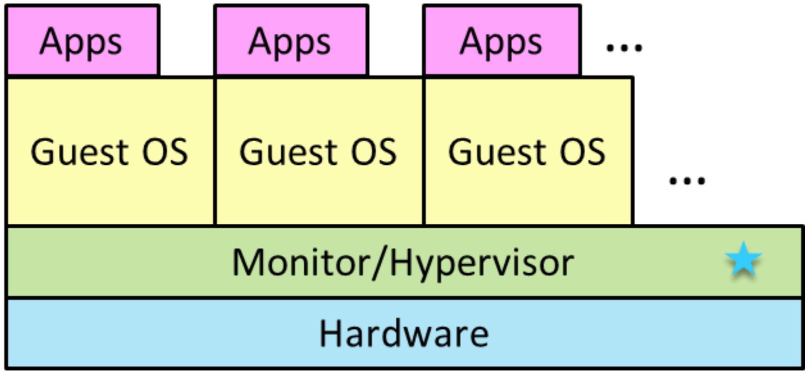


How internally is…



- A new layer is added between the operating system and hardware
  - It will 'talk' with all kind of hardware
  - Arbitrate hardware resources across all operating systems
- Each operating system is executed in privileged mode but the monitor/hypervisor intercepts its requests to server them.

How internally is…

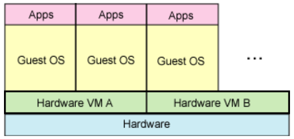


1

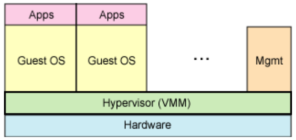
- A new layer is added between the operating system and hardware
  - 3 It will 'talk' with all kind of hardware
  - 4 Arbitrate hardware resources across all operating systems
- Each operating system is executed in privileged mode but the monitor/hypervisor intercepts its requests to server them.

2

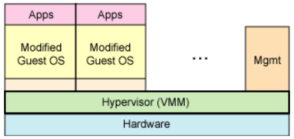
Types of Virtualization


Apps Apps Apps
Guest OS Guest OS Guest OS …
Hardware VM A Hardware VM B
Hardware

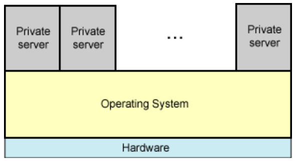
□ Hardware Emulation


Apps Apps
Guest OS Guest OS … Mgmt
Hypervisor (VMM)
Hardware

□ Full Virtualization


Apps Apps
Modified Guest OS Modified Guest OS … Mgmt
Hypervisor (VMM)
Hardware

□ Para-virtualization

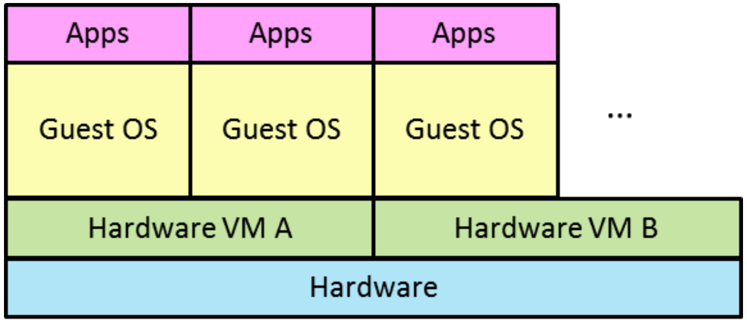

Private server Private server … Private server
Operating System
Hardware

□ Containers

http://www-128.ibm.com/developerworks/library/l-linuxvirt/index.html

Hardware emulation

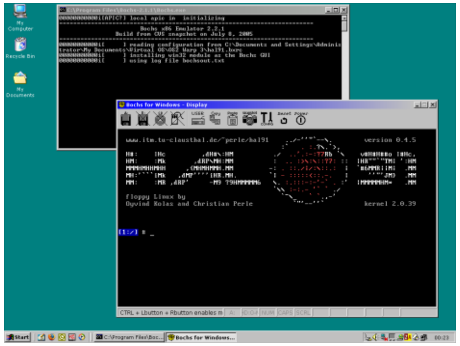
- A virtual machine on the host system is created to emulate the target hardware.
- Advantage: you can execute software for CPU1 on CPU2 without modifications.
- Disadvantage: s-l-o-w (about x100)



Hardware emulation

bochs 2.3

- Bochs



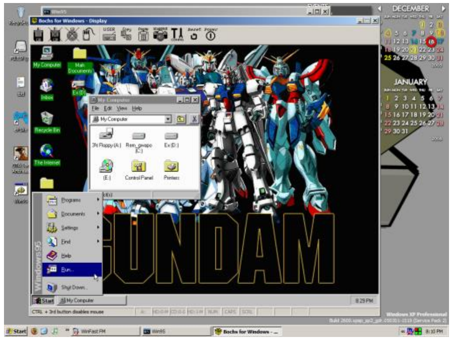
```
00000000000i[APIC?] local apic in  initializing
========================================================================
                  Bochs x86 Emulator 2.2.1
             Build from CVS snapshot on July 8, 2005
========================================================================
00000000000i[     ] reading configuration from C:\Documents and Settings\Administrator\My Documents\Virtual OS\OS2 Warp 3\hal91.bxrc
00000000000i[     ] installing win32 module as the Bochs GUI
00000000000i[     ] using log file bochsout.txt
```

```
 www.itm.tu-clausthal.de/~perle/hal91                     version 0.4.5

 floppy Linux by
 Oyvind Kolas and Christian Perle                          kernel 2.0.39

[1:/] # _
```

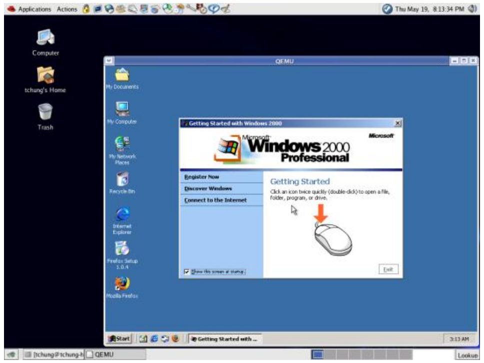
Linux/W95

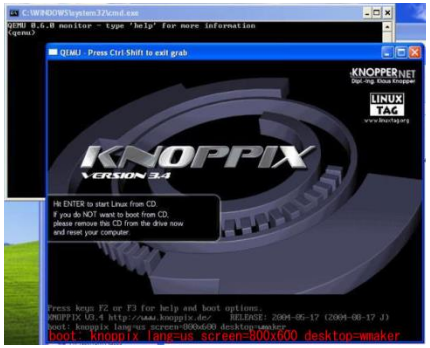


W95/WXP

- Qemu

QEMU
open source processor emulator





```
QEMU 0.6.0 monitor - type 'help' for more information
(qemu)
```

Hit ENTER to start Linux from CD.
If you do NOT want to boot from CD,
please remove this CD from the drive now
and reset your computer.

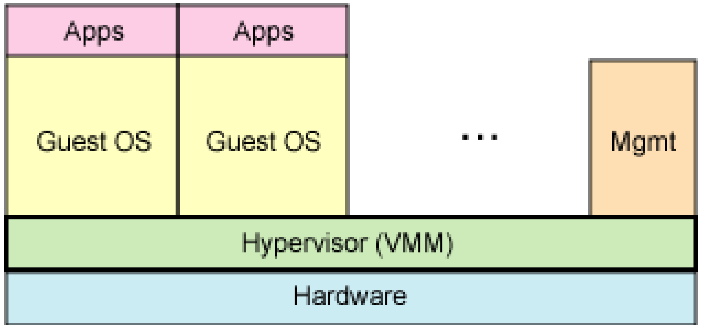
```
Press keys F2 or F3 for help and boot options.
KNOPPIX V3.4 http://www.knoppix.de/    RELEASE: 2004-05-17 (2004-08-17 J)
boot: knoppix lang=us screen=800x600 desktop=wmaker
boot: knoppix lang=us screen=800x600 desktop=wmaker
```

Full Virtualization

- Hardware is shared among all guest operating systems through a **hypervisor.**

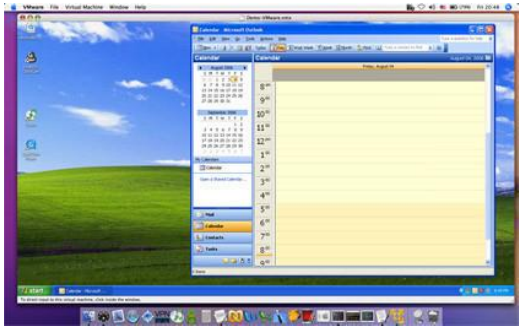
- Advantage: the operating system do not need to be modified.

- Disadvantage: it is necessary to intercept the access of the operating system to the hardware:
  - Hardware support
  - On-the-fly binary patching

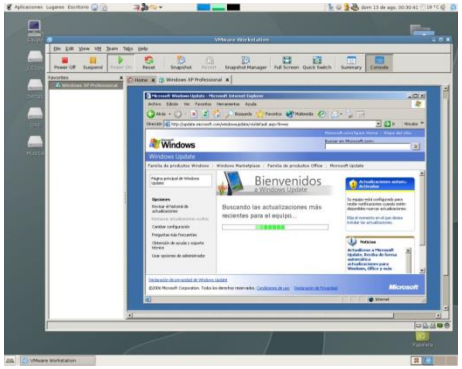


Full Virtualization

- ❑ VMware

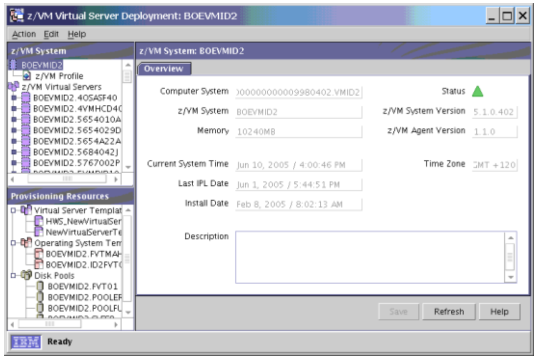


WXP/MacOS



WXP/Linux

- ❑ z/VM



z/VM Virtual Server Deployment: BOEVMID2

z/VM System
BOEVMID2
z/VM Profile
z/VM Virtual Servers
BOEVMID2.40SASF40
BOEVMID2.4VMHCD4C
BOEVMID2.5654010A
BOEVMID2.5654029D
BOEVMID2.5654A22A
BOEVMID2.5684042J
BOEVMID2.5767002P

Provisioning Resources
Virtual Server Templat
HWS_NewVirtualSer
NewVirtualServerTe
Operating System Tem
BOEVMID2.FVTMAH
BOEVMID2.ID2FVTC
Disk Pools
BOEVMID2.FVT01
BOEVMID2.POOLEF
BOEVMID2.POOLFL

z/VM System: BOEVMID2

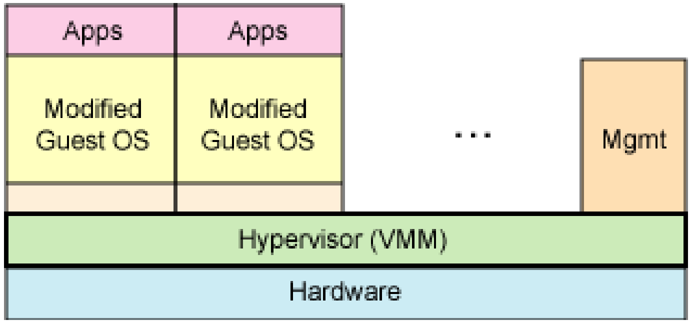
Computer System:)000000000009980402.VMID2
Status
z/VM System: BOEVMID2
z/VM System Version: 5.1.0.402
Memory: 10240MB
z/VM Agent Version: 1.1.0
Current System Time: Jun 10, 2005 / 4:00:46 PM
Time Zone: ЭMT +120
Last IPL Date: Jun 1, 2005 / 5:44:51 PM
Install Date: Feb 8, 2005 / 8:02:13 AM
Description

IBM Ready

z/Linux sobre z/VM

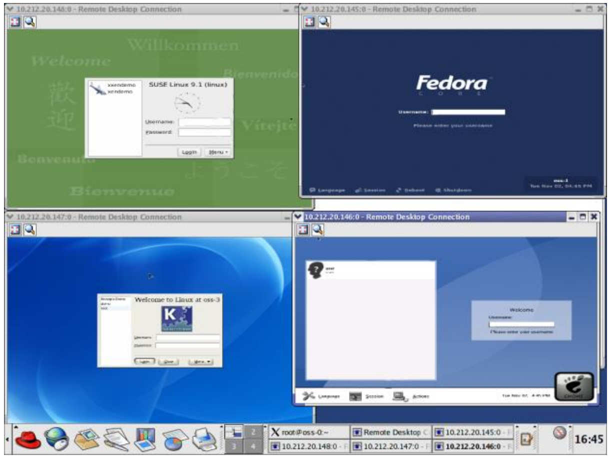
Para-Virtualization

- □ Similar to the full-virtualization but the guest software collaborate with hypervisor.
- □ Advantage: The operating system works with the hypervisor (less wasted time by interception mechanism).
- □ Disadvantage: The operating system has to be modified in order to interact with the hypervisor.



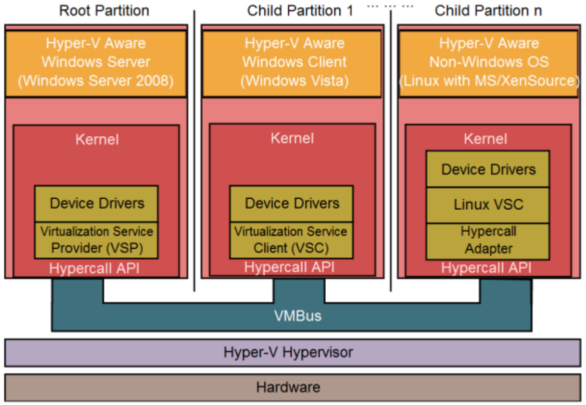
Para-Virtualization

❑ XEN


10.212.20.148:0 - Remote Desktop Connection
10.212.20.145:0 - Remote Desktop Connection
Willkommen
Welcome
Bienvenido
SUSE Linux 9.1 (linux)
Username:
Password:
Login
Menu
Vitejte
Benvenuto
Bienvenue
Fedora
Username:
10.212.20.147:0 - Remote Desktop Connection
10.212.20.146:0 - Remote Desktop Connection
Welcome to Linux at oss-3
Welcome
16:45

Linux(S,F,D)/Linux(F)

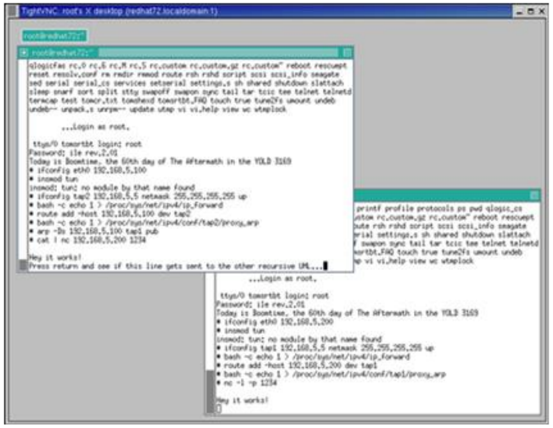
❑ Hiper-v


Root Partition
Child Partition 1
Child Partition n
Hyper-V Aware Windows Server (Windows Server 2008)
Hyper-V Aware Windows Client (Windows Vista)
Hyper-V Aware Non-Windows OS (Linux with MS/XenSource)
Kernel
Kernel
Kernel
Device Drivers
Virtualization Service Provider (VSP)
Hypercall API
Device Drivers
Virtualization Service Client (VSC)
Hypercall API
Device Drivers
Linux VSC
Hypercall Adapter
Hypercall API
VMBus
Hyper-V Hypervisor
Hardware

Windows S2008+Vista+Suse / Windows

Para-Virtualization

- ❑ User Mode Linux (UML)



```
qlogicfas rc.0 rc.6 rc.M rc.S rc.custom rc.custom.gz rc.custom~ reboot rescuept
reset resolv.conf rm rmdir rmmod route rsh rshd script scsi scsi_info seagate
sed serial serial_cs services setserial settings.s sh shared shutdown slattach
sleep snarf sort split stty swapoff swapon sync tail tar tcic tee telnet telnetd
termcap test tomcr.txt tomshexd tomsrtbt.FAQ touch true tune2fs umount undeb
undeb~~ unpack.s unrpm~~ update utmp vi vi.help view wc wtmplock

        ...Login as root.

 ttys/0 tomsrtbt login: root
Password: ile rev.2.01
Today is Boomtime, the 60th day of The Aftermath in the YOLD 3169
# ifconfig eth0 192.168.5.100
# insmod tun
insmod: tun: no module by that name found
# ifconfig tap2 192.168.5.5 netmask 255.255.255.255 up
# bash -c echo 1 > /proc/sys/net/ipv4/ip_forward
# route add -host 192.168.5.100 dev tap2
# bash -c echo 1 > /proc/sys/net/ipv4/conf/tap2/proxy_arp
# arp -Ds 192.168.5.100 tap1 pub
# cat | nc 192.168.5.200 1234

Hey it works!
Press return and see if this line gets sent to the other recursive UML...
```

```
 printf profile protocols ps pwd qlogic_cs
ustom rc.custom.gz rc.custom~ reboot rescuept
oute rsh rshd script scsi scsi_info seagate
erial settings.s sh shared shutdown slattach
f swapon sync tail tar tcic tee telnet telnetd
ortbt.FAQ touch true tune2fs umount undeb
p vi vi.help view wc wtmplock

        ...Login as root.

 ttys/0 tomsrtbt login: root
Password: ile rev.2.01
Today is Boomtime, the 60th day of The Aftermath in the YOLD 3169
# ifconfig eth0 192.168.5.200
# insmod tun
insmod: tun: no module by that name found
# ifconfig tap1 192.168.5.5 netmask 255.255.255.255 up
# bash -c echo 1 > /proc/sys/net/ipv4/ip_forward
# route add -host 192.168.5.200 dev tap1
# bash -c echo 1 > /proc/sys/net/ipv4/conf/tap1/proxy_arp
# nc -l -p 1234

Hey it works!
```

Linux/Linux

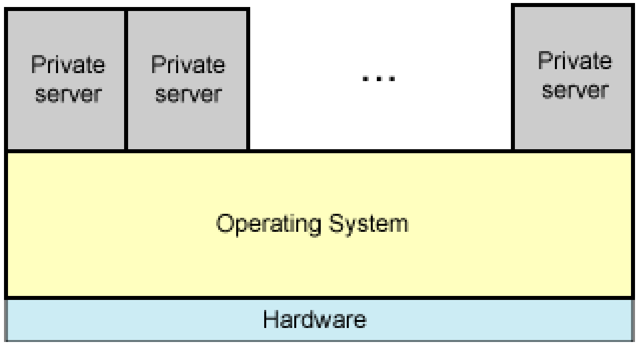
2004

SIMULACIÓN DE UN CLUSTER USANDO *USER MODE LINUX*

AUTOR: VICTOR INIESTA SAMPAYO

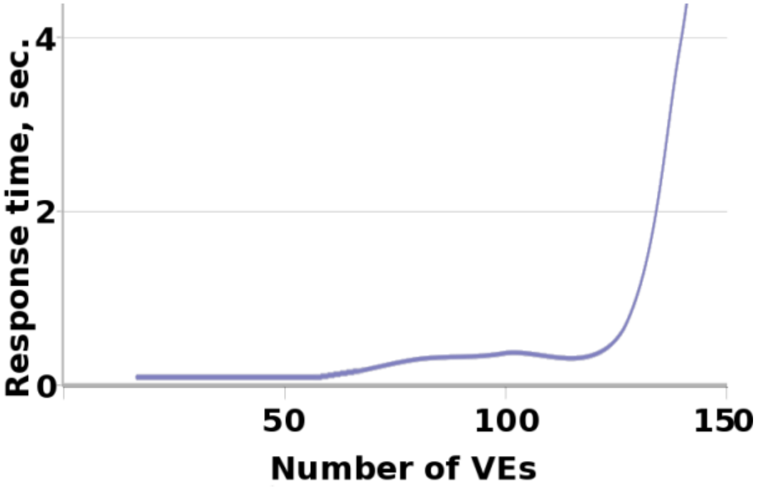
Containers

- Different approach: the operating system provides virtual copies of itself.
- Advantage: it is NOT possible to execute different operating systems.
- Disadvantage: best performance and more number of virtual machines executing (with less memory).



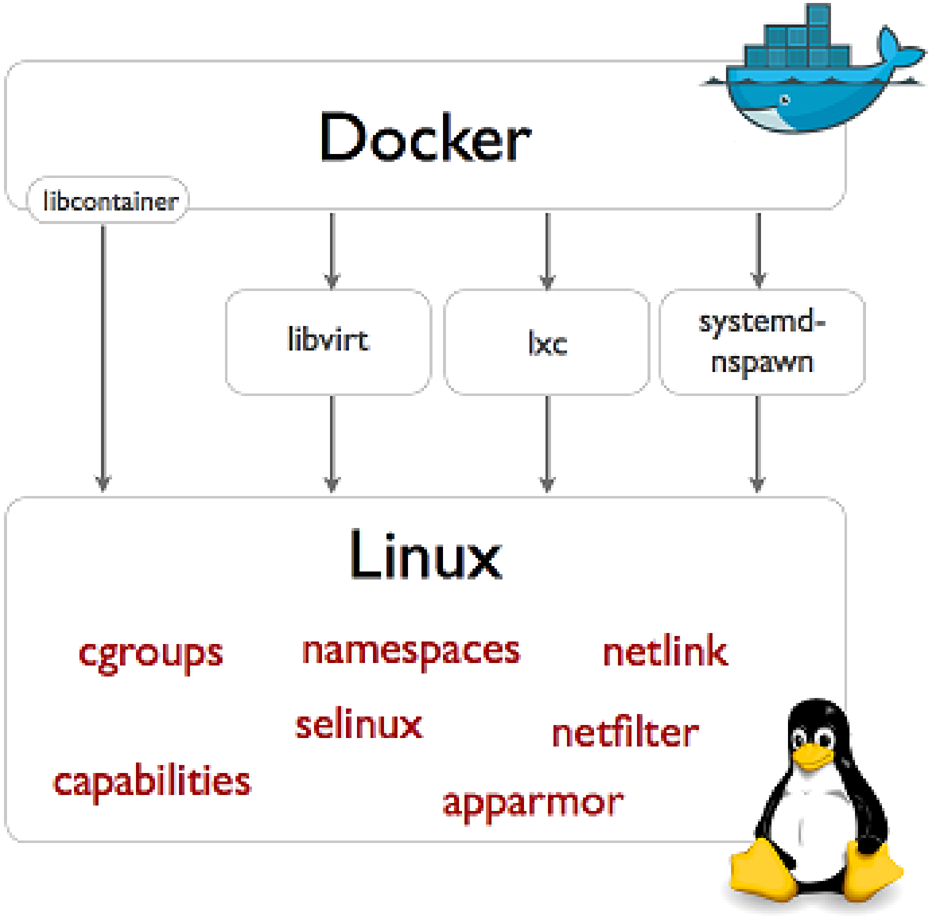
Containers

- ❑ OpenVZ



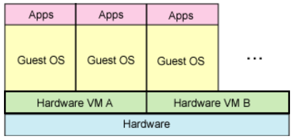
**Density of OpenVZ in a
768 MiB (¾ Gb) RAM computer**

Containers: docker

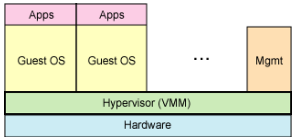


Types of Virtualization

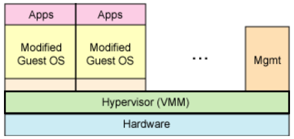
summary



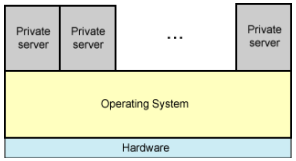
- □ Hardware Emulation



- □ Full Virtualization



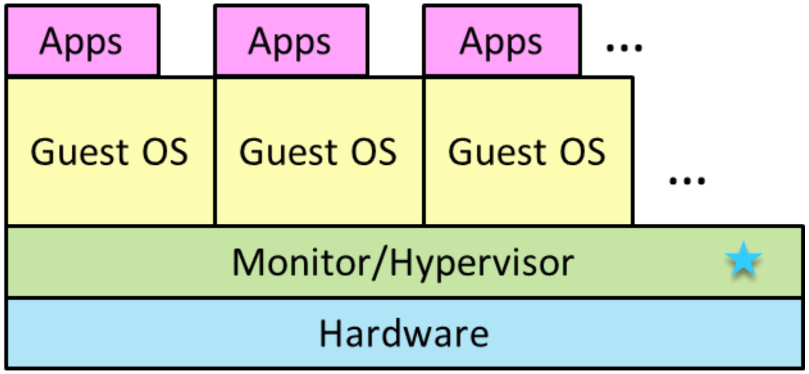
- □ Para-virtualization



- □ Containers

http://www-128.ibm.com/developerworks/library/l-linuxvirt/index.html

How internally is…

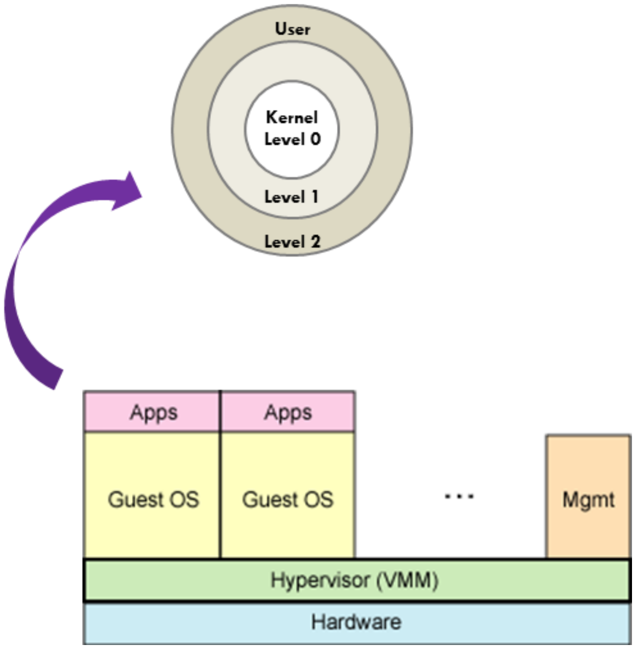


- □ A new layer is added between the operating system and hardware
  - 3 ◘ It will 'talk' with all kind of hardware
  - 4 ◘ Arbitrate hardware resources across all operating systems
- □ Each operating system is executed in privileged mode but the monitor/hypervisor intercepts its requests to server them.

  2

Full/Para-virtualization

Interception

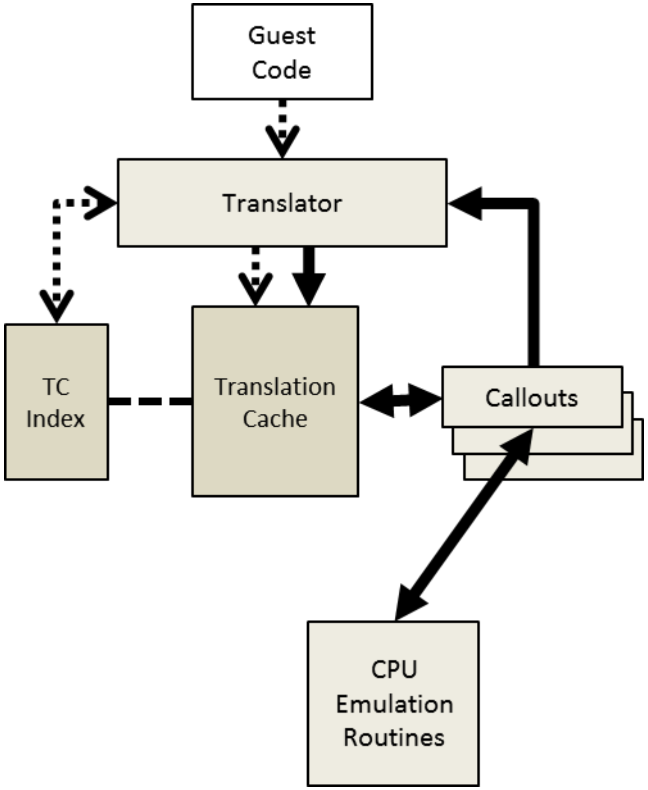


- Regular operating system has been designed to be executed in hardware in privileged mode.
  - In x86 processors, on ring 0
- But now the privileged code has to be executed without been privileged anymore (the hypervisor is now privileged)
  - Binary patching
  - New virtualization instructions

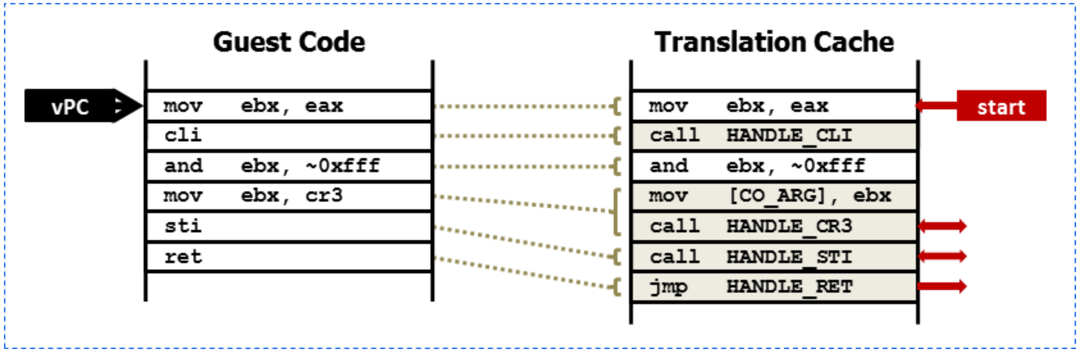
http://pdc-amd01.poly.edu/~wein/cs6243/ppts/CPUVirtualization.pptx

Full/Para-virtualization

Interception



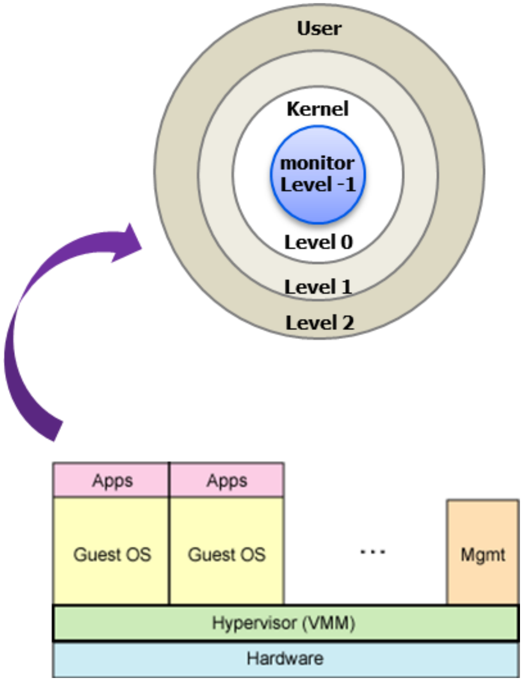
- Binary translation/patching:
  - Patching the instructions on the fly.
  - The guest code is analyzed and the privileged instructions are replaced with hypervisor calls.
  - Speed-up by caching the patched fragments.
  - Advantage: Can be used on any kind of CPU.
  - Disadvantage: S-l-o-w.



```
Guest Code                  Translation Cache
mov   ebx, eax              mov   ebx, eax
cli                         call  HANDLE_CLI
and   ebx, ~0xfff           and   ebx, ~0xfff
mov   ebx, cr3              mov   [CO_ARG], ebx
                            call  HANDLE_CR3
sti                         call  HANDLE_STI
ret                         jmp   HANDLE_RET
```

Full/Para-virtualization

Interception



- Special hardware Instructions :
  - Ring '-1' where the hypervisor is executed.
    - It reduce the performance penalty of dynamic on-the-fly translation.
  - Intel and AMD have developed instructions set extensions for virtualization. There are similar but no compatible.
  - Advantage: Fast request to the hypervisor.
  - Disadvantage: It needs special CPU support.

Intel has:

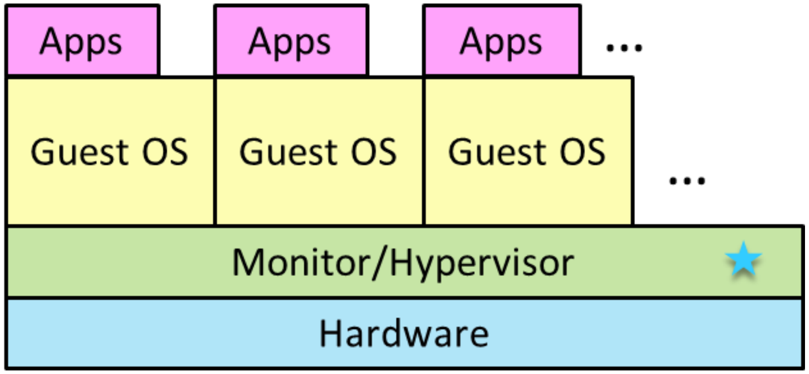
- VT-x as extensions IVT for IA-32 (Vanderpool)
- VT-i as extensions IVT for IA-64 (Silvervale)
- VT-d in 32/64 for Directed I/O

AMD has:

- AMD-V (Pacifica) for 32/64
- IOMMU as Directed I/O or PCI-Passthrough

http://en.wikipedia.org/wiki/Virtualization_Technology

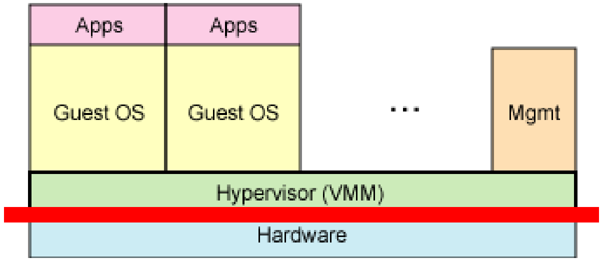
How internally is…



- A new layer is added between the operating system and hardware
  - 3 It will 'talk' with all kind of hardware
  - 4 Arbitrate hardware resources across all operating systems
- Each operating system is executed in privileged mode but the monitor/hypervisor intercepts its requests to server them.

Hypervisor

Deployment



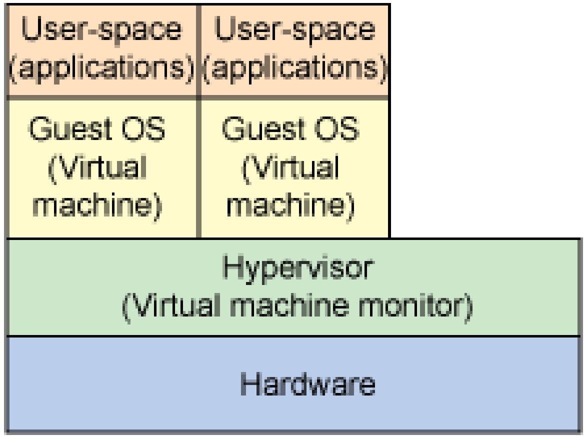
- The monitor/hypervisor have to be able to work with all kind of hardware
  - It has to have drivers for all hardware.
  - It is very difficult to get drivers for all existing (and new) hardware.
- But we can use a modified operating system as monitor/hypervisor:
  - Hosted or Split
  - “Pure” Hypervisor

http://www.ibm.com/developerworks/linux/library/l-pci-passthrough/

Type 1 hypervisor

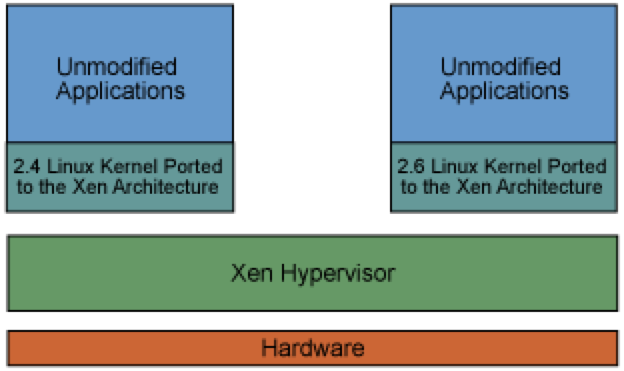
Deployment

- Hypervisor
  - To remove from an existent operating system everything but what is needed to transform it into a hypervisor.
  - The hypervisor boots first, and then every virtual machine that uses it:
    - A: less interferences between guests
    - D: no so easier to install


User-space (applications)
User-space (applications)
Guest OS (Virtual machine)
Guest OS (Virtual machine)
Hypervisor (Virtual machine monitor)
Hardware

Type 1 hypervisor

Deployment


Unmodified Applications
2.4 Linux Kernel Ported to the Xen Architecture
Unmodified Applications
2.6 Linux Kernel Ported to the Xen Architecture
Xen Hypervisor
Hardware

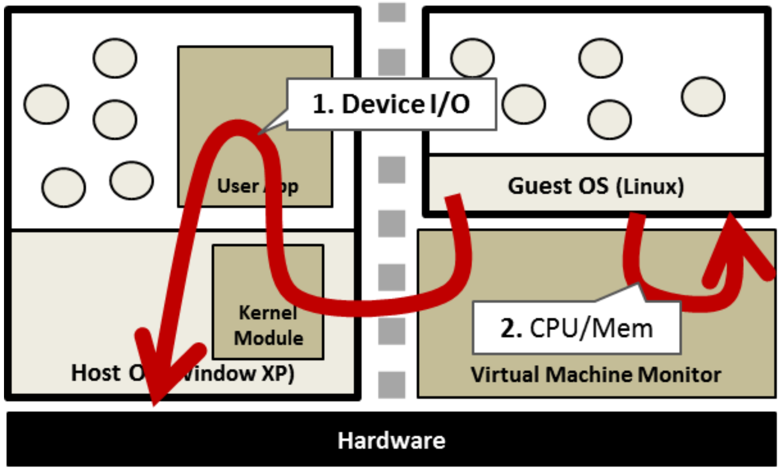
- XEN is now included in the Linux kernel (4.6 in progress)
- XEN could be described as a Linux system to which all has been remove but the base to be used as hypervisor.
  - Initially designed as para-virtualization system.

http://www.ibm.com/developerworks/linux/library/l-xen/

Type 2 hypervisor

Deployment

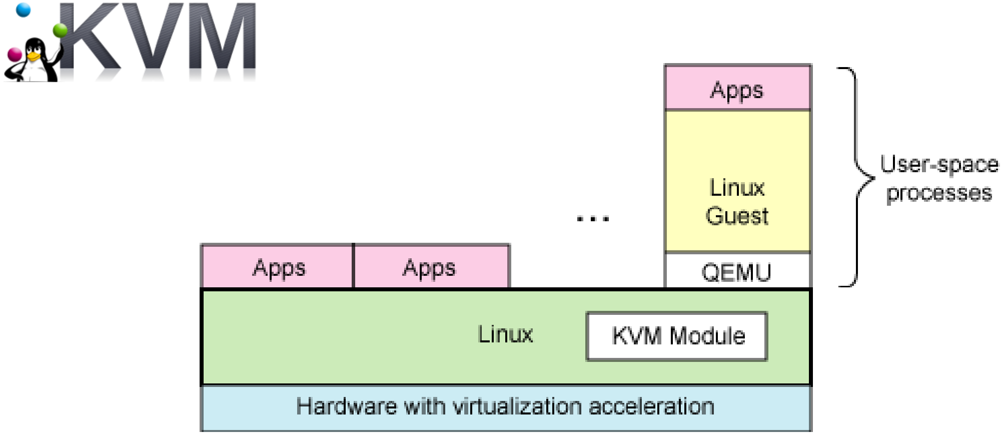
- Hosted or Split
  - Transform an existent operating system into a hypervisor.
  - A V.M. is a process in the host system:
    - D: Double scheduling
    - D: Expensive access to the hardware
    - A: Easy to install (like a familiar application)



http://www.govirtual.org/servlet/JiveServlet/download/1024-4-1042/IntroToVMs.pptx

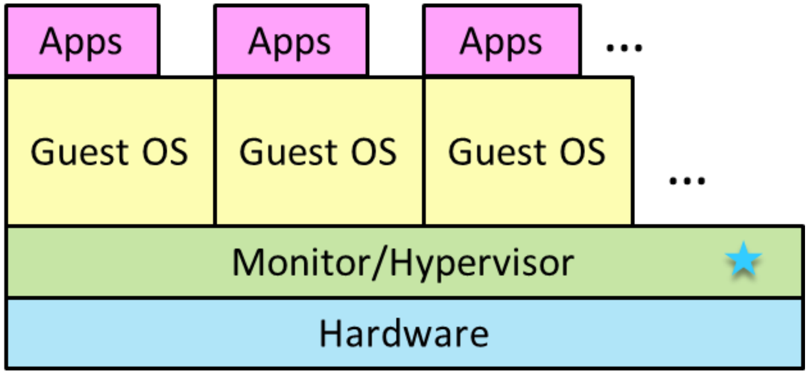
Type 2 hypervisor

Deployment



- KVM is include in the Linux kernel since version 2.6.20
- KVM transforms the Linux kernel into an hypervisor as a module
  - Other guest operating system can be executed in user-space.
  - It use a modified QEMU process.

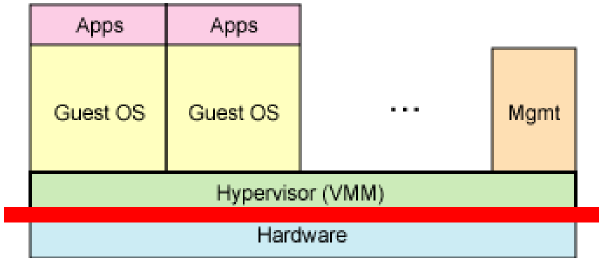
How internally is…



- A new layer is added between the operating system and hardware
  - It will ‘talk’ with all kind of hardware
  - 4 Arbitrate hardware resources across all operating systems
- Each operating system is executed in privileged mode but the monitor/hypervisor intercepts its requests to server them.

Full/Para-virtualization

device accesses



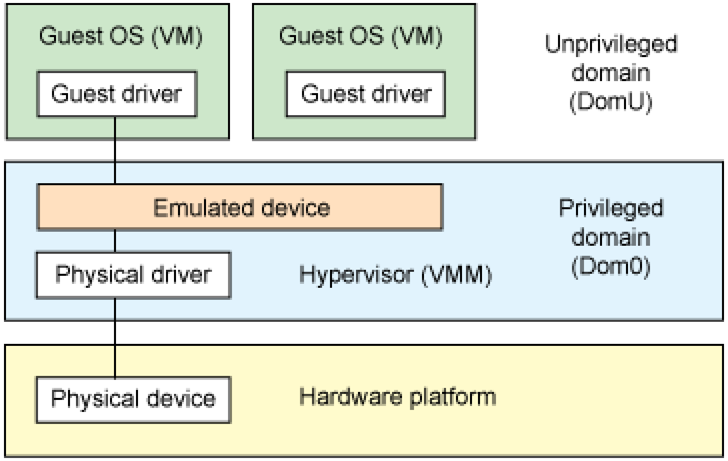
- The monitor/hypervisor must be able to deal with all types of hardware:
  - It must have driver for all devices.
  - It provides access to the underlying hardware.
- Expose the hardware to the guest operating system:
  - Hypervisor device emulation.
  - User-space device emulation.
  - Gateway to device
  - SR-IOV and MR-IOV

http://www.ibm.com/developerworks/linux/library/l-pci-passthrough/

Full/Para-virtualization

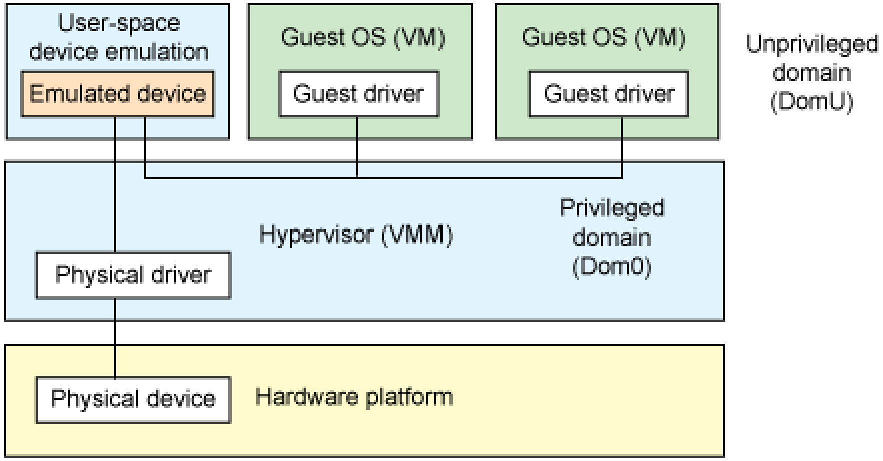
device accesses

- Hypervisor device emulation.
  - E.g.: VMware workstation
  - Advantage: easy to migrate



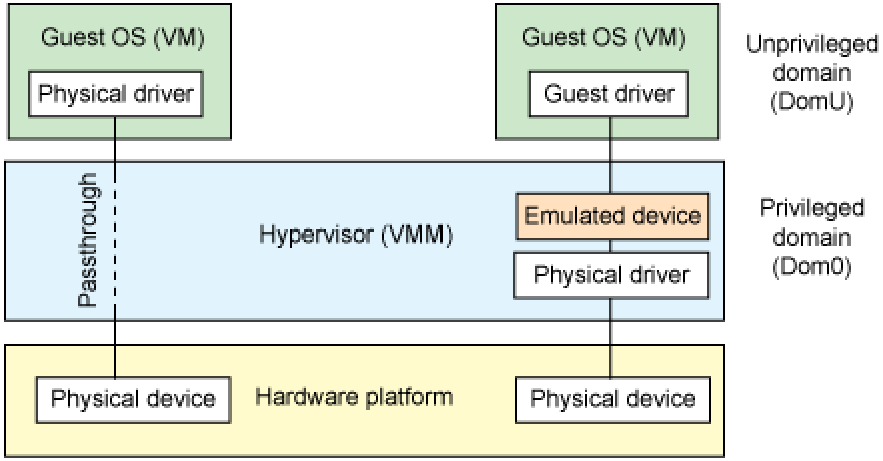
Full/Para-virtualization
device accesses

- User-space device emulation.
  - E.g.: KVM
  - Advantage: easy to migrate (even to other hypervisor) and safe (no privileged)



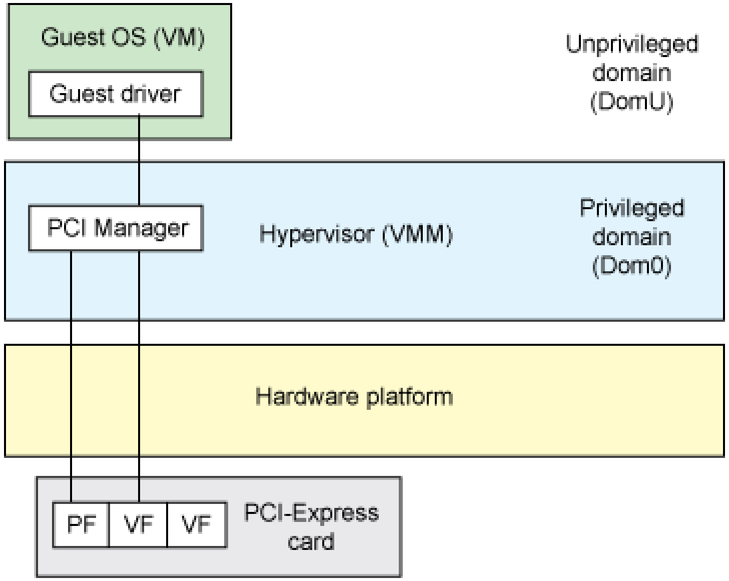
Full/Para-virtualization
device accesses

- Gateway to device.
  - E.g.: VMware, XEN, etc.
  - Advantage: efficient

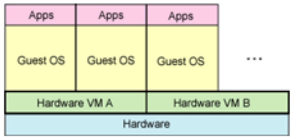


Full/Para-virtualization
device accesses

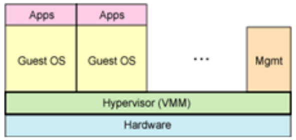
- SR-IOV and MR-IOV.
  - Single-Root I/O Virtualization (one server)
  - Multi-Root I/O Virtualization (blades)



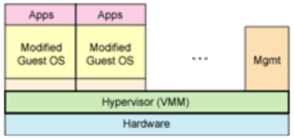
Summary



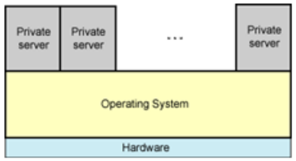
❑ Hardware emulation



❑ Full Virtualization

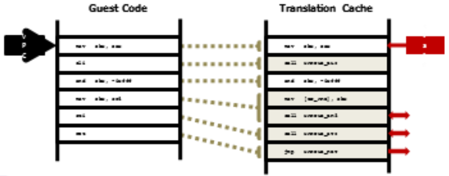


❑ Para-Virtualization

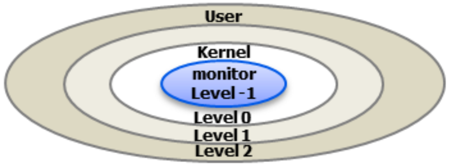


❑ O.S. Virtualization level

1

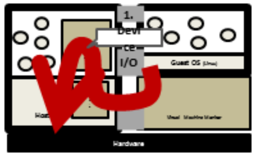


❑ Binary translation (patching)

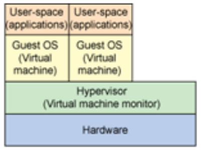


❑ New instructions

2

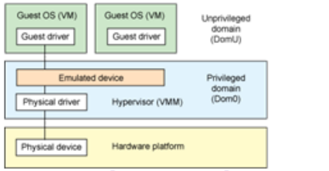


❑ Hosted or split

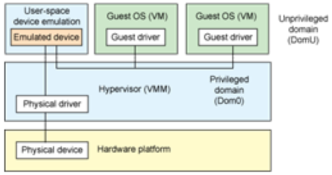

User-space (applications)
User-space (applications)
Guest OS (Virtual machine)
Guest OS (Virtual machine)
Hypervisor (Virtual machine monitor)
Hardware

❑ Hypervisor

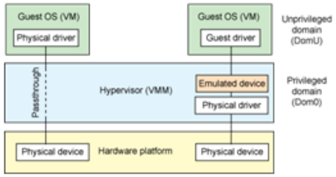
3



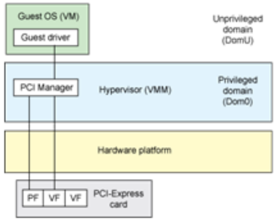
❑ Hypervisor device emulation



❑ User-space device emulation



❑ Gateway to device



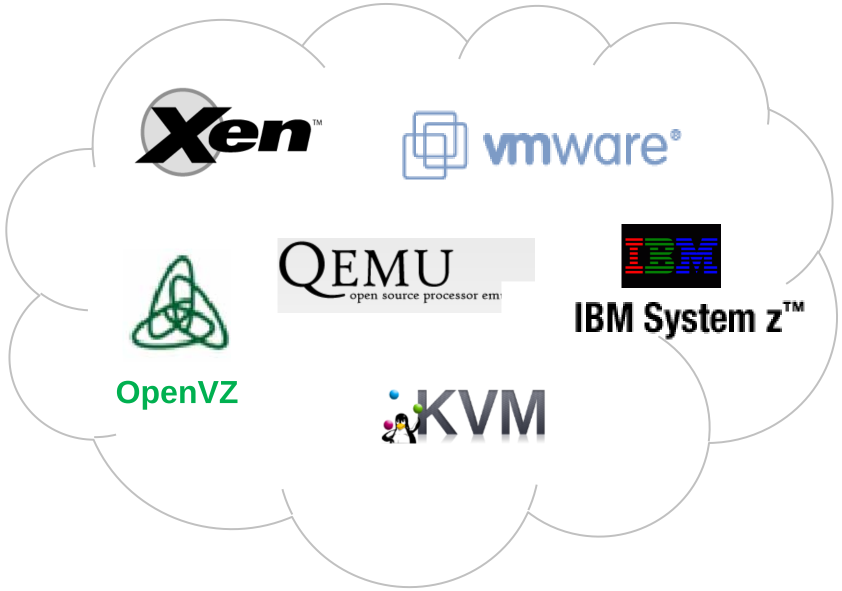
❑ SR-IOV and MR-IOV

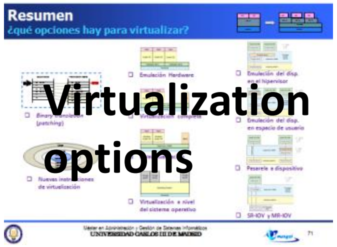
4

Contents

- What virtualization is.
- Where virtualization could be used.
- Why virtualization is used.
- Which kind of virtualization options could be used.
- **Main aspects to remember when deploying a virtualization system.**

Platform selection





Resumen

¿qué opciones hay para virtualizar?

- □ Binary translation (patching)
- □ Nuevas instru ciones de virtualización
- □ Emulación Hardware
- □ Virtualización completa
- □ Virtualización a nivel del sistema operativo
- □ Emulación del disp. en el hipervisor
- □ Emulación del disp. en espacio de usuario
- □ Pasarela a dispositivo
- □ SR-IOV y MR-IOV

Máster en Administración y Gestión de Sistemas Informáticos
UNIVERSIDAD CARLOS III DE MADRID
71

Virtualization options

To understand

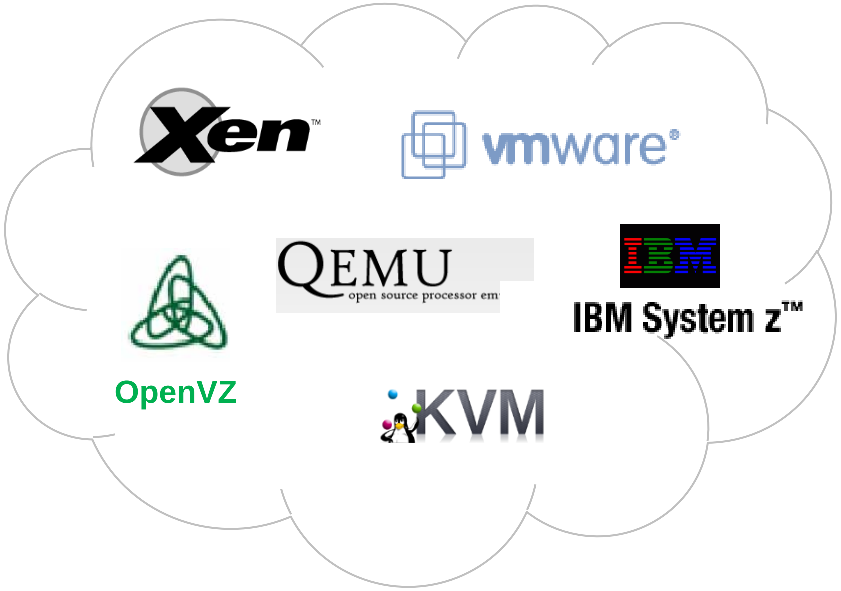
Criteria:

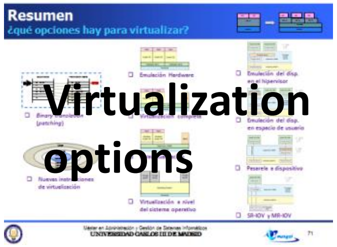
- Features
- Performance
- Information

To apply

Requirements

Platform selection





Resumen
¿qué opciones hay para virtualizar?

- □ Binary translation (patching)
- □ Nuevas instrucciones de virtualización
- □ Emulación Hardware
- □ Virtualización completa
- □ Virtualización a nivel del sistema operativo
- □ Emulación del disp. en el hipervisor
- □ Emulación del disp. en espacio de usuario
- □ Pasarela a dispositivo
- □ SR-IOV y MR-IOV

Máster en Administración y Gestión de Sistemas Informáticos
UNIVERSIDAD CARLOS III DE MADRID
71

Virtualization options

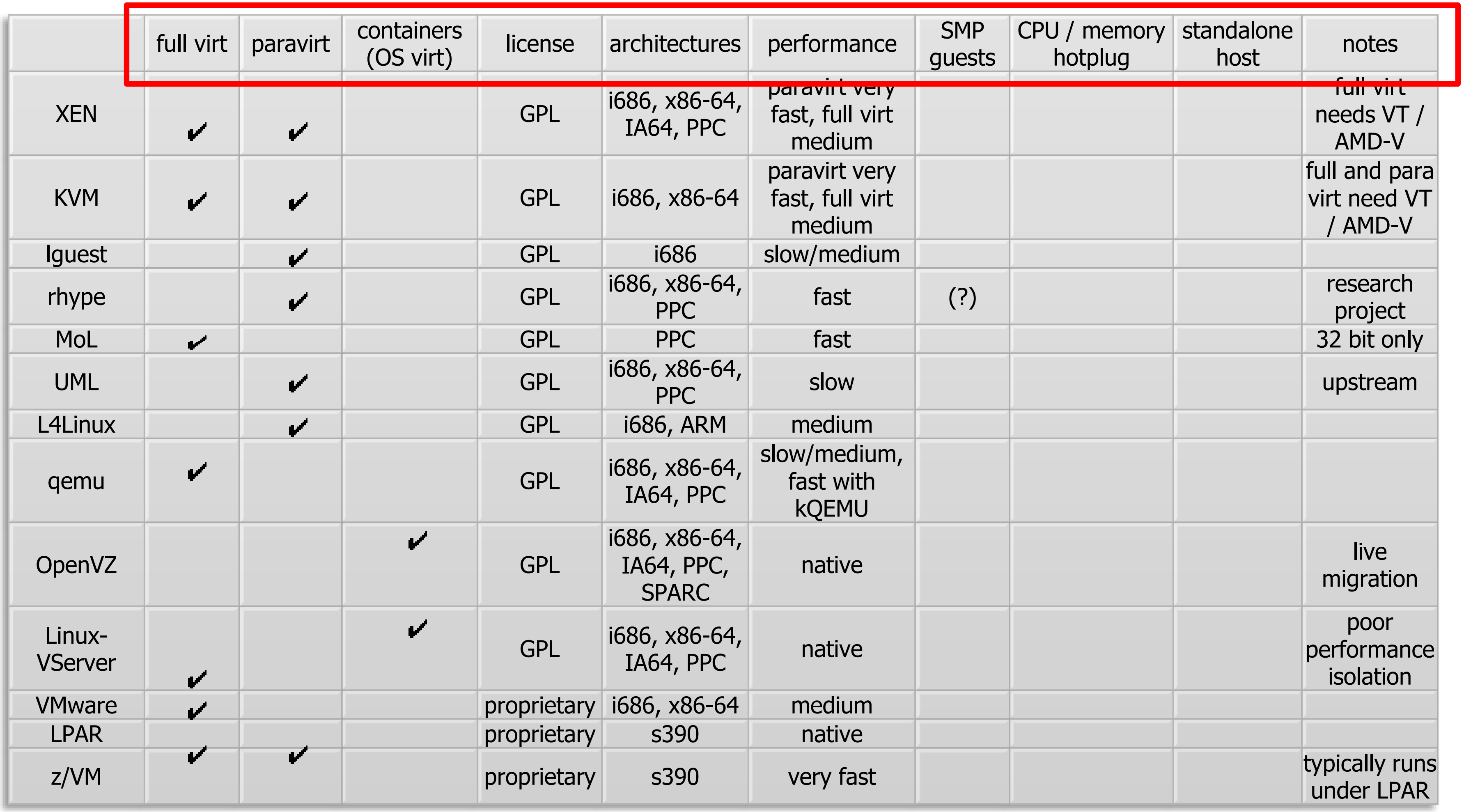
To understand

Criteria:
- Features
- Performance
- Information

To apply

Requirements

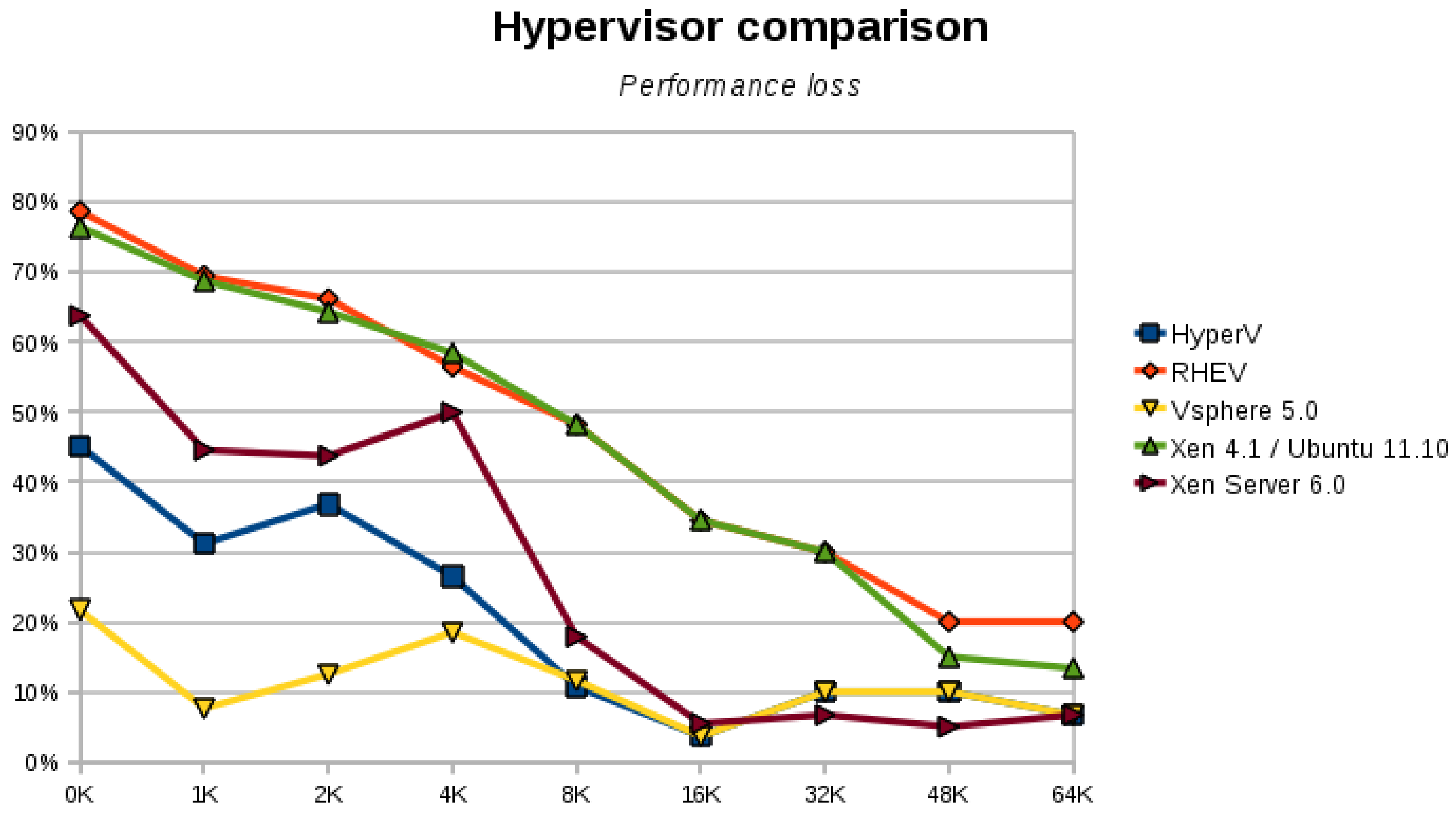
Features



	full virt	paravirt	containers (OS virt)	license	architectures	performance	SMP guests	CPU / memory hotplug	standalone host	notes
XEN	✔	✔		GPL	i686, x86-64, IA64, PPC	paravirt very fast, full virt medium				full virt needs VT / AMD-V
KVM	✔	✔		GPL	i686, x86-64	paravirt very fast, full virt medium				full and para virt need VT / AMD-V
lguest		✔		GPL	i686	slow/medium				
rhype		✔		GPL	i686, x86-64, PPC	fast	(?)			research project
MoL	✔			GPL	PPC	fast				32 bit only
UML		✔		GPL	i686, x86-64, PPC	slow				upstream
L4Linux		✔		GPL	i686, ARM	medium				
qemu	✔			GPL	i686, x86-64, IA64, PPC	slow/medium, fast with kQEMU				
OpenVZ			✔	GPL	i686, x86-64, IA64, PPC, SPARC	native				live migration
Linux-VServer	✔		✔	GPL	i686, x86-64, IA64, PPC	native				poor performance isolation
VMware	✔			proprietary	i686, x86-64	medium				
LPAR				proprietary	s390	native				
z/VM	✔	✔		proprietary	s390	very fast				typically runs under LPAR

http://virt.kernelnewbies.org/TechComparison

Performance



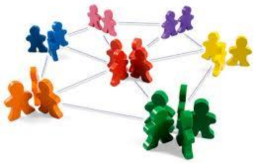
https://alohalb.wordpress.com/2012/04/24/hypervisors-virtual-network-performance-comparison-from-a-virtualized-load-balancer-point-of-view/

Information

- Support

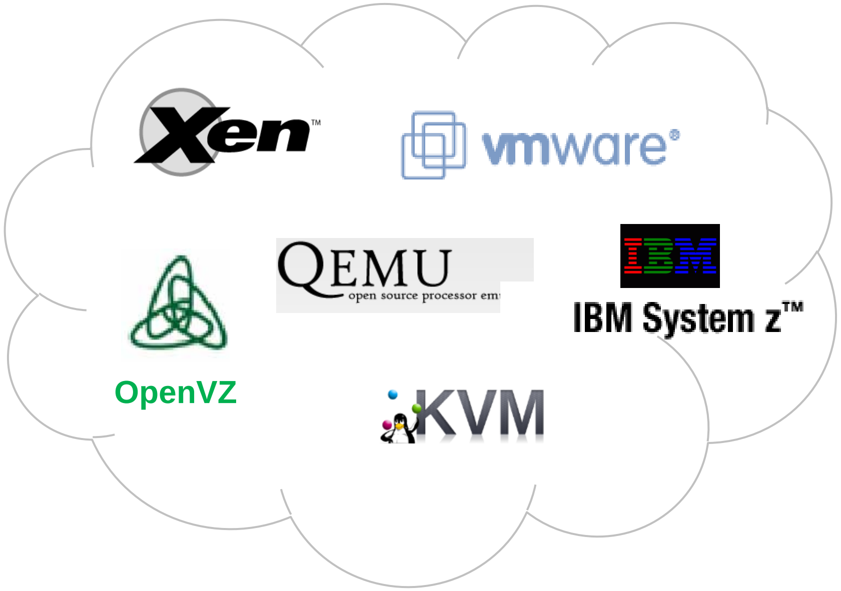
- Documentation

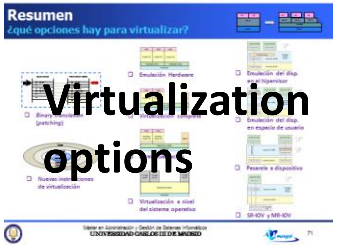
- Forums



- Recent deployments

Platform selection




Resumen
¿qué opciones hay para virtualizar?

□ Binary translation (patching)

□ Nuevas instrucciones de virtualización

□ Emulación Hardware

□ Virtualización completa

□ Virtualización a nivel del sistema operativo

□ Emulación del disp. en el hipervisor

□ Emulación del disp. en espacio de usuario

□ Pasarela a dispositivo

□ SR-IOV y MR-IOV

Máster en Administración y Gestión de Sistemas Informáticos
UNIVERSIDAD CARLOS III DE MADRID

71

Virtualization options

To understand

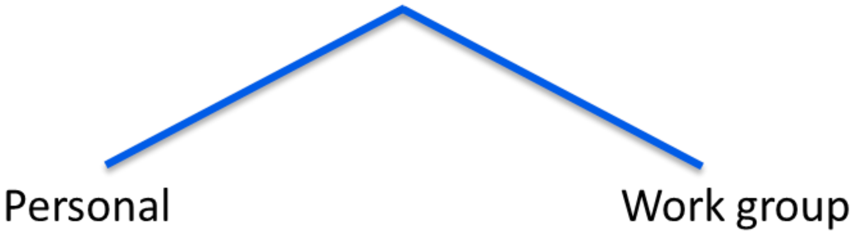
Criteria:

- Features
- Performance
- Information

To apply

Requirements

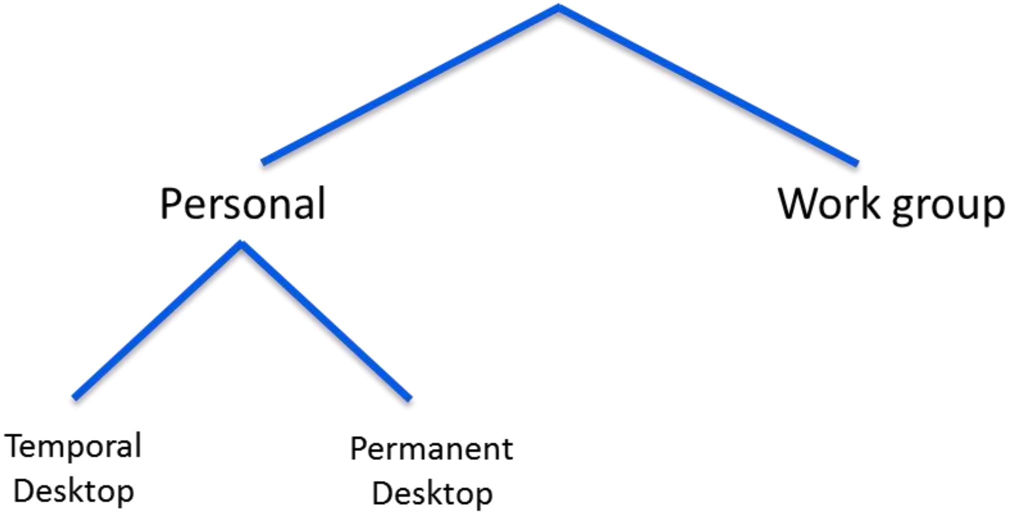
Requirement identification



Personal

Work group

Requirement identification

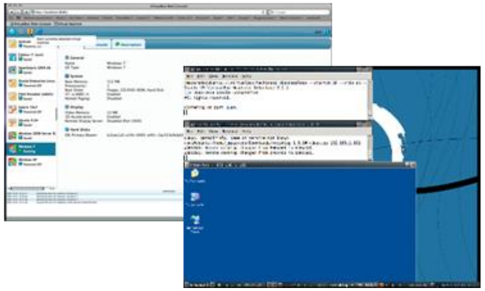


Personal

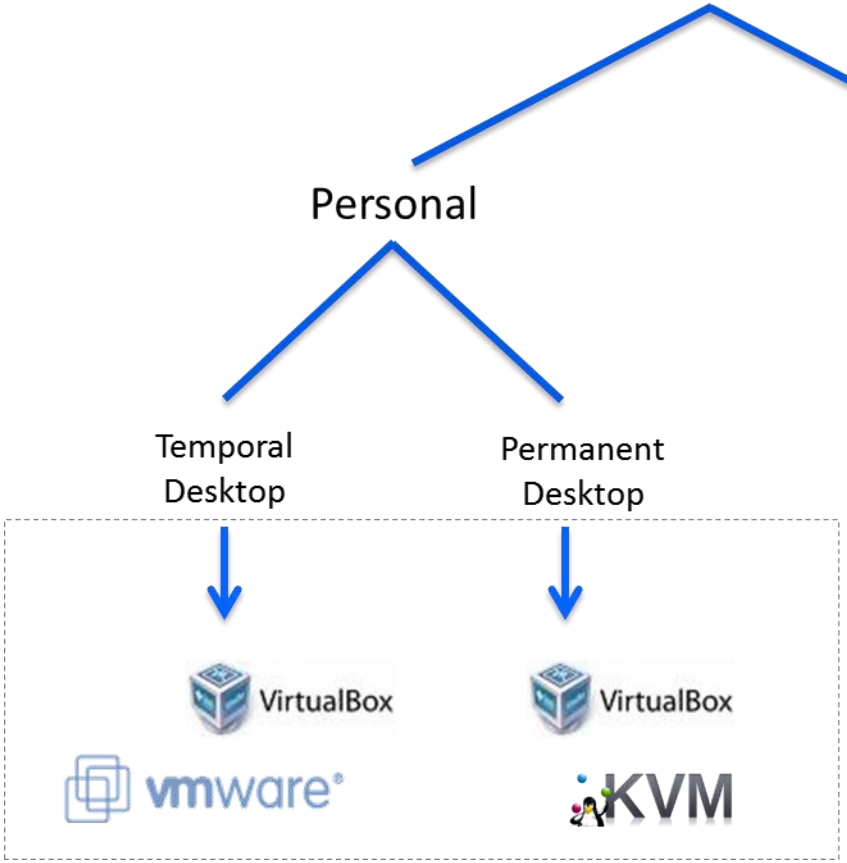
Work group

Temporal Desktop

Permanent Desktop



Requirement identification



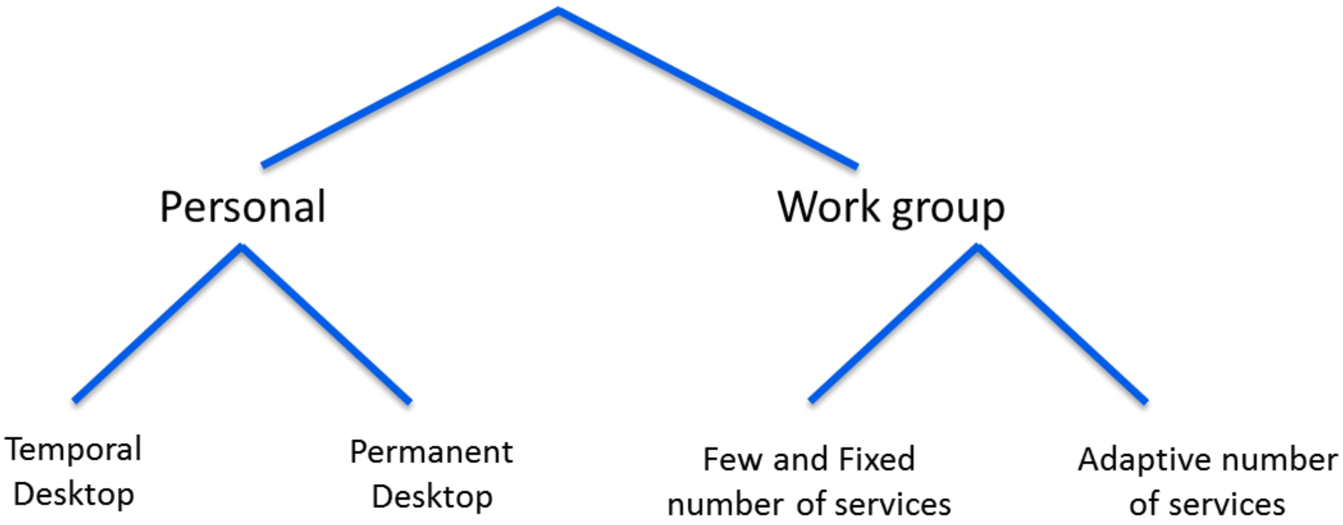
Personal

Work group

Temporal Desktop

Permanent Desktop

Requirement identification



Personal

Work group

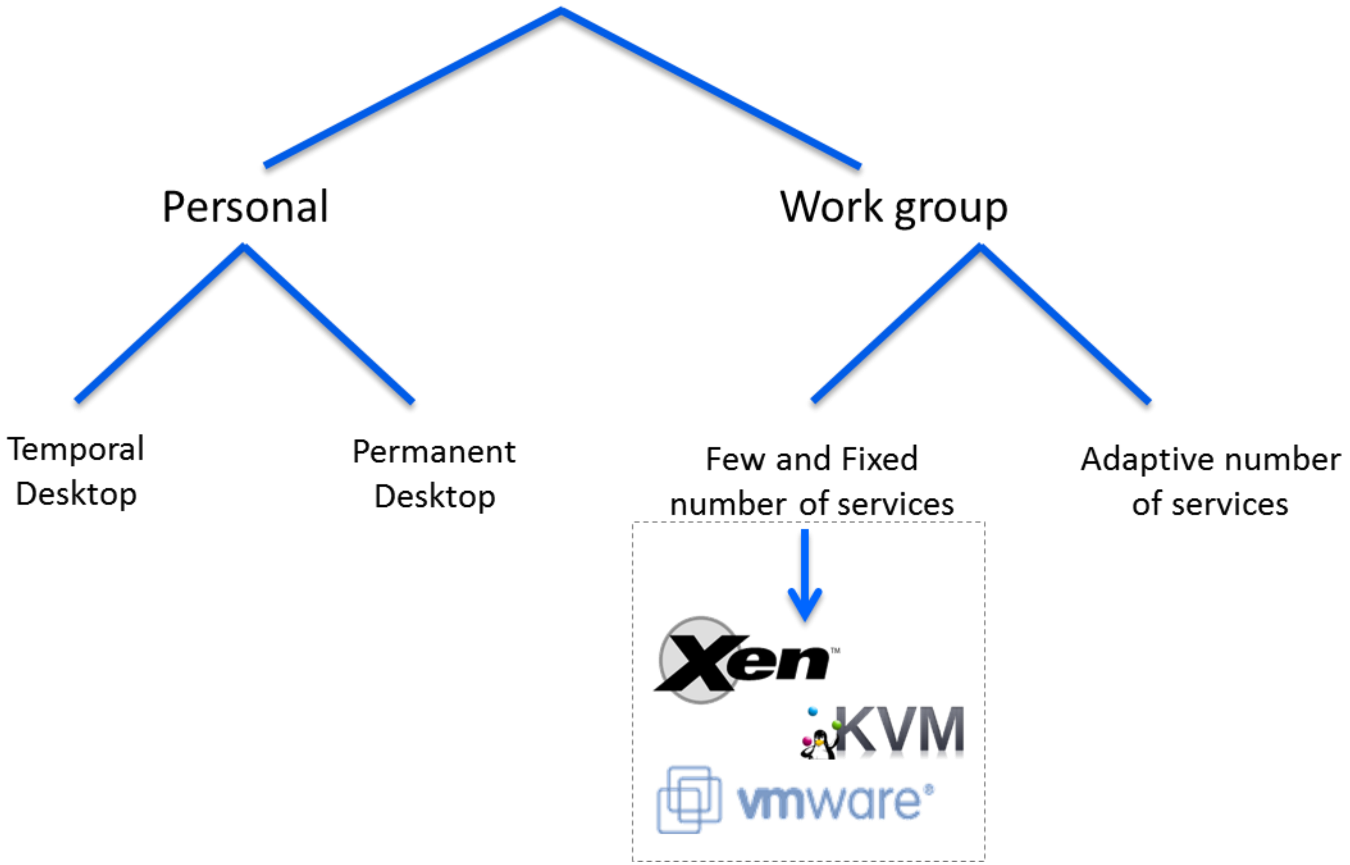
Temporal Desktop

Permanent Desktop

Few and Fixed number of services

Adaptive number of services

Requirement identification



Personal

Work group

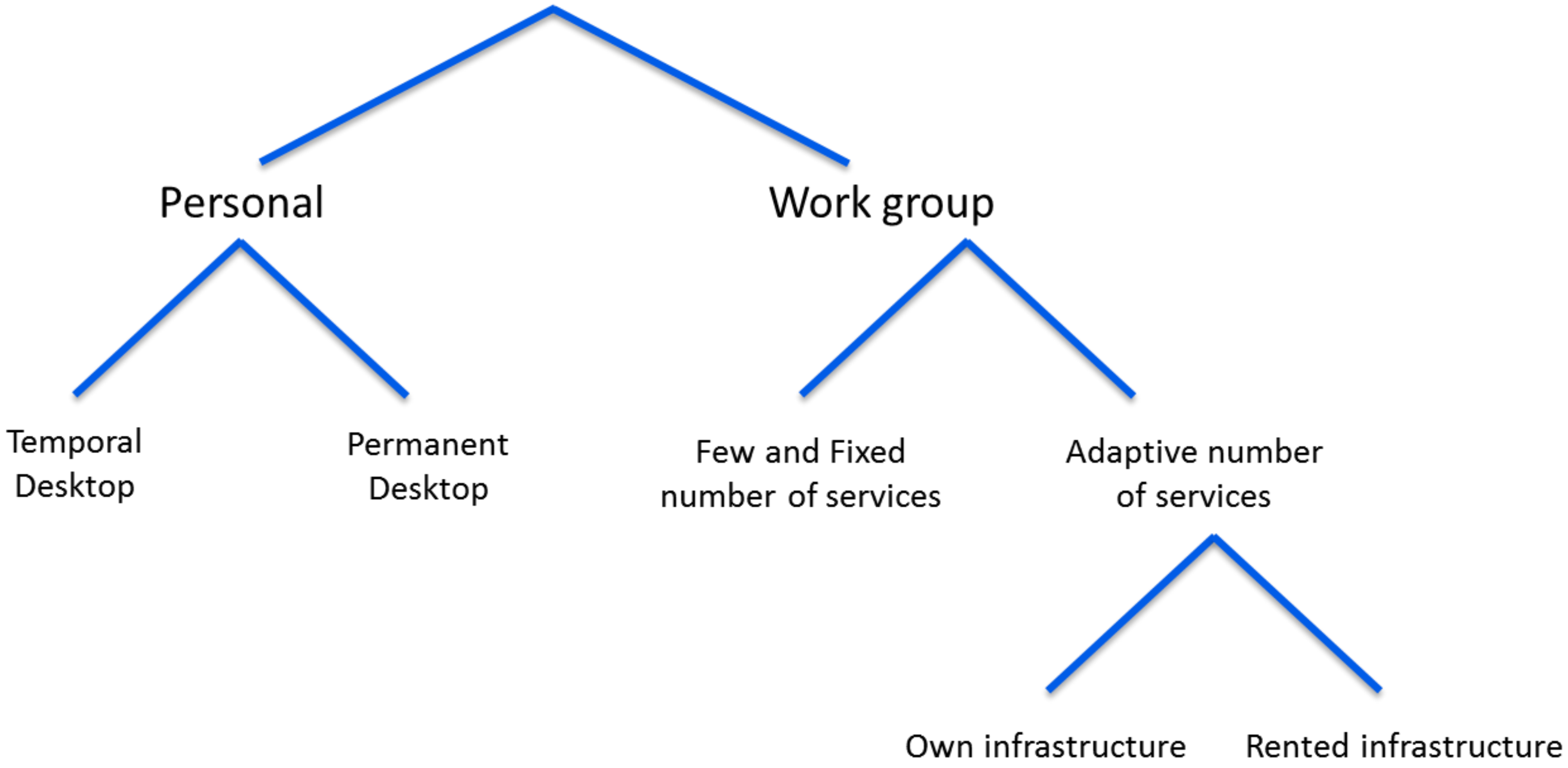
Temporal Desktop

Permanent Desktop

Few and Fixed number of services

Adaptive number of services

Requirement identification



Personal

Work group

Temporal Desktop

Permanent Desktop

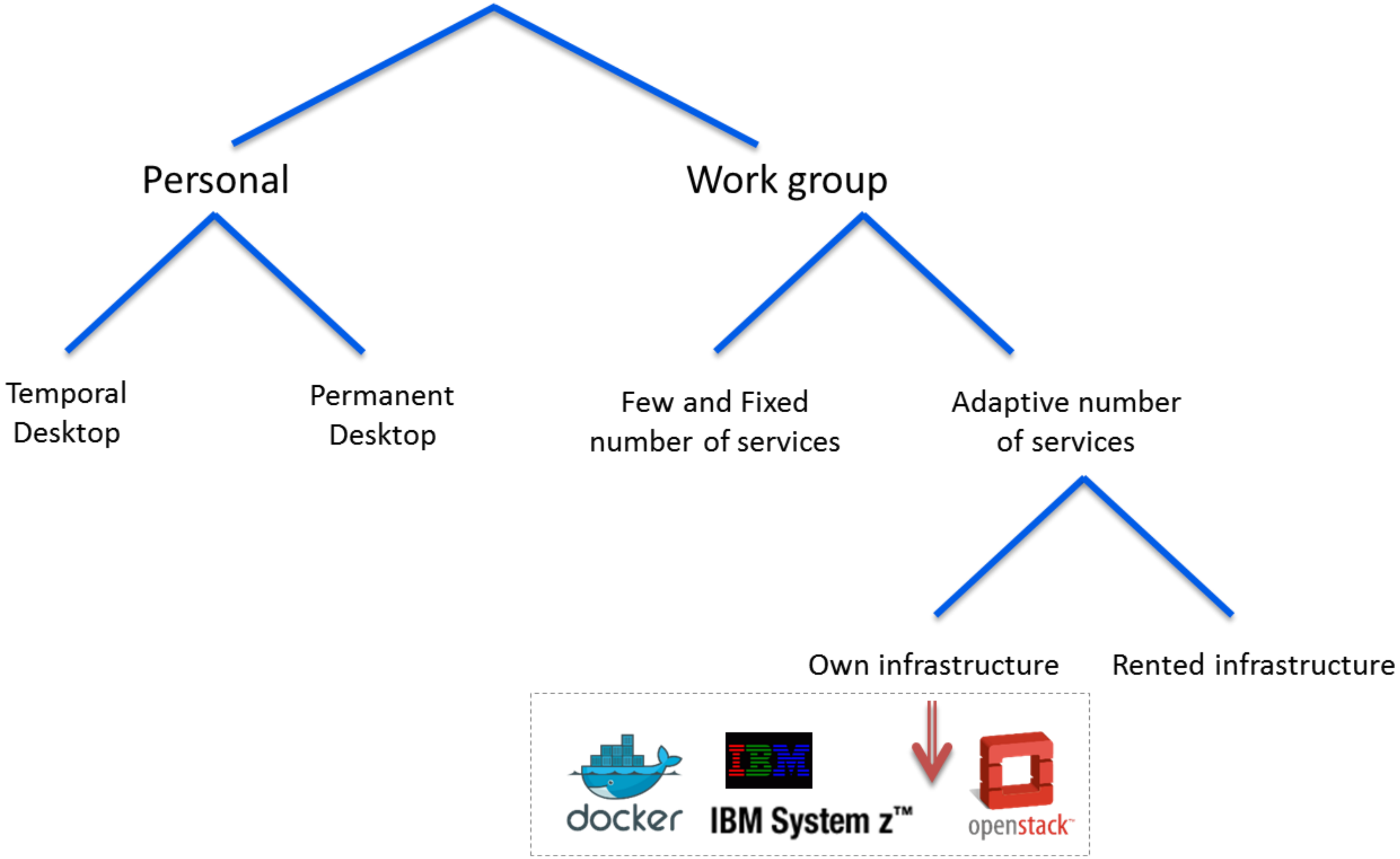
Few and Fixed number of services

Adaptive number of services

Own infrastructure

Rented infrastructure

Requirement identification



Personal

Work group

Temporal Desktop

Permanent Desktop

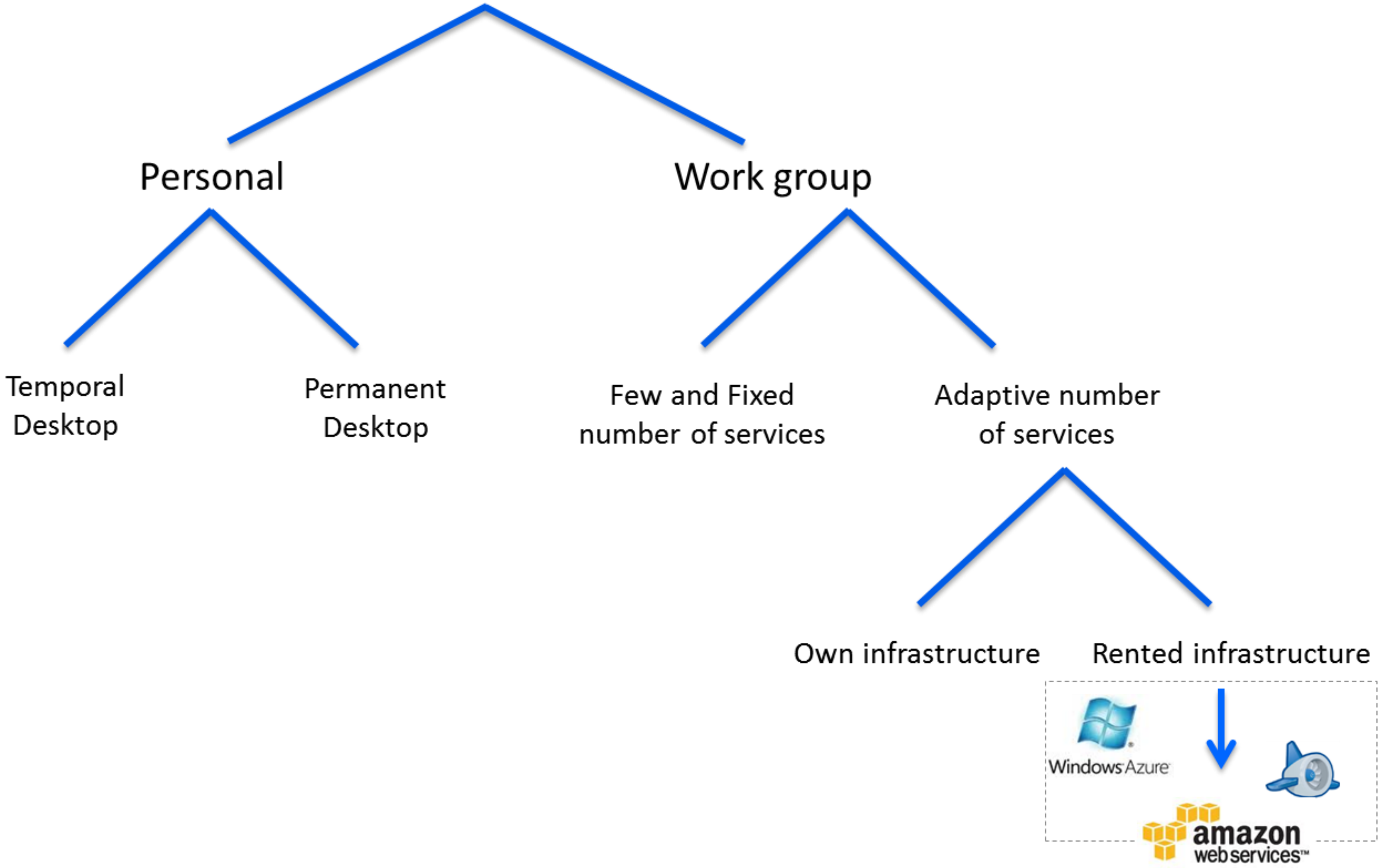
Few and Fixed number of services

Adaptive number of services

Own infrastructure

Rented infrastructure

Requirement identification



Personal

Work group

Temporal Desktop

Permanent Desktop

Few and Fixed number of services

Adaptive number of services

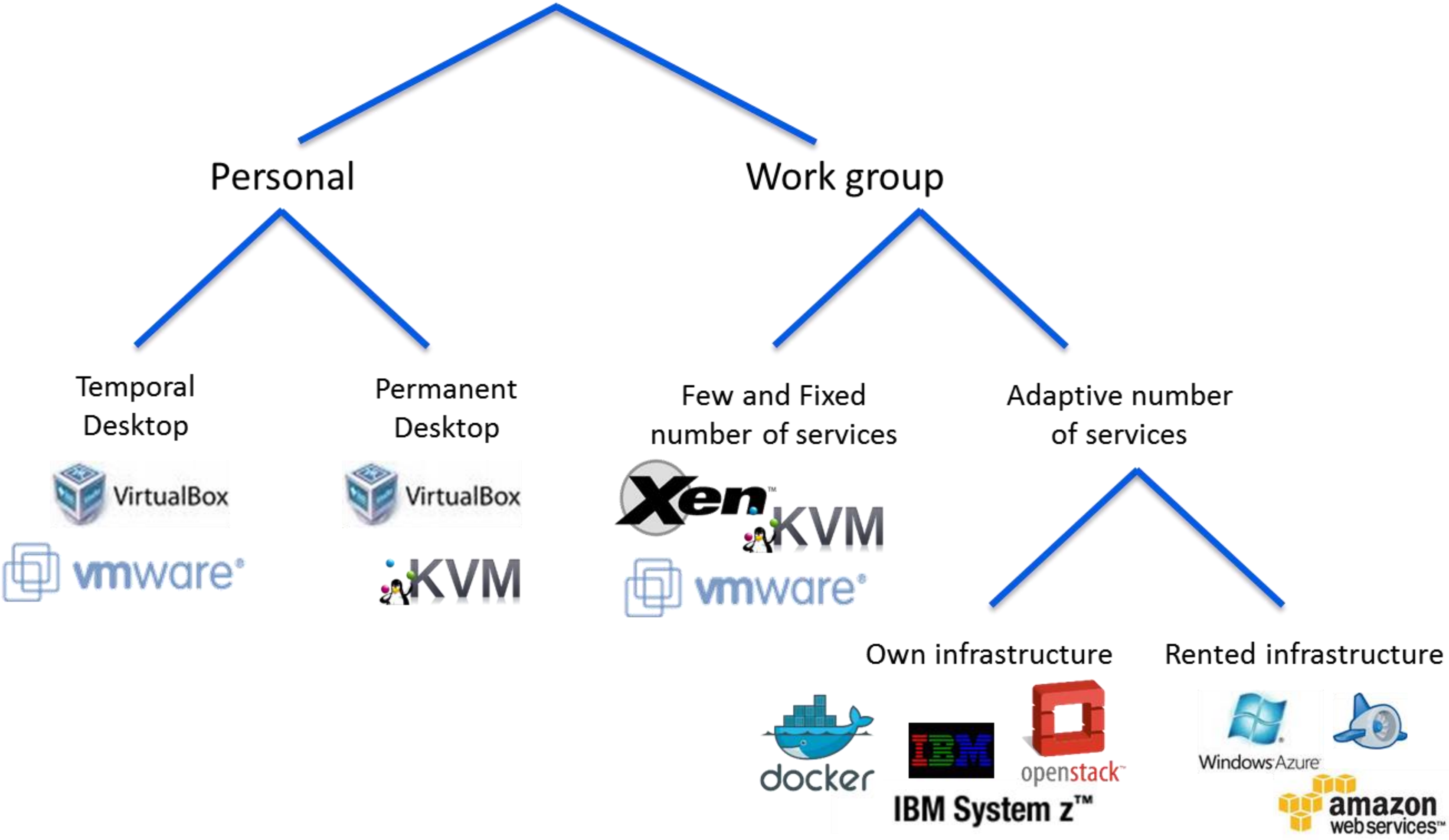
Own infrastructure

Rented infrastructure

https://groups.google.com/group/cloud-computing/web/list-of-cloud-platforms-providers-and-enablers

Requirement identification

summary



Personal

Work group

Temporal Desktop

Permanent Desktop

Few and Fixed number of services

Adaptive number of services

Own infrastructure

Rented infrastructure

Universidad
Carlos III de Madrid
www.uc3m.es

ARCOS

FUNDAMENTALS OF VIRTUALIZATION ON BIG DATA SYSTEMS

Lesson 3

Virtualization