# CPET 581 Cloud Computing: Technologies and Enterprise IT Strategies

## Lecture 4

## Virtualization of Clusters and Data Centers

Text Book: Distributed and Cloud Computing, by K. Hwang, G C. Fox, and J.J. Dongarra, published Elsevier/Morgan Kaufmann, 2012.

Spring 2015

A Specialty Course for Purdue University's M.S. in Technology Graduate Program: IT/Advanced Computer App Track

Paul I-Hai Lin, Professor

Dept. of Computer, Electrical and Information Technology

Purdue University Fort Wayne Campus



## References

[1] Chapter 3. Virtual Machines and Virtualization of Clusters for Data Centers, of the Text Book: Distributed and Cloud Computing, by K. Hwang, G C. Fox, and J.J. Dongarra, published Elsevier/Morgan Kaufmann, 2012.

[2] "Virtualization Overview," VMware White paper, 2006

[3] Understanding Full Virtualization, Paravirtualization and Hardware Assist, Nov. 10, 2007, http://www.vmware.com/resources/techresources/1008

[4] "The Architecture of Virtual Machines," J. E. Smith and R. Nair, IEEE Computer, May 2005, pp. 32-38.

[5] "Virtual Machine Monitors: Current Technology and Future Trends," IEEE Computer, May 2005, pp. 39-47.

[6] "Intel Virtualization Technology," R. Uhlig, et. al., IEEE Computer, May 2005, pp. 48-56.

[7] "Rethinking the Design of Virtual Machine Monitor," A. Whitaker, R. S. Cox, M. Shaw, and S. D. Gribble, IEEE Computer, May 2005, pp. 57-62.

[8] "Virtual Distributed Environments in a Shared Infrastructure," P. Ruth, X. Jiang, D. Xu, and S. Goasguen, IEEE Computer, May 2005, pp. 63-69.

## Topics of Discussion

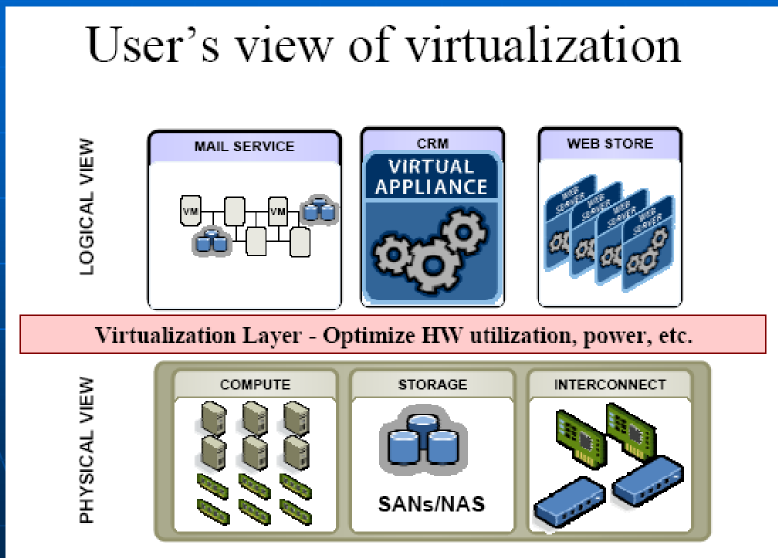
- Implementation Levels of Virtualization
- Virtualization Structures/Tools and Mechanisms
- Virtualization of CPU, Memory, and I/O Devices
- Virtual Clusters and Resource Management
  - Cloud OS for Building Private Cloud
- Virtualization for Datacenter Automation



## Virtualization

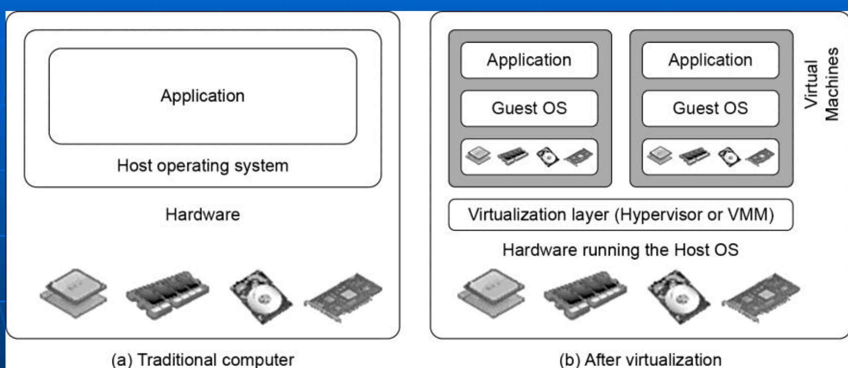
- Virtualization
  - **Emulation of a computer hardware environment in software**.
  - A **virtualization software layer** known as Hypervisor or Virtual Machine Monitor (VMM) is needed.
  - It presents all the necessary hardware components such as CPU, memory, storage, etc., to an hosting environment to make it appear that its is running on a real hardware device.
  - **Full** Virtualization
  - **Para** Virtualization
  - **Hardware Assisted** Virtualization
- Virtual Machine
- Levels of Virtualization
- Virtualization Technologies
- Server Virtualization
  - Window Servers
  - Linux Servers
  - Desktop virtualization

## User's View of Virtualization (Source: VMWare 2008)






## Figure 3.1 The Architecture of a Computer System before and after Virtualization



## Primitive Operations in Virtual Machines

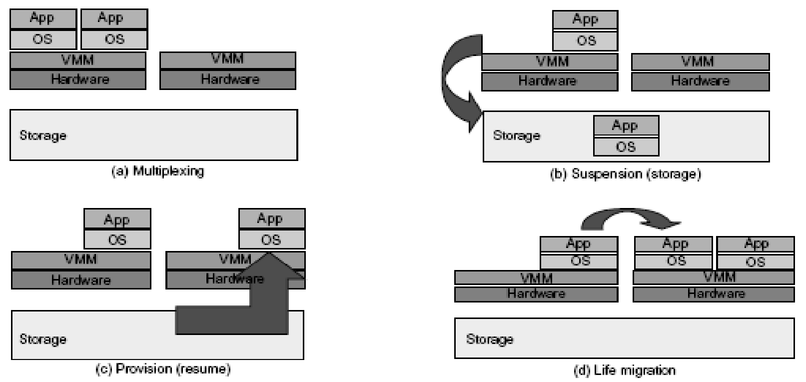



**FIGURE 1.13**

VM multiplexing, suspension, provision, and migration in a distributed computing environment.

*(Courtesy of M. Rosenblum, Keynote address, ACM ASPLOS 2006 [41])*



## Virtual Machine, VMM or Virtualization Layer

- Virtual Machine
  - A representation of a real machine using software that provides an operating environment which can run or host a guest operating system.
- Guest Operating System
  - An operating system running in a virtual machine environment that would otherwise run directly on a separate physical system.
- Virtualization Layer or Virtual Machine Monitor
  - The Virtualization layer is the middleware between the underlying hardware and virtual machine represented in the system, also known as virtual machine monitor (VMM)

## Figure 1.12 Three VM Architectures

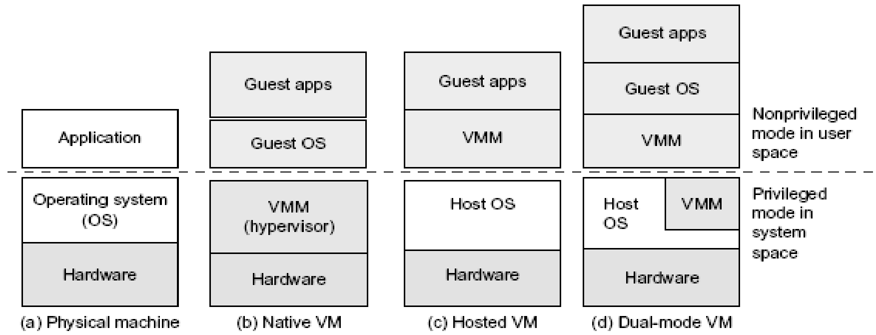
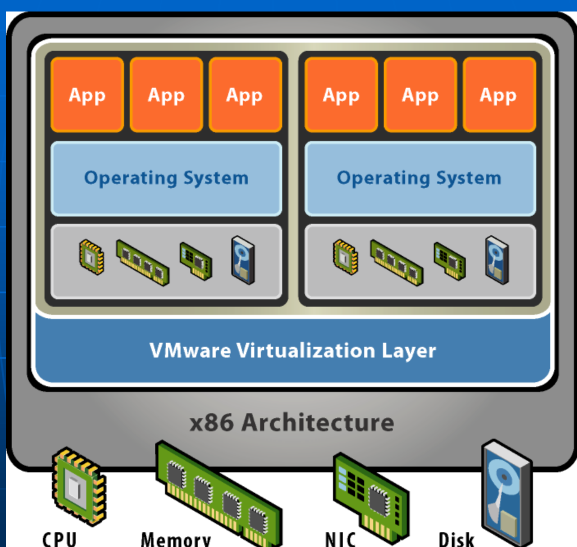



**FIGURE 1.12**

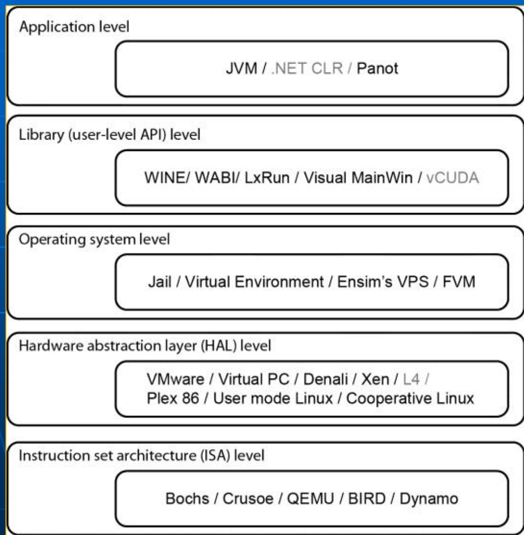
Three VM architectures in (b), (c), and (d), compared with the traditional physical machine shown in (a).



## X86 Virtualization Layer (source: VMWare [2])



## Figure 3.2 Five Levels of Virtualization ranging from Hardware to Applications






## Table 3.1 Relative Merits of Virtualization at Various Levels (More "X"'s means Higher Merits, with a Maximum of 5 X's)

**Table 3.1** Relative Merits of Virtualization at Various Levels

| Level of Implementation | Higher Performance | Application Flexibility | Implementation Complexity | Application Isolation |
|---|---|---|---|---|
| ISA | X | XXXXX | XXX | XXX |
| Hardware-level virtualization | XXXXX | XXX | XXXXX | XXXX |
| OS-level virtualization | XXXXX | XX | XXX | XX |
| Runtime library support | XXX | XX | XX | XX |
| User application level | XX | XX | XXXXX | XXXXX |

## Virtualization on Linux or Windows Platforms

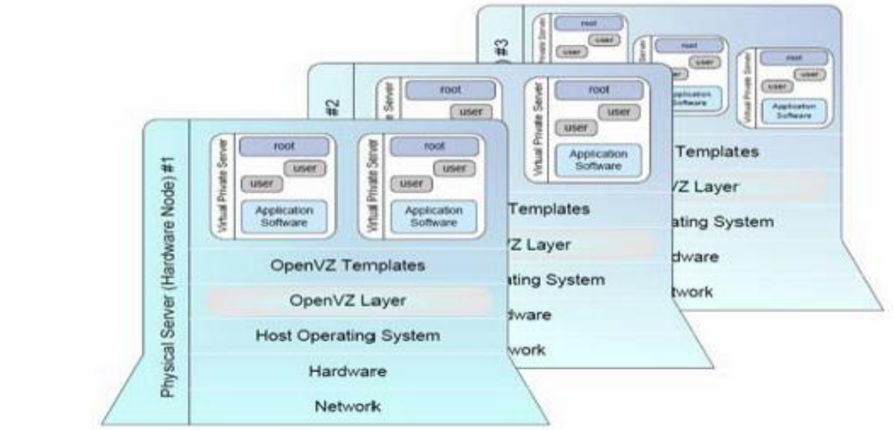



**Figure 6.4 OpenVZ inserts a virtualization layer called OpenVZ inside the host OS. This layer provides some OS images to create VMs quickly** (Courtesy of OpenVZ User's Guide, http://ftp.openvz.org/doc/OpenVZ-Users-Guide.pdf )



## VM Technology for Supporting Cloud Computing

- Challenges for VM technology to support Cloud Computing
  - 1. Ability to use a variable number of physical machines and VM instances
  - 2. Slow operation of instantiating new VMs
- Why OS-Level Virtualization (benefits)
  - Minimum startup/shutdown costs, Low resource requirement, and high scalability
  - Synchronizing state changes when necessary

# Full vs. Para Virtualization

- Full Virtualization
  - Guest OS and critical instructions are emulated through “Binary Code Translation”
  - VMware Workstation applies full virtualization
  - Slow down the performance
- Para Virtualization
  - Guest OS, and non-virtualizable instructions are replaced by hypercalls (kernel calls) that communicate directly with the Hypervisor or VMM
  - Guest OS has to be modified
  - Disadvantages:
    - Reduced Compatibility, Portability because of the modified OS
    - Higher cost of maintenance due to deep OS modifications
  - Supported by Xen, Denali and VMWare ESX



# Full Virtualization

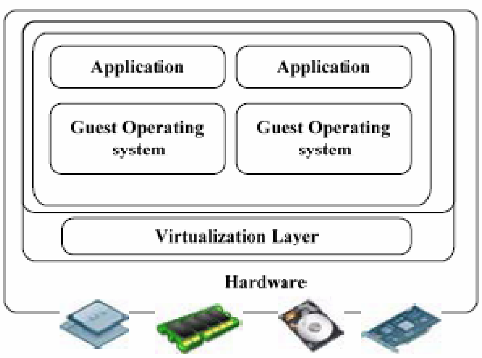
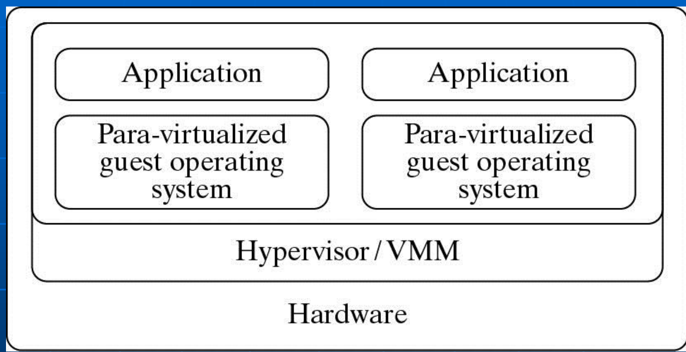



**Figure 6.9** The concept of full virtualization using a hypervisor or a VMM directly sitting on top of the bare hardware devices. Note that no host OS is used here as in Figure 6.11.

# Para Virtualization






# Table 3.2 Comparison of Four VMM and Hypervisor Software Packages

| VMM Provider | Host CPU | Guest CPU | Host OS | Guest OS | VM Architecture |
|---|---|---|---|---|---|
| VMware Work-station | X86, x86-64 | X86, x86-64 | Windows, Linux | Windows, Linux, Solaris, FreeBSD, Netware, OS/2, SCO, BeOS, Darwin | Full Virtualization |
| VMware ESX Server | X86, x86-64 | X86, x86-64 | No host OS | The same as VMware workstation | Para-Virtualization |
| XEN | X86, x86-64, IA-64 | X86, x86-64, IA-64 | NetBSD, Linux, Solaris | FreeBSD, NetBSD, Linux, Solaris, windows XP and 2003 Server | Hypervisor |
| KVM | X86, x86-64, IA64, S390, PowerPC | X86, x86-64, IA64, S390, PowerPC | Linux | Linux, Windows, FreeBSD, Solaris | Para-Virtualization |

- Vmware, www.vmware.com
- http://xen.org, Xen Hypervisor, Xen Cloud platform, Xen Arm
- KVM (kernel-based Virtual Machine), http://www.linux-kvm.org/page/Main_Page

## Example 3-1. Virtualization on Linux or Windows Platforms

- Most reported OS-level virtualization systems are Linux-based.
- New hardware may need different Linux patched kernels to provide special support for extended functionality.

Table 3.3 Virtualization Support for Linux and Windows NT Platforms

| Virtualization Support and Source of Information | Brief Introduction on Functionality and Application Platforms |
|---|---|
| **Linux vServer** for Linux platforms (http://linux-vserver.org/) | Extends Linux kernels to implement a security mechanism to help build VMs by setting resource limits and file attributes and changing the root environment for VM isolation |
| **OpenVZ** for Linux platforms [65]; http://ftp.openvz.org/doc/OpenVZ-Users-Guide.pdf) | Supports virtualization by creating *virtual private servers (VPSes)*; the VPS has its own files, users, process tree, and virtual devices, which can be isolated from other VPSes, and checkpointing and live migration are supported |
| **FVM** (Feather-Weight Virtual Machines) for virtualizing the Windows NT platforms [78]) | Uses system call interfaces to create VMs at the NY kernel space; multiple VMs are supported by virtualized namespace and copy-on-write |



## Hypervisor

- Hypervisor (VMM)
  - A hardware virtualization techniques allowing multiple guest OSs to run on a host machine.
  - Provides hypercalls for the guest OSs and applications
  - Depending on the functionalities, a hypervisor can
    - Assume a micro-kernel architecture
    - Or assume a monolithic hypervisor architecture like VMware ESX for server virtualization
- Types of Hypervisor
  - Type 1 Hypervisor
    - Run on the bare metal hypervisor
  - Examples:
    - IBM CP/CMS hypervisor
    - Microsoft Hyper-V
  - Type 2 (Hosted Hypervisor)
    - Run over a host OS
    - The hypervisor is the second layer over the hardware
    - Examples: FreeBSD

## Hypervisor and XEN Architecture

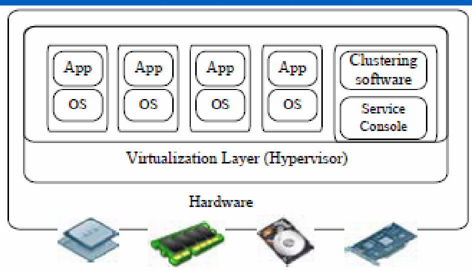



Figure 6.7 A hypervisor is the software layer for virtualization of the bare metal hardware. This layer can be implemented as a micro-kernel of the OS. It converts physical devices into virtual resources for user applications to run on the virtual machines deployed.



## The XEN Architecture

- An open source hypervisor program developed by Cambridge University, http://xen.org
- Xen hypervisor (a micro kernel)
- Commercial Xen Hypervisors: Citrix XenServer, Oracle VM
- Domain 0 – Privileged guest OS of Xen

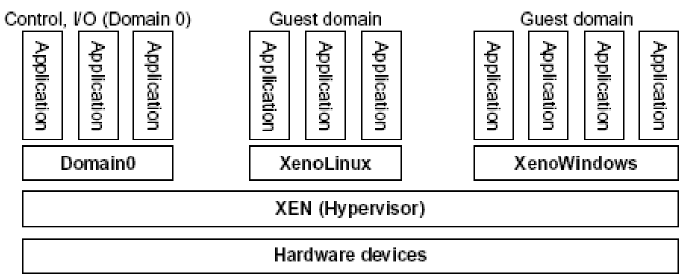



FIGURE 3.5

The Xen architecture's special domain 0 for control and I/O, and several guest domains for user applications.

## Para-Virtualization with Compiler Support

- The KVM builds offers Kernel-based VM on the Linux platform, based on Para-virtualization

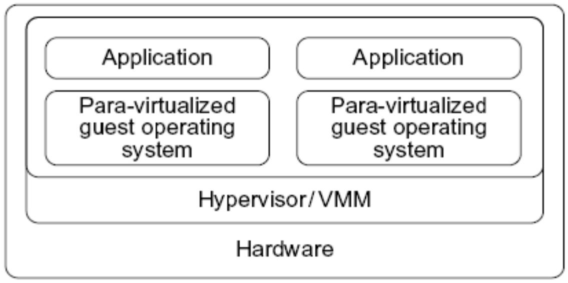
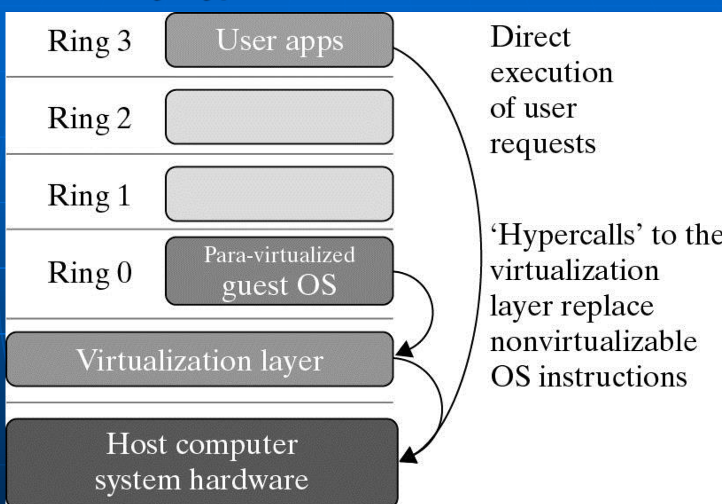



**FIGURE 3.7**

Para-virtualized VM architecture, which involves modifying the guest OS kernel to replace nonvirtualizable instructions with hypercalls for the hypervisor or the VMM to carry out the virtualization process (See Figure 3.8 for more details).

## Figure 3.8 The use of a Para-Virtualized OS assisted by an Intelligent Compiler to replace Non-virtualized OS Instructions by Hypercalls



## Example 3-3: VMWare ESX Server for Para-Virtualization

- SCSI (Small Computer System Interface); NIC (Network Interface Card); SMP (Symmetrical Multiprocessing)

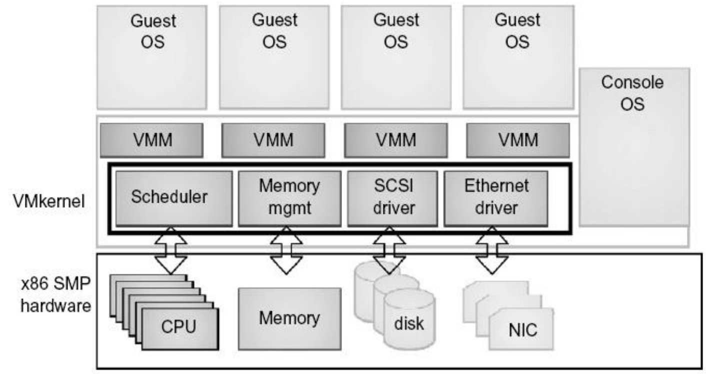



**FIGURE 3.9**

The VMware ESX server architecture using para-virtualization.

(*Courtesy of VMware [71]*)

## Virtualization of CPU, Memory, and I/O Devices

- Hardware Support for Virtualization
- CPU Virtualization
- Memory Virtualization
- I/O Virtualization
- Virtualization in Multi-Core Processors

## Example 3-4: Virtualization Support at Intel

- EPT (Extended Page Table); VT-x (Intel's Virtualization Technology)

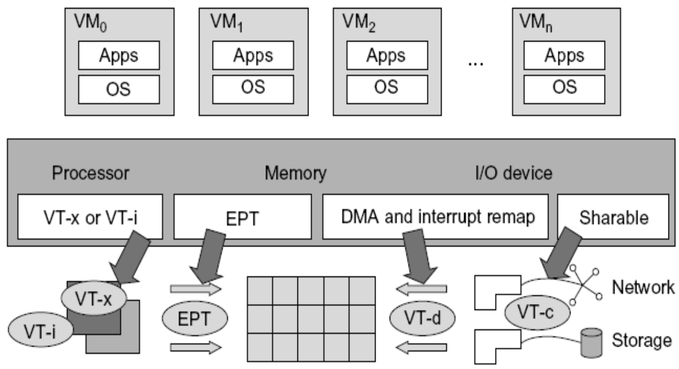



**FIGURE 3.10**

Intel hardware support for virtualization of processor, memory, and I/O devices.

*(Modified from [68], Courtesy of Lizhong Chen, USC)*

## Example 3.5: Intel Hardware-assisted CPU Virtualization

- VMCS (Virtual Machine Control Structure)

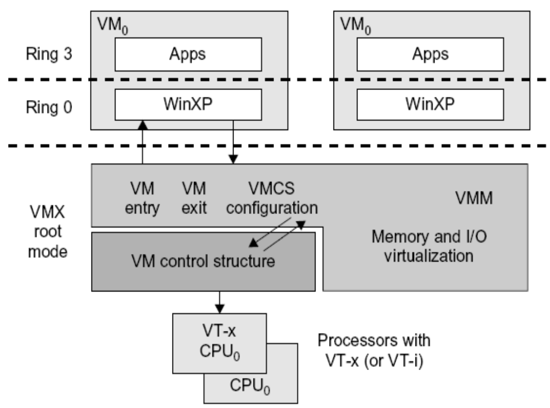



**FIGURE 3.11**

Intel Hardware-assisted CPU virtualization.

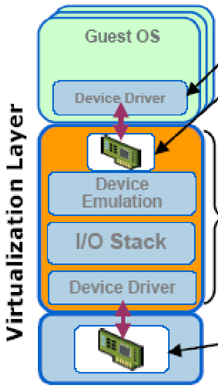
*(Modified from [68], Courtesy of Lizhong Chen, USC)*

8

## Intel Hardware-assisted CPU Virtualization

- COW – Copy On Write

### Current virtual I/O devices




- Guest device driver
- Virtual device
- Virtualization layer
  - emulates the virtual device
  - remaps guest and real I/O addresses
  - multiplexes and drives the physical device
  - I/O features, *e.g.*, COW disks,
- Real device
  - may be different from virtual device



## Virtual Clusters vs. Physical Clusters

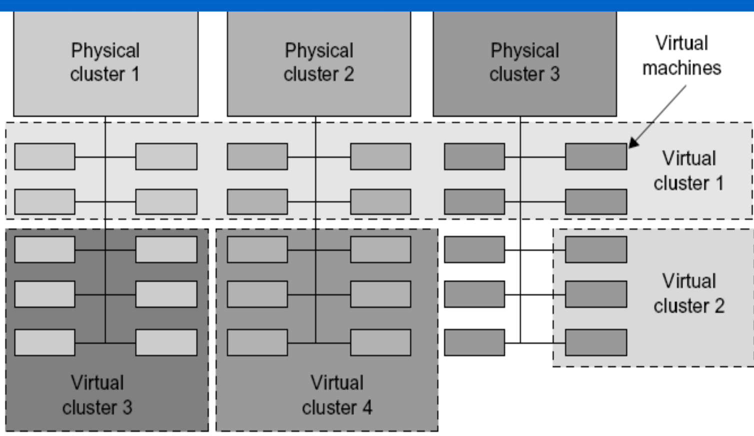



**FIGURE 3.18**

A cloud platform with 4 virtual clusters over 3 physical clusters shaded differently.

## A Virtual Clusters based on Application Partitioning

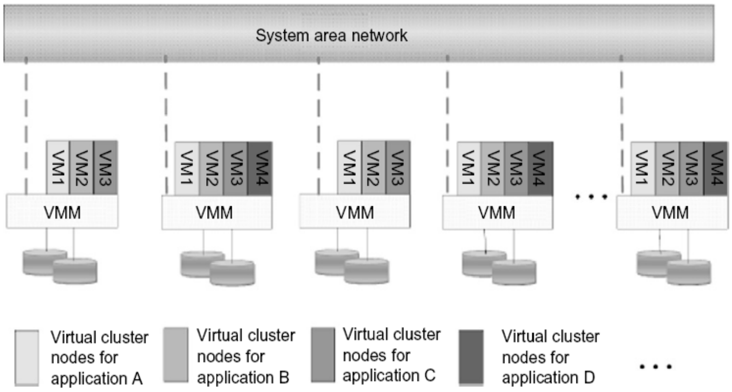



**FIGURE 3.19**

The concept of a virtual cluster based on application partitioning.

*(Courtesy of Kang, Chen, Tsinghua University 2008)*

## Virtual Clusters Projects

**Table 3.5** Experimental Results on Four Research Virtual Clusters

| Project Name | Design Objectives | Reported Results and References |
|---|---|---|
| Cluster-on-Demand at Duke Univ. | Dynamic resource allocation with a virtual cluster management system | Sharing of VMs by multiple virtual clusters using Sun GridEngine [12] |
| Cellular Disco at Stanford Univ. | To deploy a virtual cluster on a shared-memory multiprocessor | VMs deployed on multiple processors under a VMM called Cellular Disco [8] |
| VIOLIN at Purdue Univ. | Multiple VM clustering to prove the advantage of dynamic adaptation | Reduce execution time of applications running VIOLIN with adaptation [25,55] |
| GRAAL Project at INRIA in France | Performance of parallel algorithms in Xen-enabled virtual clusters | 75% of max. performance achieved with 30% resource slacks over VM clusters |

## Example 3.9: COD (Cluster-on-Demand) Project at Duke University

- DHCP (Dynamic Host Configuration Protocol); VCM (configuration Manager); NIS (Network Information Service)

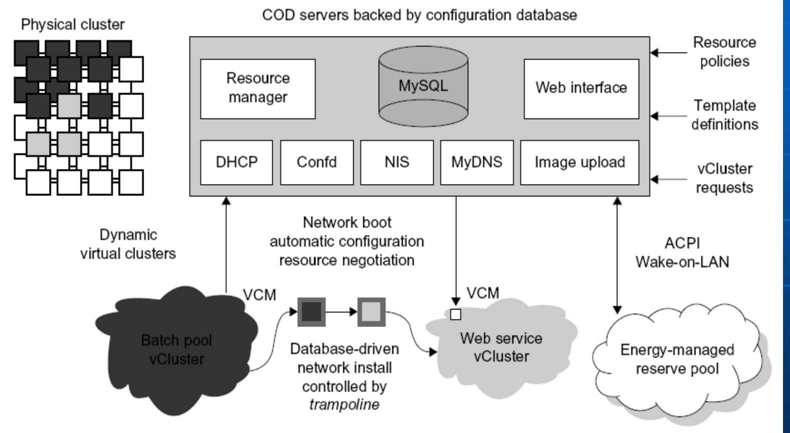



**FIGURE 3.23**

COD partitioning a physical cluster into multiple virtual clusters.

(*Courtesy of Jeff Chase, et al, HPDC-2003 [12]*)

## Example 3.10: VIOLIN Project at Purdue University

- Live VM migration to reconfigure a virtual cluster environment
- Five concurrent virtual environment, labeled VIOLIN 1-5, sharing two physical clusters.

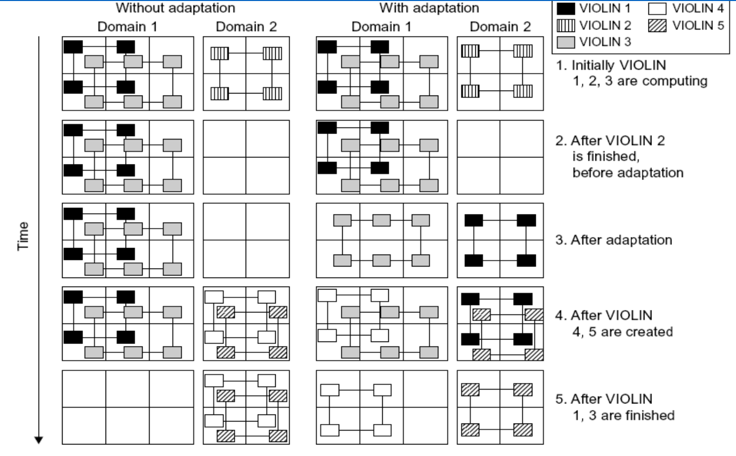



**FIGURE 3.25**

VIOLIN adaptation scenario of five virtual environments sharing two hosted clusters; Note that there are more idle squares (blank nodes) before and after the adaptation.

(*Courtesy of P. Ruth, et al. [24,51]*)

34

# 3.5 Virtualization for Data Center Automation

- Server Consolidation in Data Centers
- Virtual Storage Management
- Cloud OS for Virtualized Data Centers
- Trust Management in Virtualized Data Centers
  - VM-based Intrusion Detection



## Example 3.11: Parallax Providing Virtual Disks to Clients VMs from a Large Common Shared Physical Disk

- .

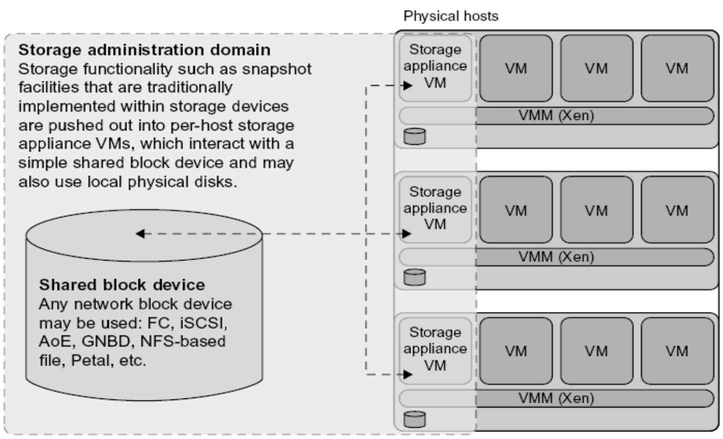



**FIGURE 3.26**

Parallax is a set of per-host storage appliances that share access to a common block device and presents virtual disks to client VMs.

*(Courtesy of D. Meyer, et al. [43])*

## Cloud OS for Virtualizing Data Centers (VI: Virtual Infrastructure, EC2: Elastic Compute Cloud)

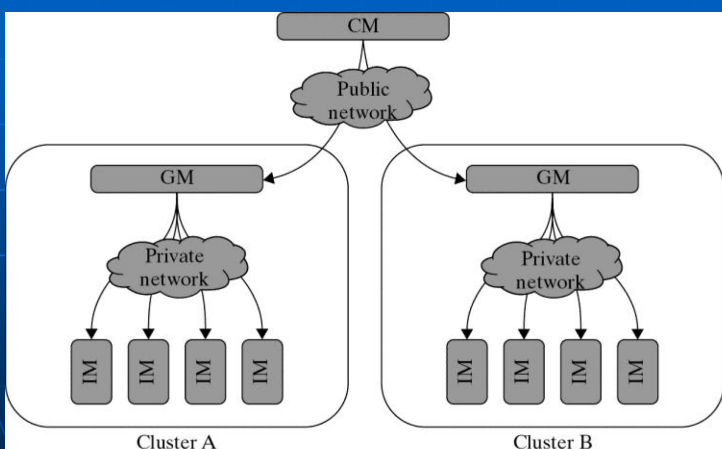
**Table 3.6** VI Managers and Operating Systems for Virtualizing Data Centers [9]

| Manager/ OS, Platforms, License | Resources Being Virtualized, Web Link | Client API, Language | Hypervisors Used | Public Cloud Interface | Special Features |
|---|---|---|---|---|---|
| **Nimbus** Linux, Apache v2 | VM creation, virtual cluster, www .nimbusproject.org/ | EC2 WS, WSRF, CLI | Xen, KVM | EC2 | Virtual networks |
| **Eucalyptus** Linux, BSD | Virtual networking (Example 3.12 and [41]), www .eucalyptus.com/ | EC2 WS, CLI | Xen, KVM | EC2 | Virtual networks |
| **OpenNebula** Linux, Apache v2 | Management of VM, host, virtual network, and scheduling tools, www.opennebula.org/ | XML-RPC, CLI, Java | Xen, KVM | EC2, Elastic Host | Virtual networks, dynamic provisioning |
| **vSphere 4** Linux, Windows, proprietary | Virtualizing OS for data centers (Example 3.13), www .vmware.com/ products/vsphere/ [66] | CLI, GUI, Portal, WS | VMware ESX, ESXi | VMware vCloud partners | Data protection, vStorage, VMFS, DRM, HA |

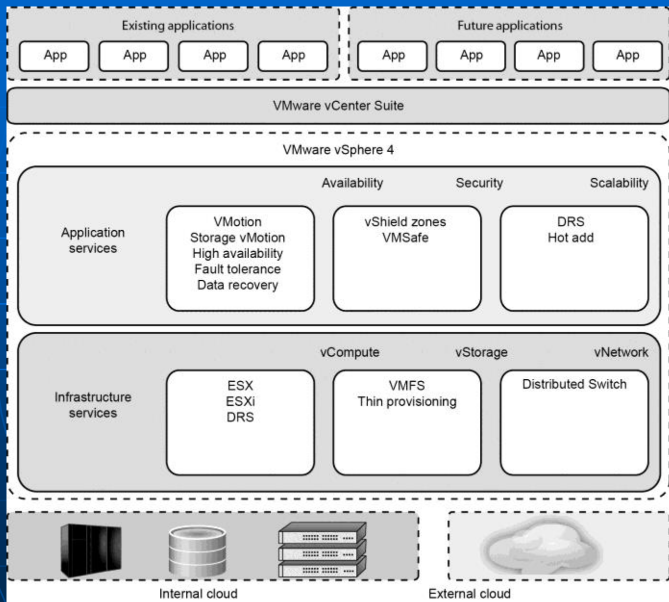


## Example 3.12: Eucalptus, An Open-SourceOS for Setting Up and Managing Private Clouds (IaaS)

- Three Resource Managers: CM (Cloud Manager), GM (Group Manager), and IM (Instance Manager)
- Works like AWS APIs



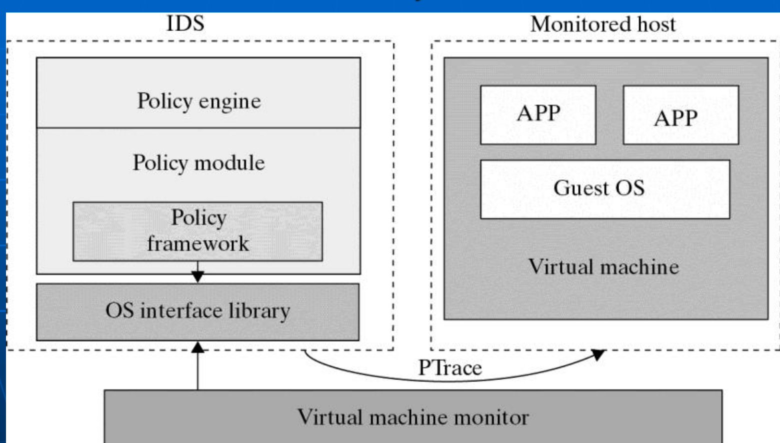
## Example 3.13: VMware vSphere 4 – A Commercial Cloud OS




39

## Trust Managmenet: VM-based Intrusion Detection

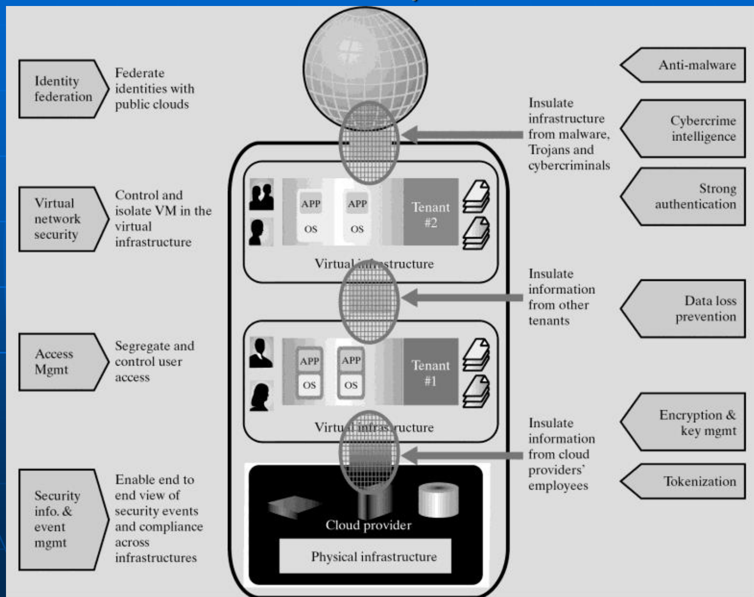
- IDS – Intrusion Detection System




Prof. Paul Lin

40

## Example 3.14: Figure 3.30 Techniques for establishing trusted zones for virtual cluster insulation and VM isolation (EMC and VMware)






# Conclusion and Summary