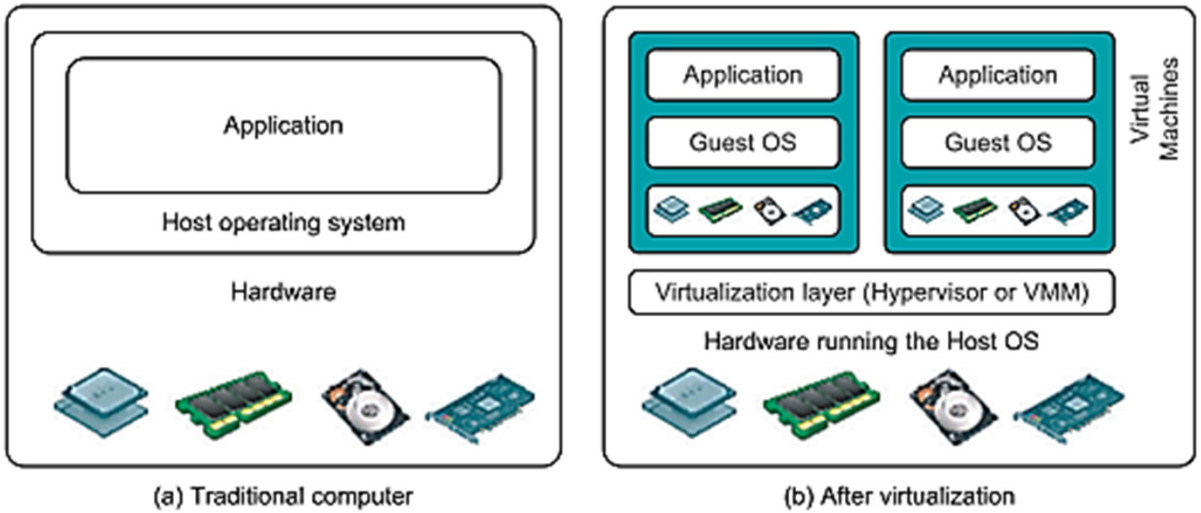
§2. VIRTUAL MACHINES AND DATA CENTERS

Nguyen Hong Son, PhD

1. VIRTUALIZATION IMPLEMENTATION

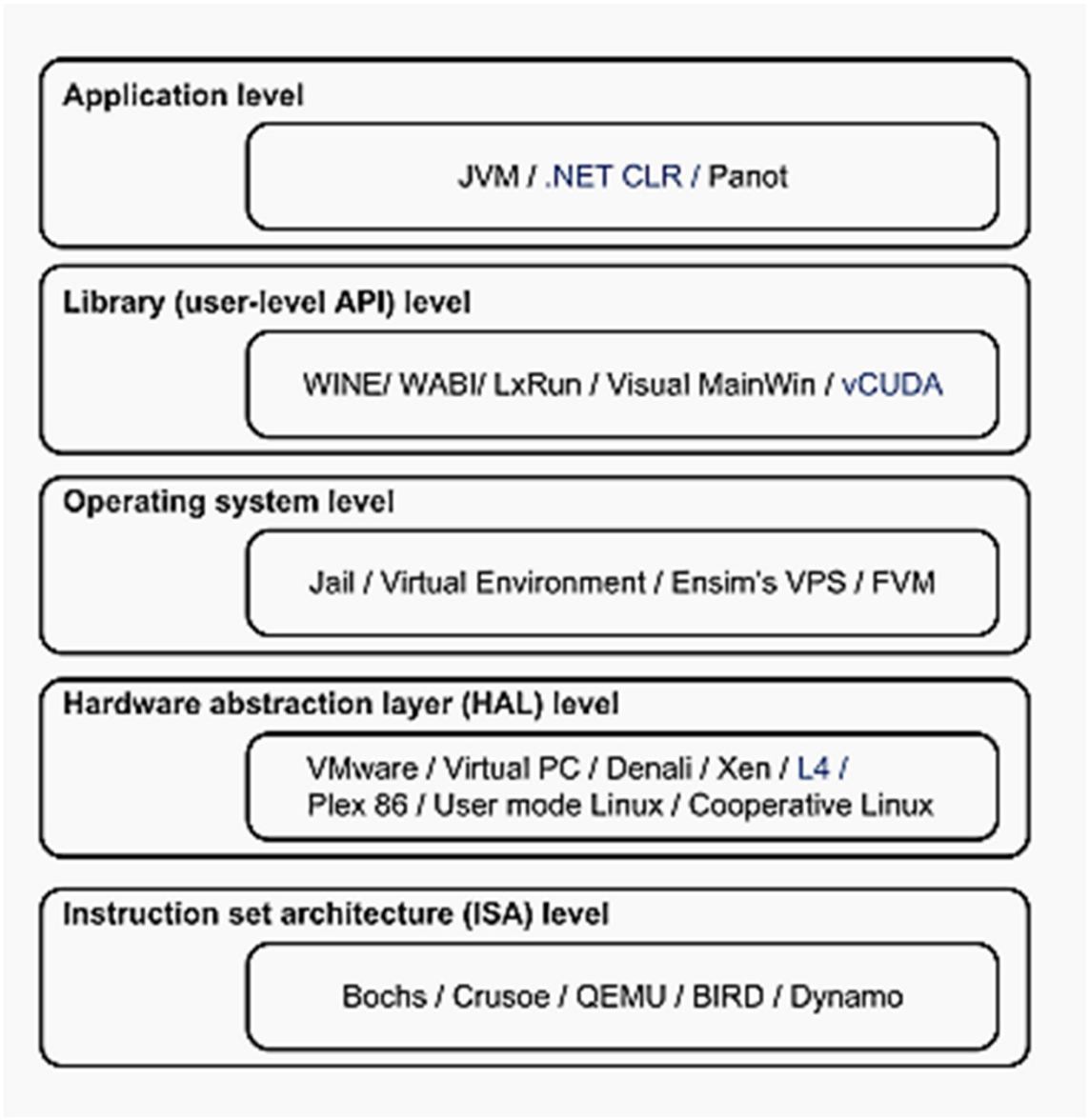
- Virtualization is a computer architecture technology by which multiple virtual machines (VMs) are multiplexed in the same hardware machine
- Hardware resources (CPU, memory, I/O devices, etc.) or software resources (operating system and software libraries) can be virtualized in various functional layers.

Levels of Virtualization Implementation



The virtualization software creates the abstraction of VMs by interposing a virtualization layer at various levels of a computer system.

Common virtualization layers include the instruction set architecture (ISA) level, hardware level, operating system level, library support level, and application level


Application level
JVM / .NET CLR / Panot
Library (user-level API) level
WINE/ WABI/ LxRun / Visual MainWin / vCUDA
Operating system level
Jail / Virtual Environment / Ensim's VPS / FVM
Hardware abstraction layer (HAL) level
VMware / Virtual PC / Denali / Xen / L4 /
Plex 86 / User mode Linux / Cooperative Linux
Instruction set architecture (ISA) level
Bochs / Crusoe / QEMU / BIRD / Dynamo

Instruction Set Architecture Level

- Virtualization is performed by emulating a given ISA by the ISA of the host machine
- The basic emulation method is through code interpretation.
- An interpreter program interprets the source instructions to target instructions one by one→slow→ dynamic binary translation
- A virtual instruction set architecture (V-ISA) requires adding a processor-specific software translation layer to the compiler

Hardware Abstraction Level

- The idea is to virtualize a computer's resources, such as processors, memory, and I/O devices
- Generating a virtual hardware environment for a VM.

Operating System Level

- An abstraction layer between traditional OS and user applications
- OS-level virtualization creates isolated containers on a single physical server and the OS instances to utilize the hardware and software in data centers.
- The containers behave like real servers.

Library Support Level

- Virtualization with library interfaces is possible by controlling the communication link between applications and the rest of a system through API hooks
- The software tool WINE
- vCUDA

User-Application Level

- To be also known as process-level virtualization
- The most popular approach is to deploy high level language (HLL) VMs
- Any program written in the HLL and compiled for this VM will be able to run on it
- Microsoft .NET CLR and Java Virtual Machine (JVM) are two good examples of this class of VM.

2. VIRTUALIZATION STRUCTURES/TOOLS AND MECHANISMS

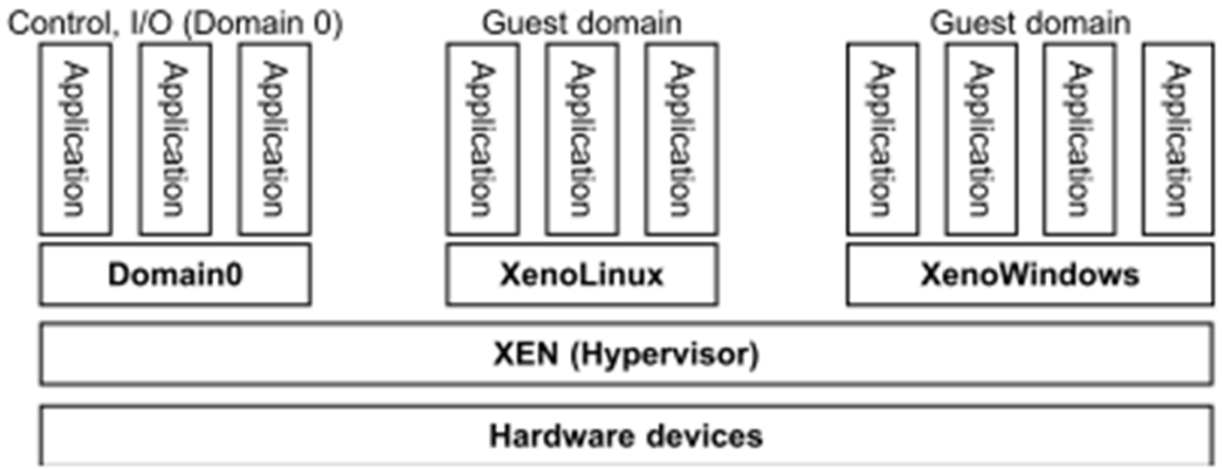
- Depending on the position of the virtualization layer, there are several classes of VM architectures:
  - Hypervisor
  - Para-virtualization,
  - Host-based virtualization

Hypervisor

- The hypervisor supports hardware-level virtualization
- The hypervisor provides hypercalls for the guest OSes and applications
- A hypervisor can assume a micro-kernel architecture like the Microsoft Hyper-V or can assume a monolithic hypervisor architecture like the VMware ESX for server virtualization
- A micro-kernel hypervisor includes only the basic and unchanging functions (physical memory management and processor scheduling)
- A monolithic hypervisor implements all the aforementioned functions, including the device drivers

Xen Architecture

- Xen is a micro-kernel hypervisor
- Separating the policy from the mechanism. The Xen hypervisor implements all the mechanisms, leaving the policy to be handled by Domain 0.



Hardware virtualization categories

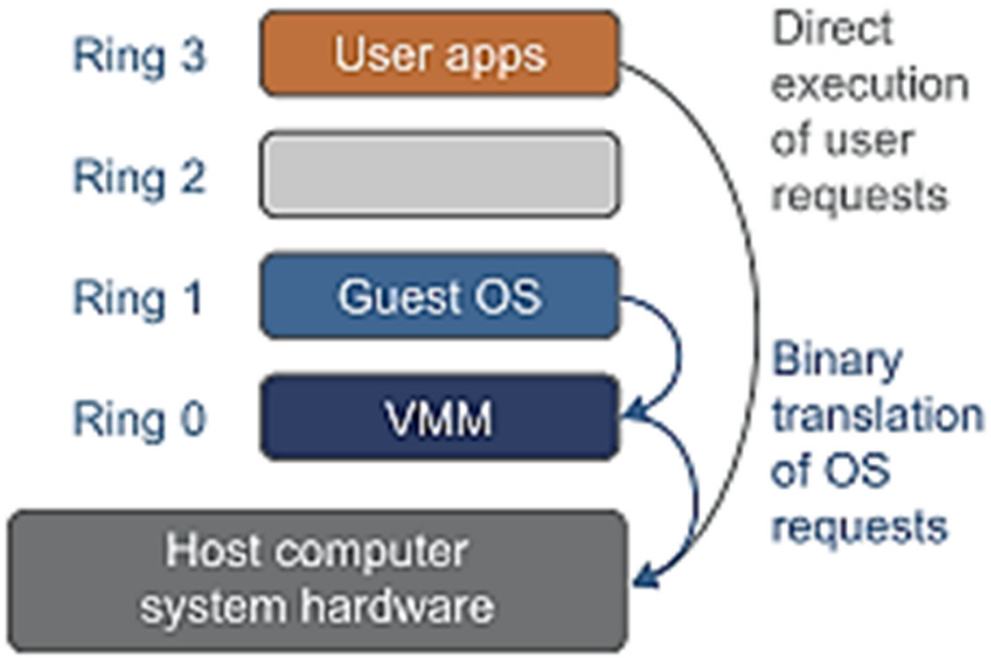
- Hardware virtualization can be classified into two categories:
  - full virtualization
  - host-based virtualization.

Full Virtualization

- No need to modify the host OS. It relies on binary translation to trap and to virtualize the execution of certain sensitive, nonvirtualizable instructions
- Noncritical instructions run on the hardware directly while critical instructions are discovered and replaced with traps into the VMM to be emulated by software
- Trapping only crictial instructions?

Host-Based Virtualization

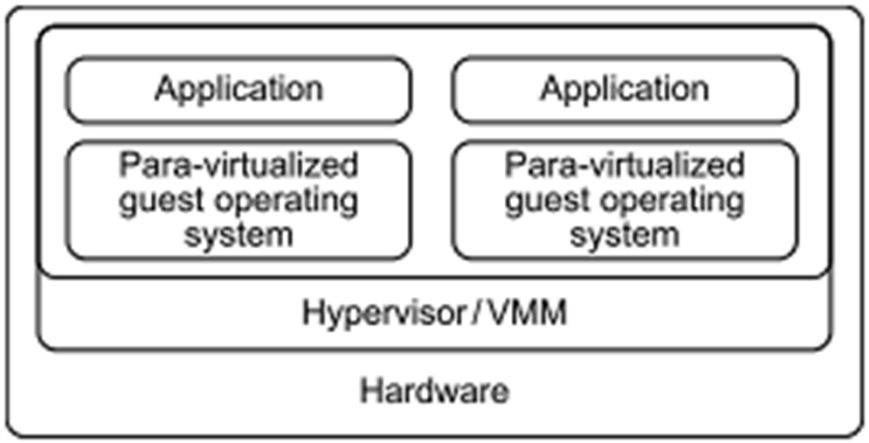
- Both a host OS and a guest OS are used.
- Host OS is still responsible for managing the hardware.
- Installing a virtualization layer on top of the host OS



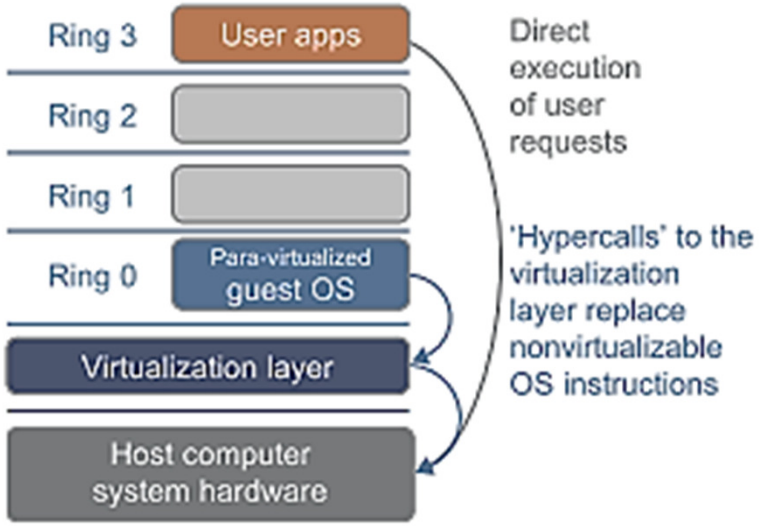
Host-based architecture

Para-Virtualization Architecture

- Para-virtualization needs to modify the guest operating systems. A para-virtualized VM provides special APIs requiring substantial OS modifications in user applications



- The traditional x86 processor offers four instruction execution rings: Rings 0, 1, 2, and 3.
- The lower the ring number, the higher the privilege of instruction being executed.
- The OS is responsible for managing the hardware and the privileged instructions to execute at Ring 0, while user-level applications run at Ring 3

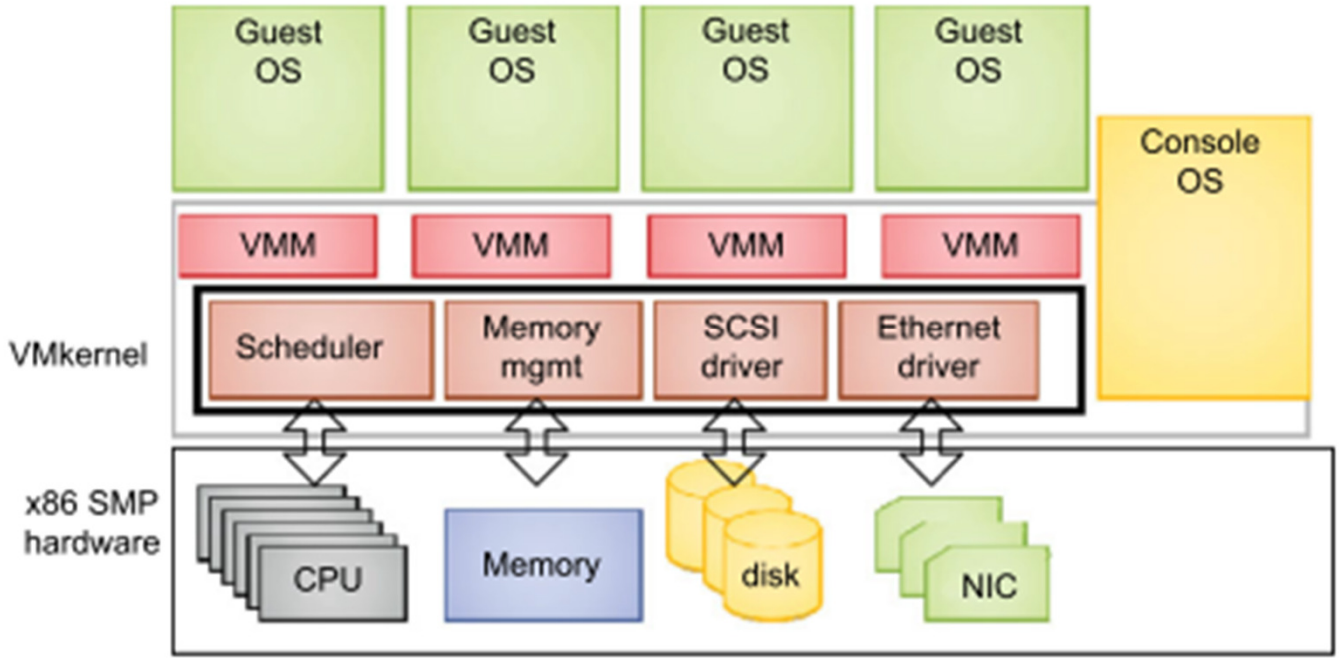


For example: KVM (Kernel-Based VM)

- A Linux para-virtualization system—a part of the Linux version 2.6.20 kernel.
- Memory management and scheduling activities are carried out by the existing Linux kernel.
- The KVM does the rest, which makes it simpler than the hypervisor that controls the entire machine.
- KVM is a hardware-assisted para-virtualization tool, which improves performance and supports unmodified guest OSes such as Windows, Linux, Solaris, and other UNIX variants

For example: VMware ESX Server

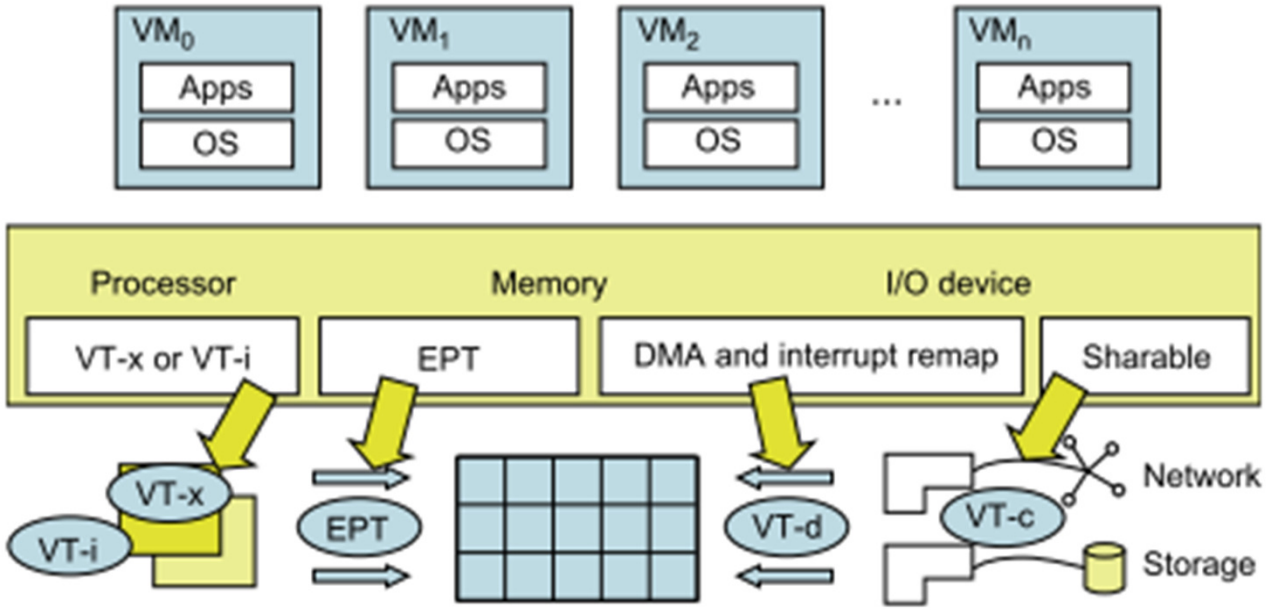
- An ESX-enabled server consists of four components: a virtualization layer, a resource manager, hardware interface components, and a service console
- To improve performance, the ESX server employs a para-virtualization architecture in which the VM kernel interacts directly with the hardware without involving the host OS.



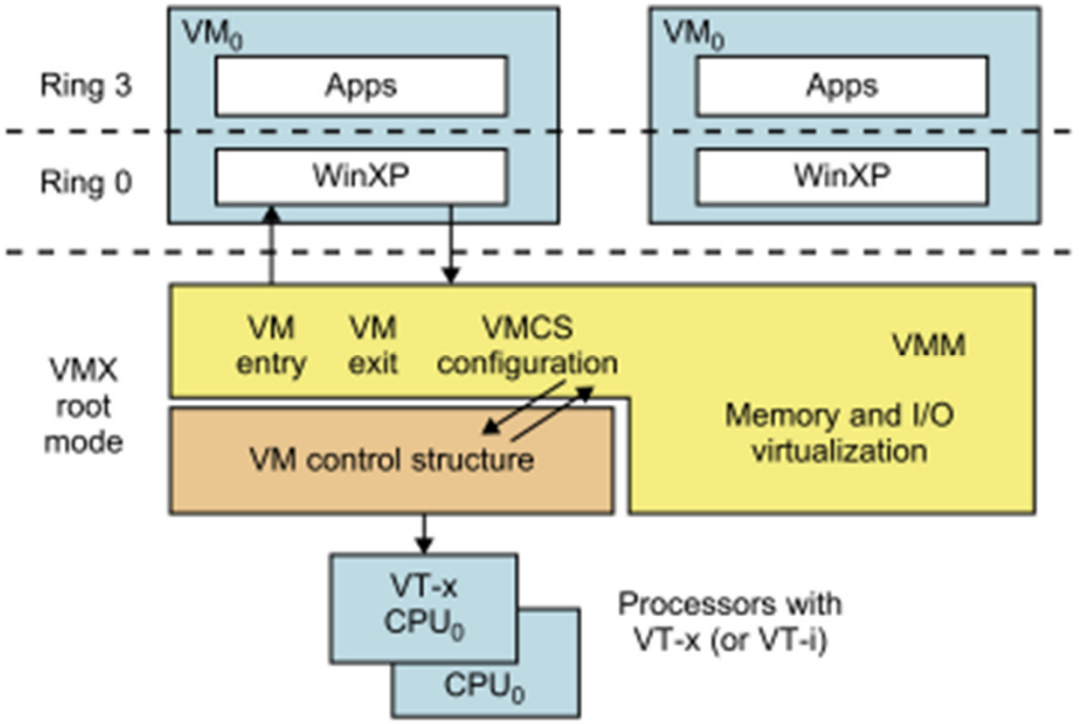
The VMware ESX server architecture using para-virtualization.

Hardware Support for Virtualization

- To support virtualization, processors such as the x86 employ a special running mode and instructions, known as hardware-assisted virtualization.
- VMM and guest OS run in different modes and all sensitive instructions of the guest OS and its applications are trapped in the VMM.
- To save processor states, mode switching is completed by hardware.
- For the x86 architecture, Intel and AMD have proprietary technologies for hardware-assisted virtualization.

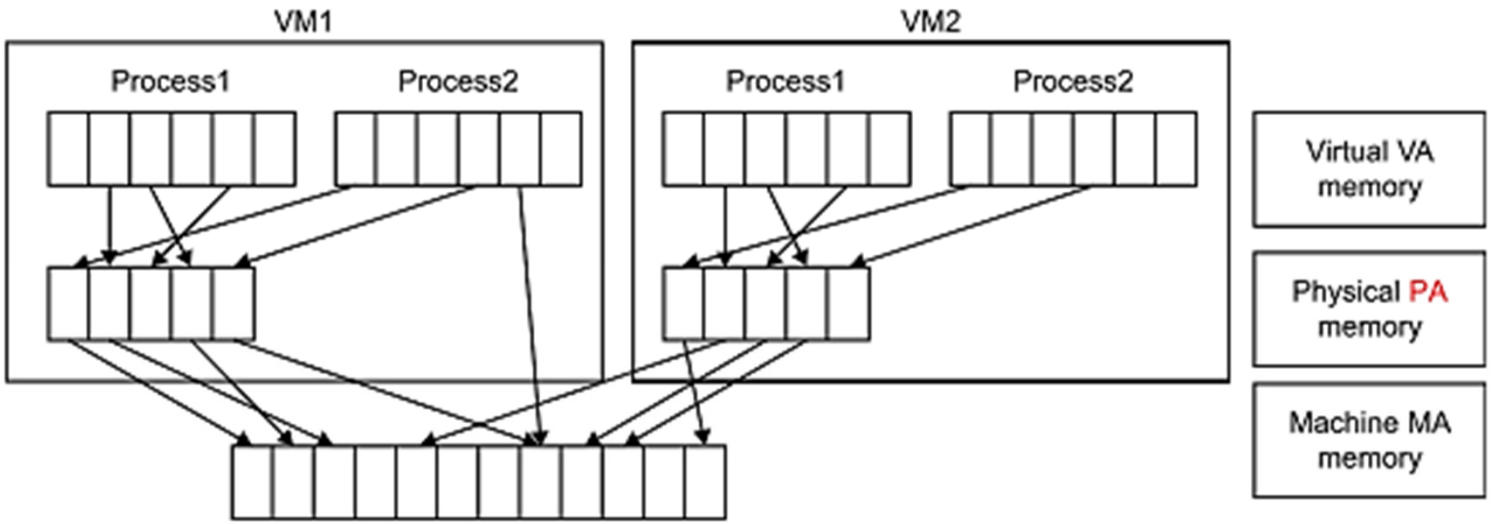


Intel hardware support for virtualization of processor, memory, and I/O devices.

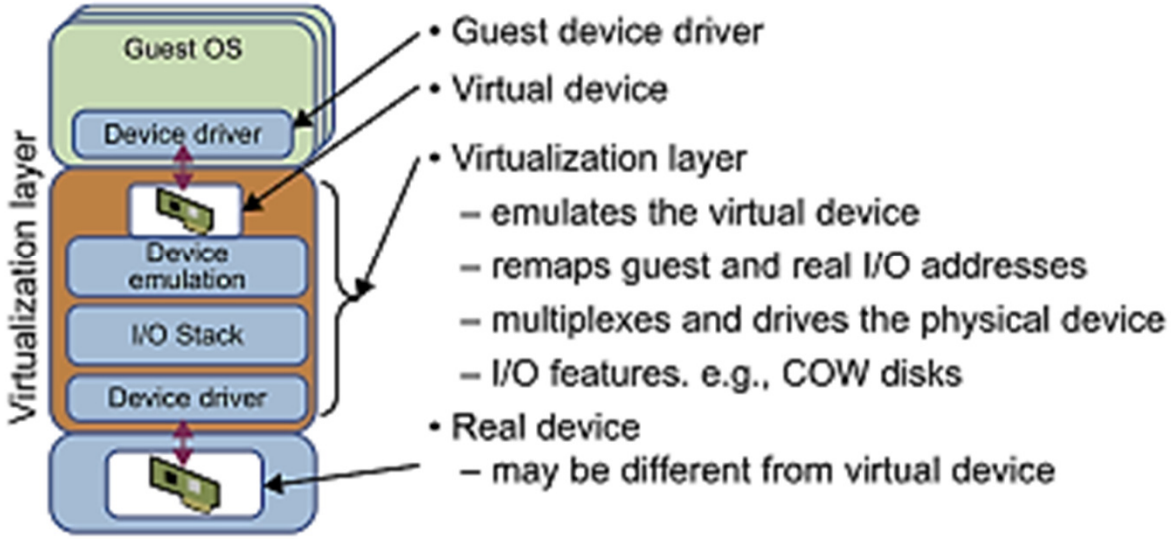


Intel hardware-assisted CPU virtualization

- In a virtual execution environment, virtual memory virtualization involves sharing the physical system memory in RAM and dynamically allocating it to the physical memory of the VMs.



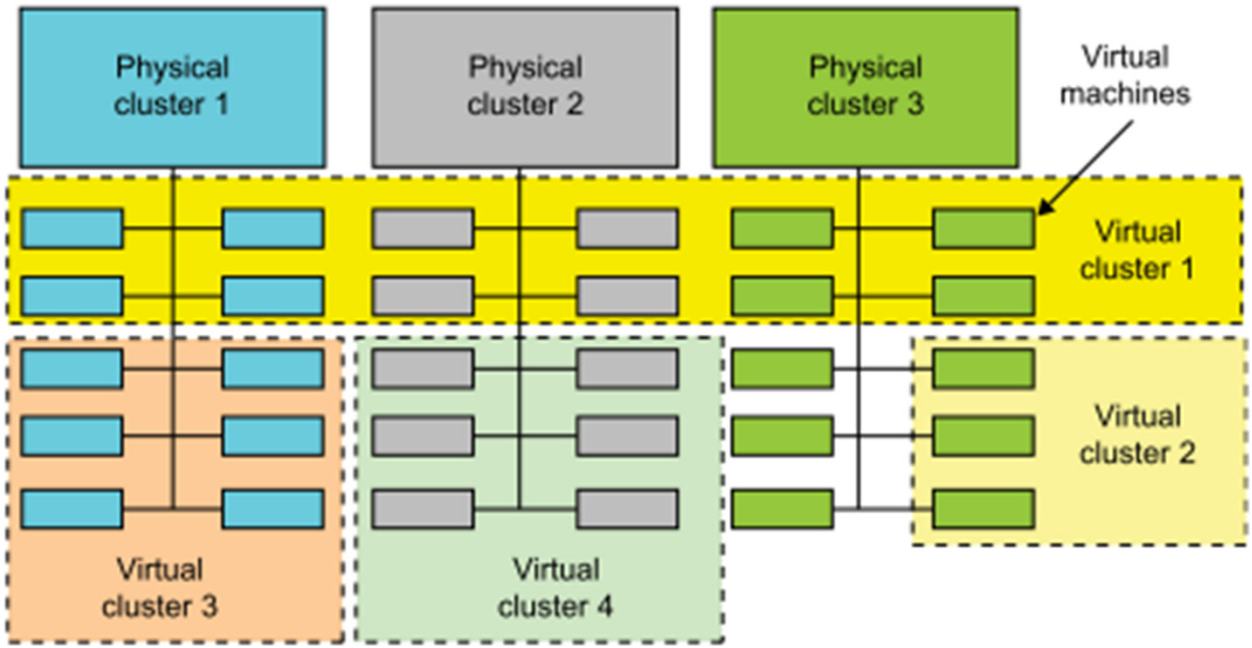
Two-level memory mapping procedure



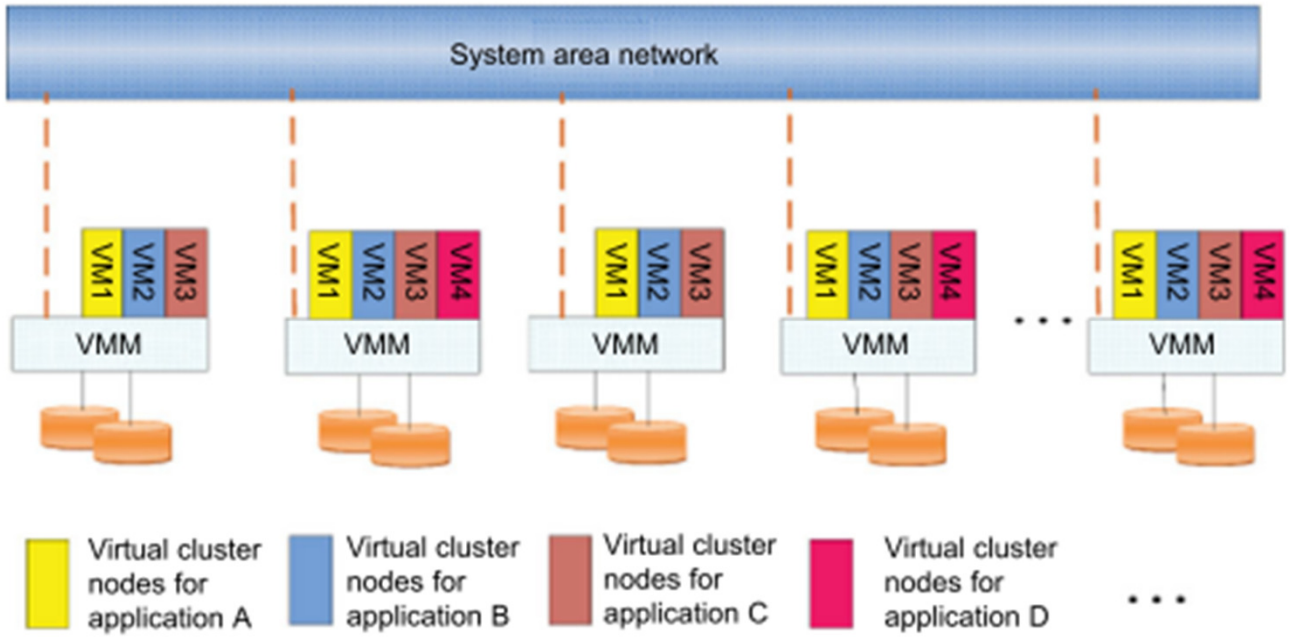
- Device emulation for I/O virtualization implemented inside the middle layer that maps real I/O devices into the virtual devices for the guest device driver to use

3. VIRTUAL CLUSTERS

- Virtual clusters are built with VMs installed at distributed servers from one or more physical clusters.
- The VMs in a virtual cluster are interconnected logically by a virtual network across several physical networks
- Each virtual cluster is formed with physical machines or a VM hosted by multiple physical clusters



A cloud platform with four virtual clusters over three physical clusters shaded differently



The concept of a virtual cluster based on application partitioning

Fast Deployment

- Fast deployment means two things:
  - to construct and distribute software stacks (OS, libraries, applications) to a physical node inside clusters as fast as possible
  - to quickly switch runtime environments from one user's virtual cluster to another user's virtual cluster.

High-Performance Virtual Storage

- Template VM, existing software packages need disk spaces
- There are four steps to deploy a group of VMs onto a target cluster: preparing the disk image, configuring the VMs, choosing the destination nodes, and executing the VM deployment command on every host.
- Many systems use templates to simplify the disk image preparation process.
- Templates could implement the COW (Copy on Write) format. A new COW backup file is very small and easy to create and transfer. Therefore, it definitely reduces disk space consumption

4. VIRTUALIZATION FOR DATA CENTER AUTOMATION

- Data-center automation means that huge volumes of hardware, software, and database resources in these data centers can be allocated dynamically to millions of Internet users simultaneously, with guaranteed QoS and cost-effectiveness
- This automation process is triggered by the growth of virtualization products and cloud computing services

Server Consolidation in Data Centers

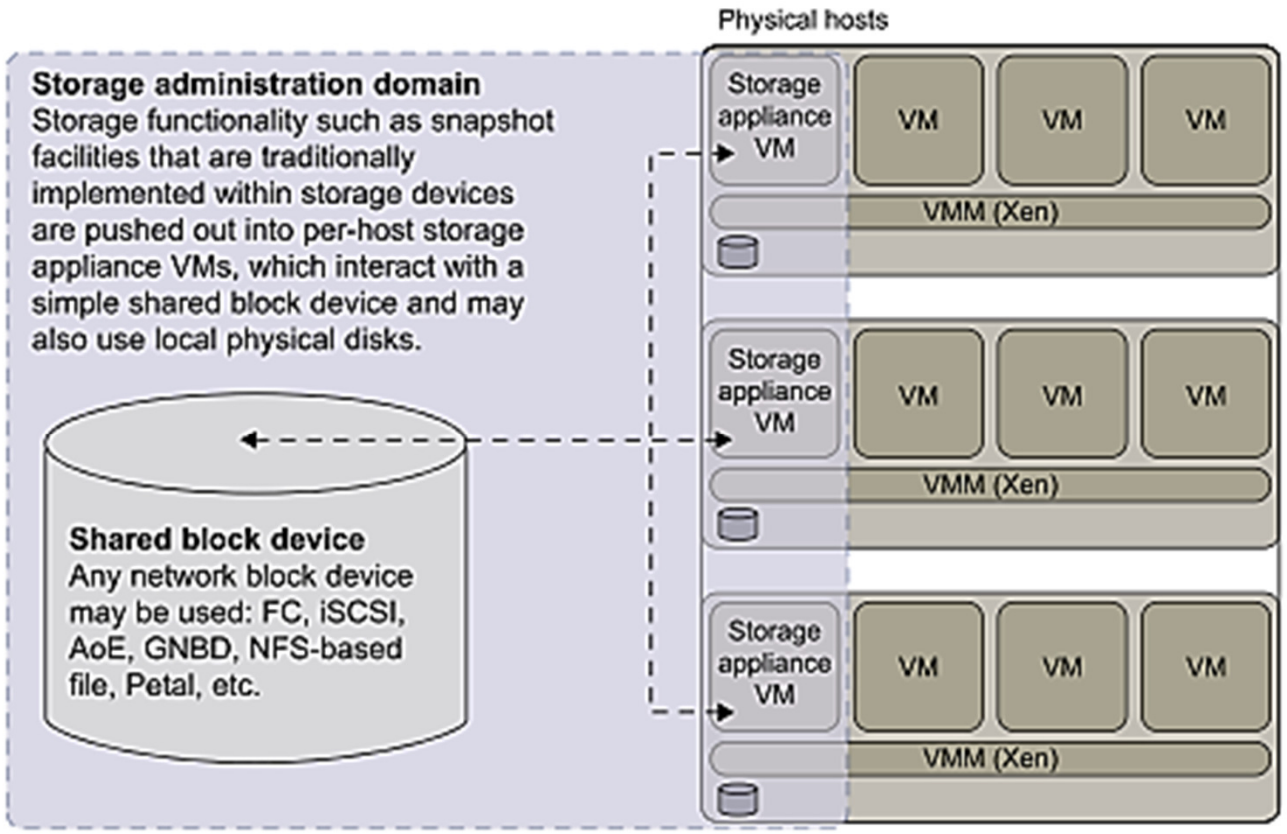
- Most servers in data centers are underutilized. A large amount of hardware, space, power, and management cost of these servers is wasted
- Server consolidation is an approach to improve the low utility ratio of hardware resources by reducing the number of physical servers
- Among several server consolidation techniques such as centralized and physical consolidation, virtualization-based server consolidation is the most powerful

Virtual Storage Management

- The storage systems become the main bottleneck of VM deployment
- In virtualization environments, a virtualization layer or modifying traditional operating system to support virtualization complicate storage operations

Parallax system architecture

- Parallax is a distributed storage system customized for virtualization environments
- For each physical machine, Parallax customizes a special storage appliance VM. The storage appliance VM acts as a block virtualization layer between individual VMs and the physical storage device. It provides a virtual disk for each VM on the same physical machine



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

Cloud OS for Virtualized Data Centers

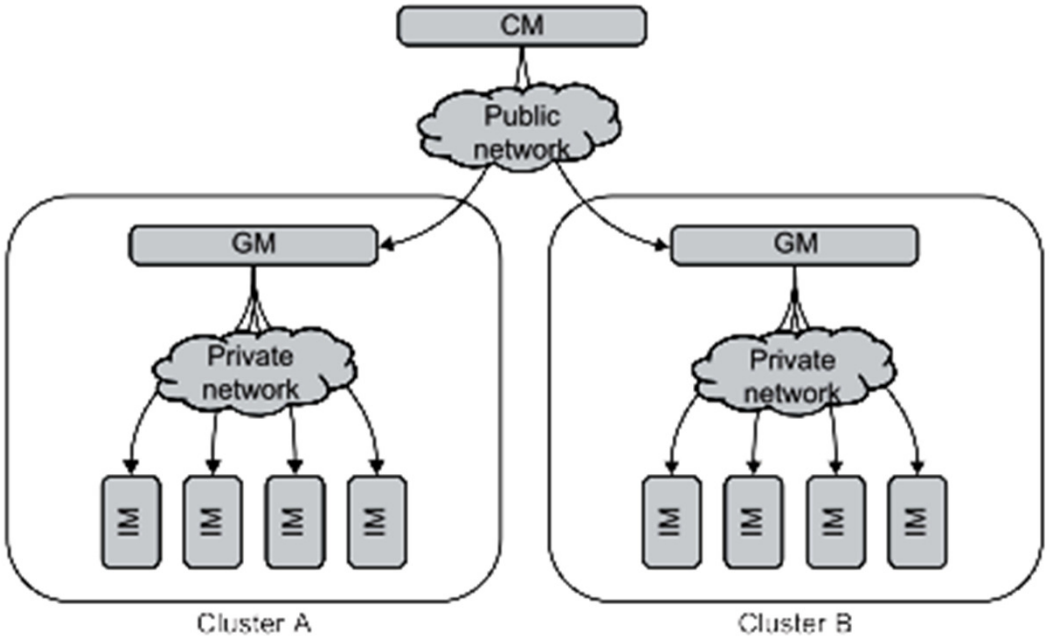
- Data centers must be virtualized to serve as cloud providers
- Following summarizes four virtual infrastructure (VI) managers and OSes:

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

- These VI managers are used to create VMs and aggregate them into virtual clusters as elastic resources.
- Nimbus and Eucalyptus support essentially virtual networks. OpenNebula has additional features to provision dynamic resources and make advance reservations.
- All three public VI managers apply Xen and KVM for virtualization.
- vSphere 4 uses the hypervisors ESX and ESXi from VMware.
- Only vSphere 4 supports virtual storage in addition to virtual networking and data protection

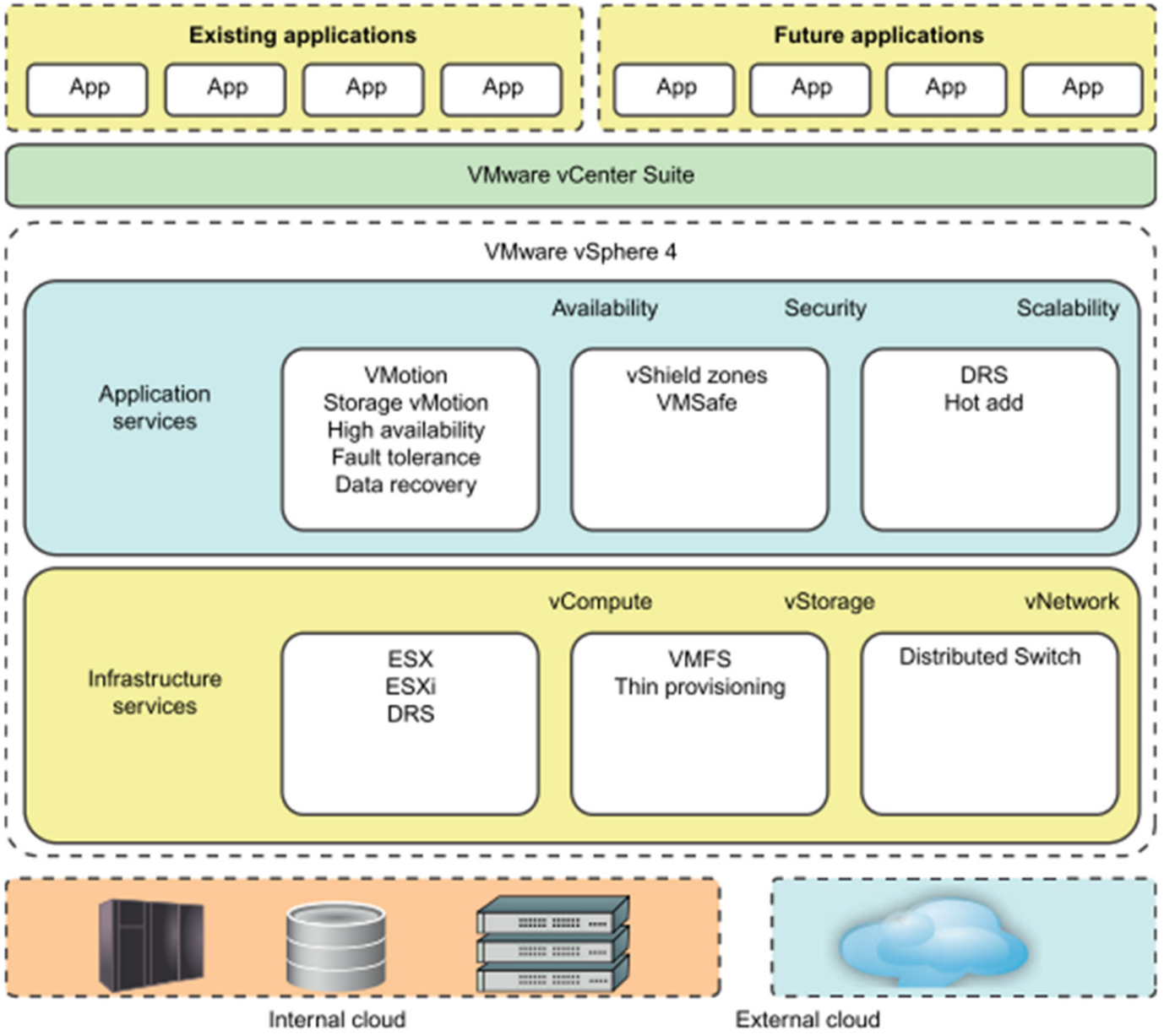
Eucalyptus for Virtual Networking of Private Cloud

- Eucalyptus is an open source software system intended mainly for supporting Infrastructure as a Service (IaaS) clouds
- Each high-level system component as a stand-alone web service. Each web service exposes a well-defined language-agnostic API in the form of a WSDL document containing both operations that the service can perform and input/output data structures



VMware vSphere 4 as a Commercial Cloud OS

- vSphere 4 offers a hardware and software ecosystem
- vSphere extends earlier virtualization software products by VMware, namely the VMware Workstation,
  - ESX for server virtualization, and
  - Virtual Infrastructure for server clusters
- The system interacts with user applications via an interface layer, called vCenter.

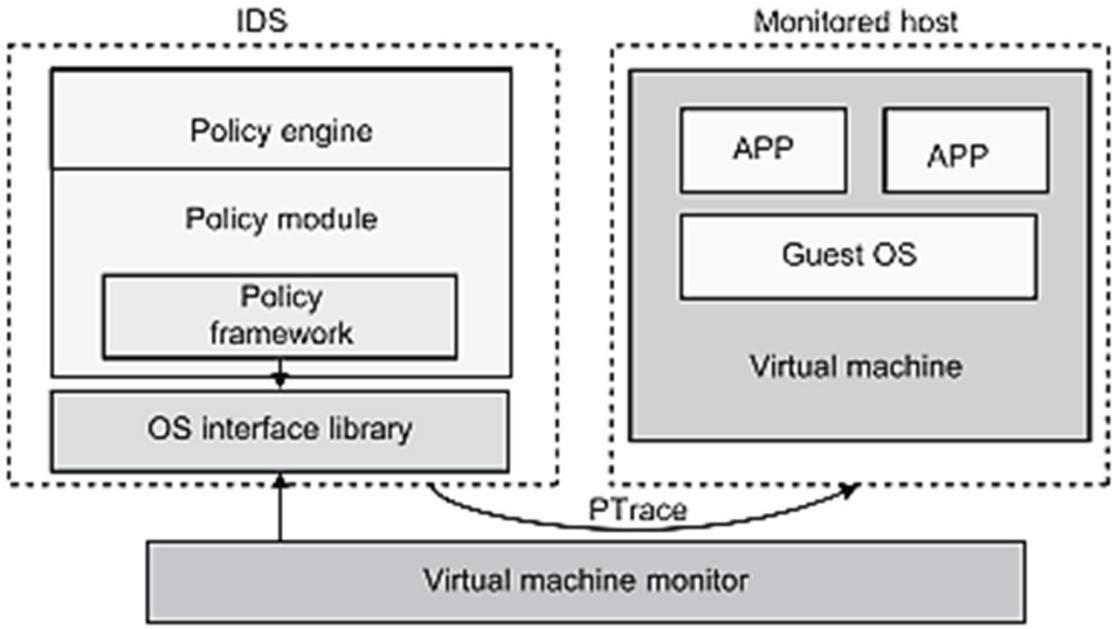


vSphere/4 overall architecture

- The vSphere 4 is built with two functional software suites: infrastructure services and application services.
- Three component packages intended mainly for virtualization purposes:
  - vCompute: is supported by ESX, ESXi, and DRS virtualization libraries from VMware;
  - vStorage is supported by VMS and thin provisioning libraries;
  - vNetwork offers distributed switching and networking functions
- The application services are also divided into three groups: availability, security, and scalability.
- To fully understand the use of vSphere 4, must also learn how to use the vCenter interfaces in order to link with existing applications or to develop new applications

Trust Management in Virtualized Data Centers

- VMM is the base of the security of a virtual system. Normally, one VM is taken as a management VM to have some privileges such as creating, suspending, resuming, or deleting a VM
- Once a hacker successfully enters the VMM or management VM, the whole system is in danger
- Virtualization-based intrusion detection can isolate guest VMs on the same hardware platform



An IDS to run on a VMM as a high-privileged VM

The End