BIT4440

CLOUD COMPUTING AND BIG DATA

MR LISTONE KAPUTULA

Cloud Computing and Big Data

- Introduction
- Cloud Computing Technologies
- Big Data Technologies
- Data Management in the Cloud
- Big Data Analytics in the Cloud
- Virtualization for Cloud
- Cloud Security
- Cloud Application Development Tools and Frameworks
- Cloud Application Development
- Cloud Deployment and Management
- Cloud Project Development

Virtualization for Cloud

Introduction

- Virtualization is a technology **that helps us to install different Operating Systems on hardware**.
- Virtualization is **the process of creating a virtual representation of physical resources such as servers, storage, and networks**.
- It allows **multiple virtual instances to run on a single physical machine or infrastructure**.
- Virtualization abstracts hardware resources from the underlying physical infrastructure.

Virtualization for Cloud

Definition

- Virtualization in cloud computing refers to the abstraction of physical computing resources into virtual instances that can be easily provisioned, managed, and scaled on-demand.
- Physical computing resources may include:
  - ✓Servers, storage, and networks, operating system or desktop

Virtualization for Cloud

Importance of Virtualization

- Virtualization plays a pivotal role in cloud computing:
- **Resource Efficiency**: Virtualization enables efficient resource utilization by allowing multiple virtual machines (VMs) or containers to share physical resources, reducing resource wastage.
- **Isolation**: Virtualization ensures isolation between VMs or containers, preventing one from affecting the others, which is crucial for security and stability.

Virtualization for Cloud

Importance of Virtualization

- **Scalability**: Virtualization facilitates easy scalability by adding or removing virtual instances based on demand, making cloud infrastructure highly adaptable.
- **Cost Savings**: Virtualization reduces hardware procurement costs, as you can get more out of existing infrastructure.
- **Migration and Disaster Recovery**: Virtualization simplifies migration and disaster recovery processes by encapsulating entire VMs for easy transport and backup.

Virtualization for Cloud

Virtualization Technologies

- Three fundamental virtualization technologies: **Virtual Machines (VMs), Hypervisors, and Containers**.
- Virtual Machines (VMs):
  - ✓VMs are complete virtualized computing environments **that mimic physical hardware, including a virtual CPU, RAM, storage, and network interfaces**.
  - ✓Each VM runs its own full-fledged operating system.

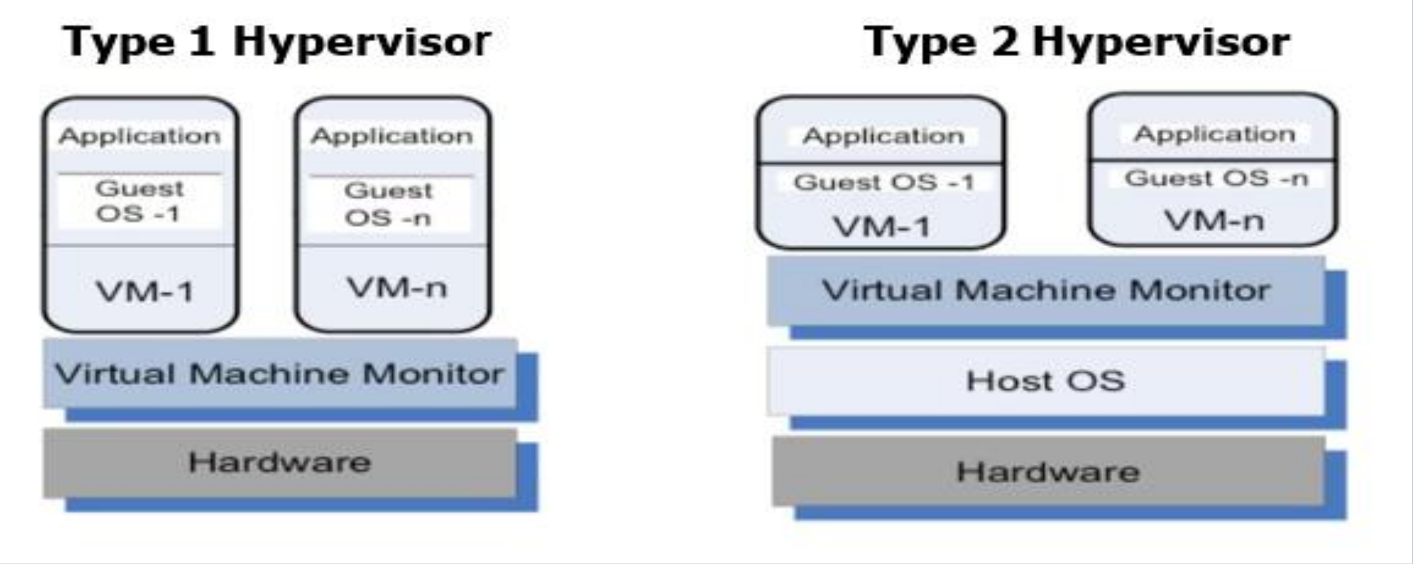
Virtualization for Cloud

Virtualization Technologies

- Hypervisors:
  - ✓Also known as Virtual Machine Manager or Virtual Machine Monitor (VMM).
  - ✓Hypervisors are software or hardware-based virtualization platforms that manage and run VMs.
  - ✓They sit between the physical hardware and VMs, allocating resources and ensuring isolation.
  - ✓There two types of hypervisors

Virtualization for Cloud

Virtualization Technologies



- **Taxonomy of VMMs:**
  1. Type 1 Hypervisor (**bare metal, native**): supports multiple virtual machines and runs directly on the hardware (e.g., VMware ESX , Xen, Denali)
  2. Type 2 Hypervisor (**hosted**) VM - runs under a host operating system (e.g., user-mode Linux)

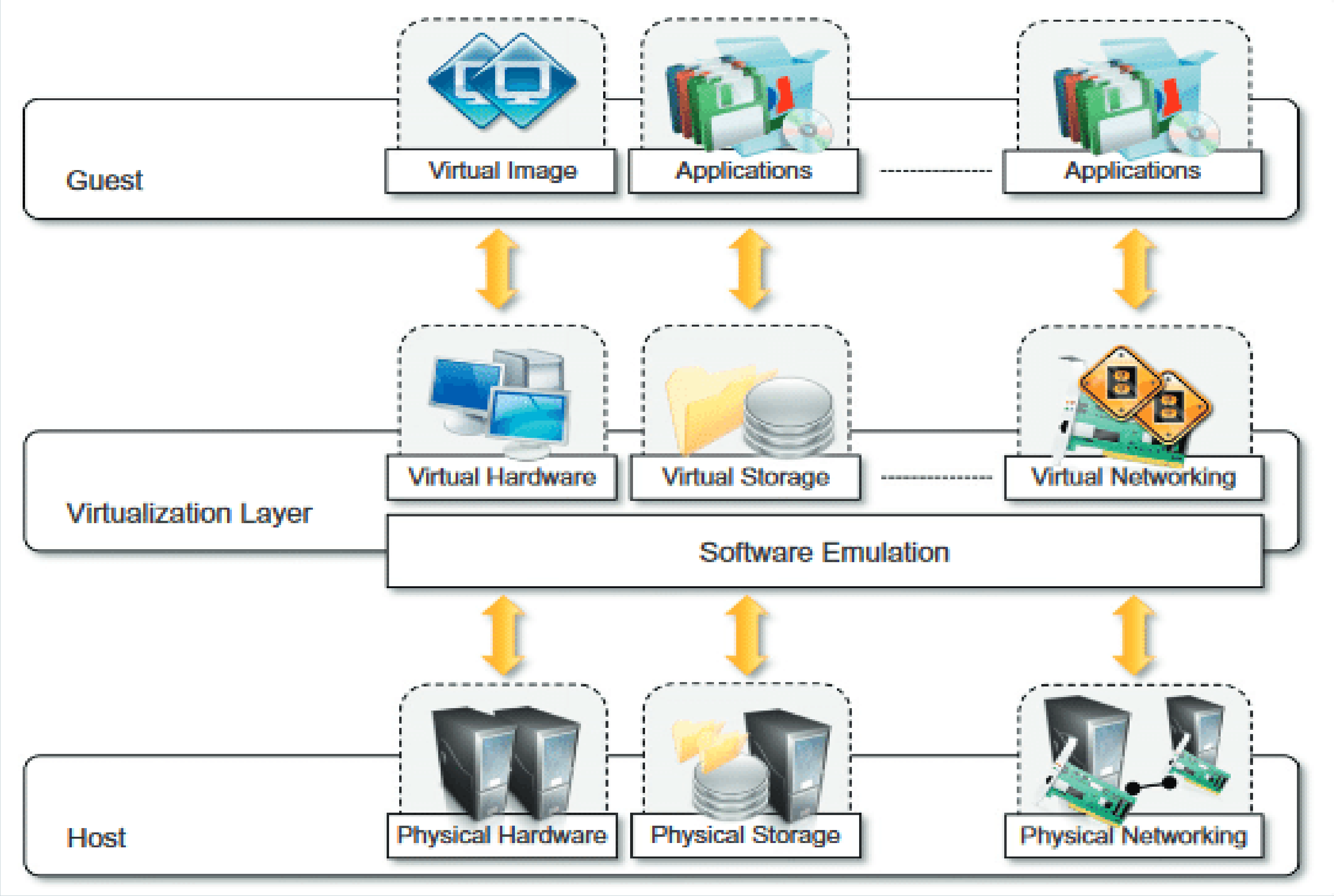
Virtualization for Cloud

Virtualization Technologies

- Containers:
  - ✓Containers are lightweight, portable, and isolated application environments that share the host OS kernel.
  - ✓They package an application and its dependencies into a single unit, making it easy to deploy and manage.

Virtualization for Cloud

Virtualization Reference Model



Virtualization for Cloud

Virtualization Reference Model

Guest

- The guest represents **the system component that interacts with the virtualization layer** rather than with the host, as would normally happen.
- Guests usually **consist of one or more virtual disk files, and a VM definition file**.
- Virtual Machines are centrally **managed by a host application that sees and manages each virtual machine as a different application.**

Virtualization for Cloud

Virtualization Reference Model

Virtualization Layer

- The virtualization layer is responsible **for recreating the same or a different environment where the guest will operate.**
- It is an additional abstraction layer between a network and storage hardware, **computing, and the application running on it.**
- Usually it helps to **run a single operating system per machine which can be very inflexible compared to the usage of virtualization**.

Virtualization for Cloud

Virtualization Reference Model

Host

- The host represents the **original environment where the guest is supposed to be managed**.
- Each guest runs on the host using shared resources donated to it by the host.
- The operating system, **works as the host and manages the physical resource management, and the device support.**

Virtualization for Cloud

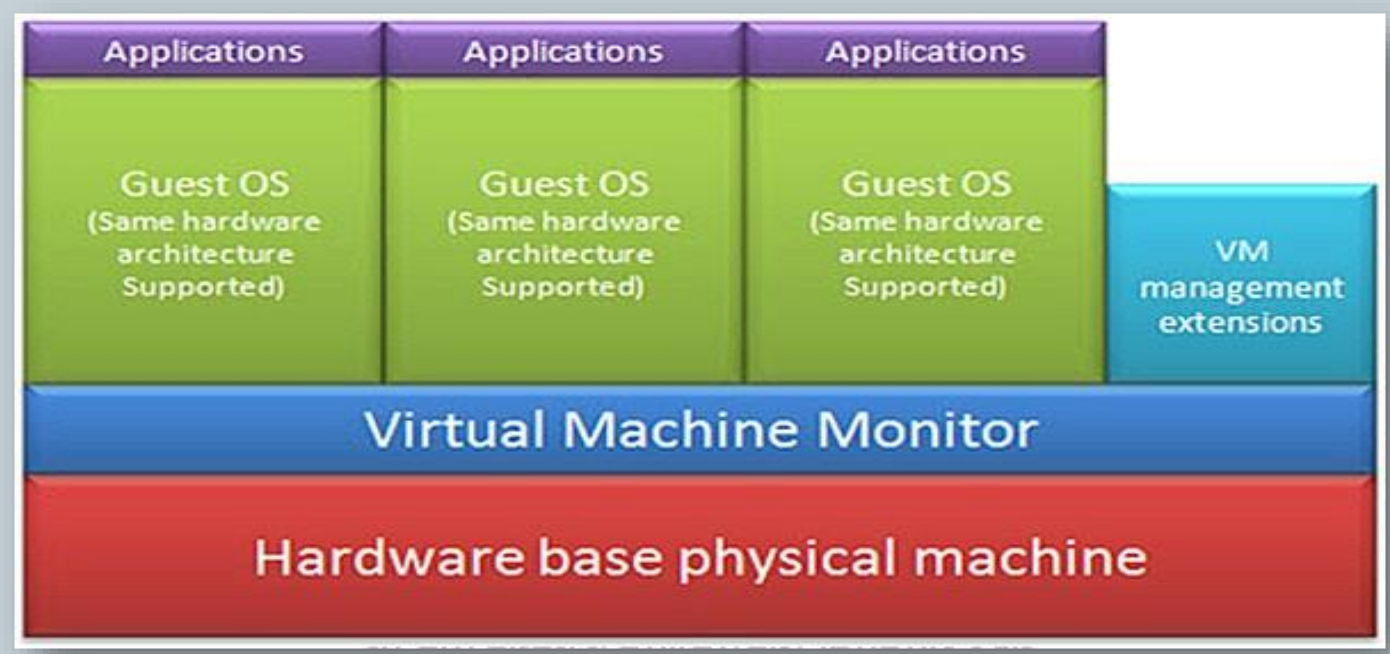
Virtualization Approaches

- **Full-Virtualization**
  - ✓Full-virtualization works with unmodified guest operating systems.
  - ✓It doesn't require any modifications to the guest OS.
- **Use Cases:**
  - ✓Full-virtualization is suitable when you need to run **a variety of guest operating systems**, including legacy and proprietary ones, without modification.
  - ✓It is commonly used in **desktop virtualization, server virtualization, and cloud computing environments**.

Virtualization for Cloud

Virtualization Approaches

- Full-Virtualization



Pros	**Need not to modify guest OS**
Cons	**Significant performance hit**

Virtualization for Cloud

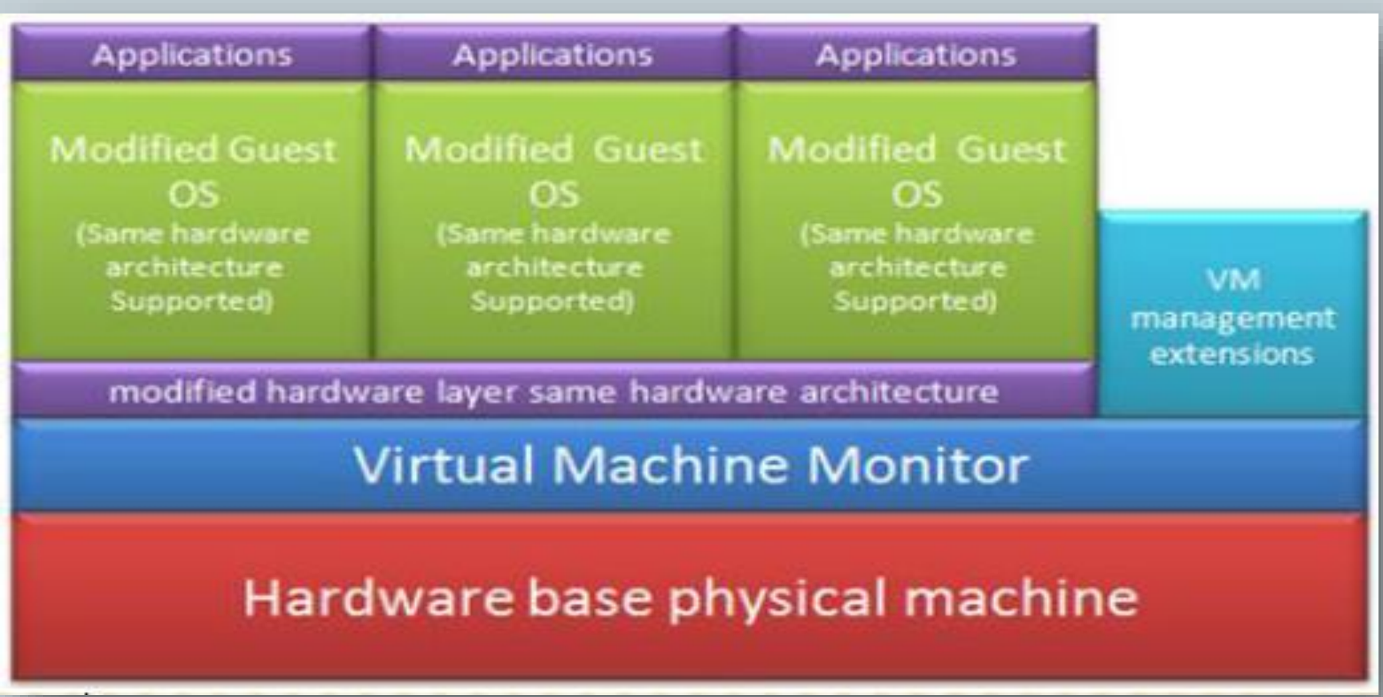
Virtualization Approaches

- **Para-Virtualization**
  - ✓In para-virtualization, the guest OS is modified to interact with the hypervisor using special API calls.
- Use Cases:
  - ✓Para-virtualization is suitable **when you have control over the guest OS and can modify it** to improve virtualization performance.
  - ✓It is commonly used in scenarios where performance is a critical factor, such as **high-performance computing (HPC) clusters and data centers**.

Virtualization for Cloud

Virtualization Approaches

- Para-Virtualization

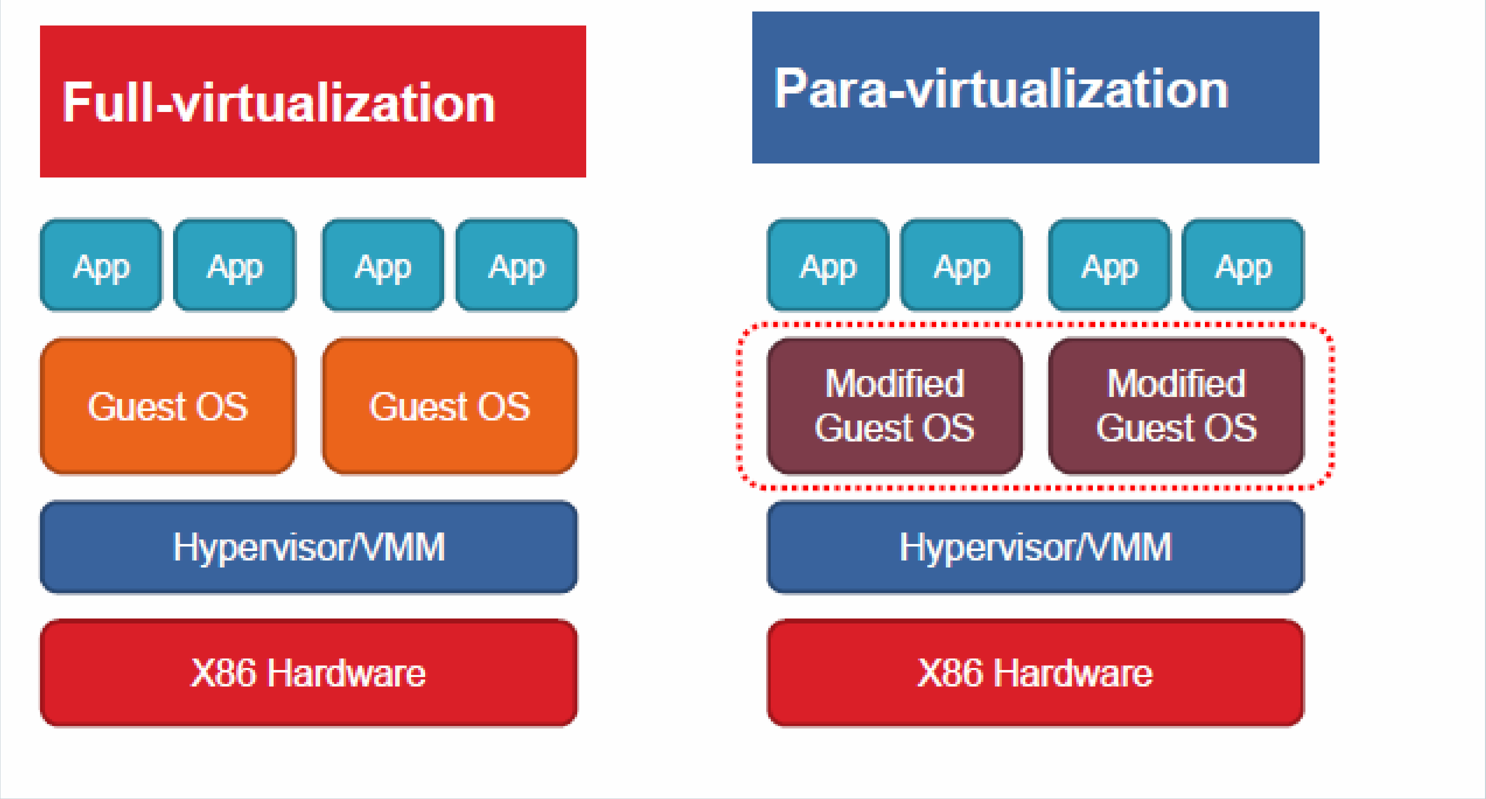


Pros	**Light weight and high performance**
Cons	**Require modification of guest OS**

Virtualization for Cloud

Virtualization Approaches

- Full-Virtualization Vs Para-Virtualization



Virtualization for Cloud

Types of Virtual Machine

- The two main types are ***System and Process Virtual Machines.***
- **System VMs** emulate an entire physical computer, providing a complete operating system environment and virtualized hardware to run multiple guest operating systems concurrently.
- A System Virtual Machine is also known as **Hardware Virtual Machine** or full virtualization

Virtualization for Cloud

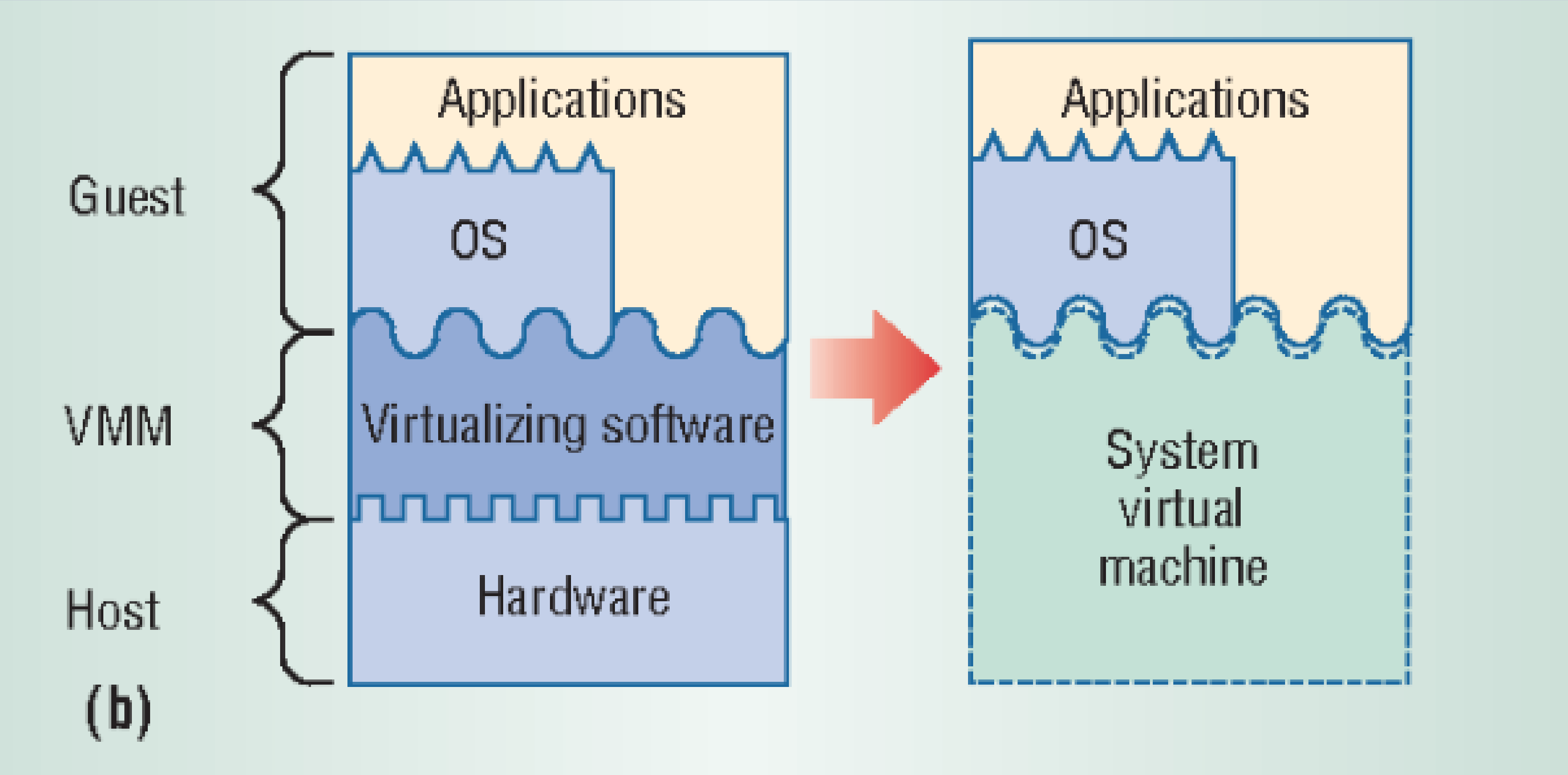
Types of Virtual Machine

- Examples of System VMs software - **VMware, VirtualBox, Windows Virtual PC, Parallels, QEMU, Citrix Xen.**
- Use Cases:
  - ✓ They are commonly used in **data centers and cloud computing environments** for server consolidation, workload isolation, and running multiple operating systems on the same physical server.

Virtualization for Cloud

Types of Virtual Machine

System Virtual Machine



Virtualization for Cloud

Types of Virtual Machine

- **Process VMs** are designed to run individual processes or applications in an isolated environment, rather than emulating an entire operating system.
- A Process Virtual Machine **is also called a Language Virtual Machine or an Application Virtual Machine or Managed Runtime Environment**.

Virtualization for Cloud

Types of Virtual Machine

- Process virtual machines **are implemented using an interpreter**;
- For improving performance **these virtual machines will use just-in-time compilers internally**.
- Examples of Process VMs –
  - ✓**Parrot Virtual Machine**: It is a process VM designed to support multiple high-level programming languages.

Virtualization for Cloud

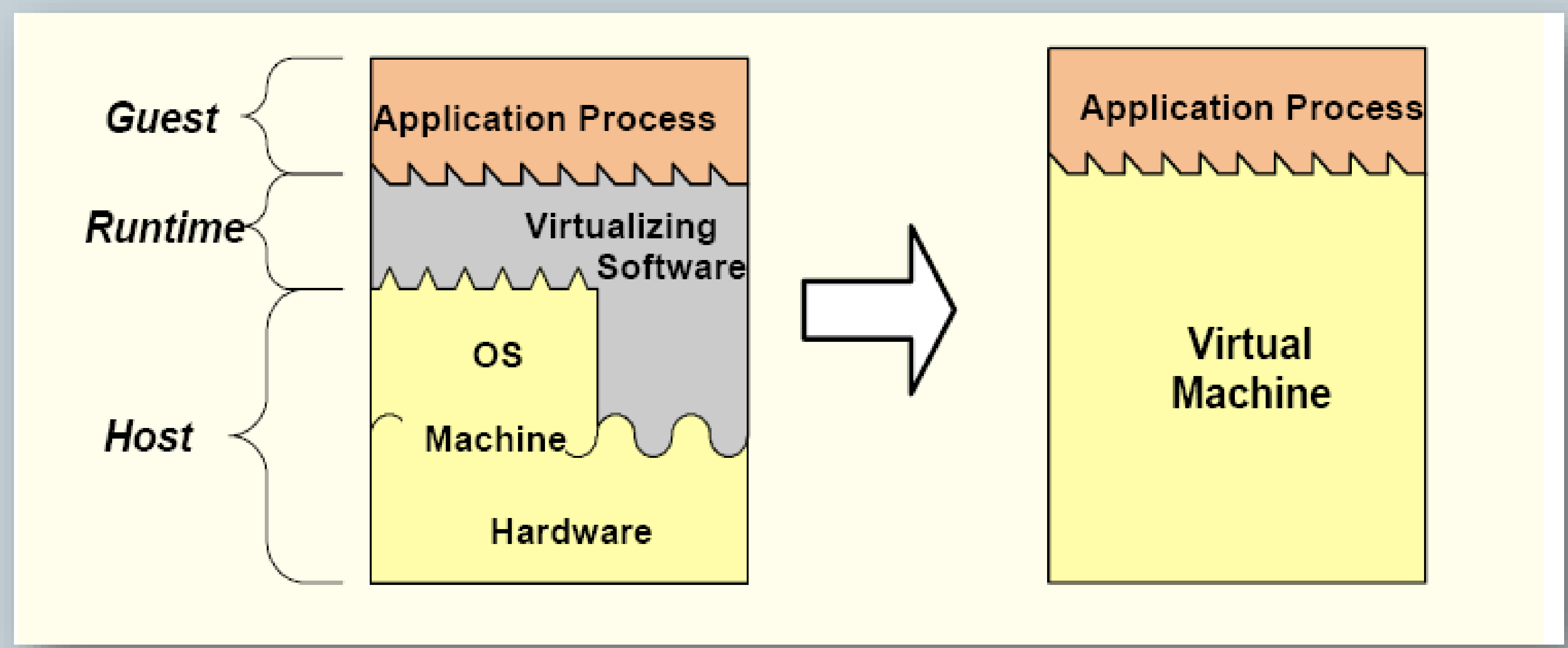
Types of Virtual Machine

- Examples of Process VMs –
  - ✓**Parrot Virtual Machine**: It is a process VM designed to support multiple high-level programming languages.
  - ✓**QEMU**: It is a process VM that can emulate different hardware platforms and run various operating systems.
  - ✓**Wine**: It is a process VM that allows running Windows applications on Linux and other Unix-like operating systems.

Virtualization for Cloud

Types of Virtual Machine

- It provides a runtime environment **to execute a single program and supports a single process**.



Virtualization for Cloud

HLL VM

- HLL VM (High-Level Language Virtual Machine)
  - ✓ It is a system that provides **a process with an execution environment that does not correspond to any particular hardware platform**.
- HLL VMs are designed **to execute programs written in high-level programming languages**, such as Java, C#, Python, and Ruby.

Virtualization for Cloud

HLL VM

- They **provide language-specific runtime environments** that abstract hardware details and enable cross-platform compatibility.
- HLL VMs are language-agnostic and can **run code written in multiple high-level languages**.
- Examples include the **Java Virtual Machine (JVM) and the Common Language Runtime (CLR) for .NET languages**.

Virtualization for Cloud

HLL VM

- Examples of HLL VM:
  - ✓Java Virtual Machine (JVM): It provides an execution environment for Java programs.
  - ✓.NET Common Language Runtime (CLR): It provides an execution environment for .NET applications, including those written in C#, Visual Basic, and F#.
  - ✓Python Virtual Machine (PVM): It provides an execution environment for Python programs.
  - ✓Ruby Virtual Machine (YARV): It is the virtual machine for the Ruby programming language.

Cloud Computing and Big Data

- Introduction
- Cloud Computing Technologies
- Big Data Technologies
- Data Management in the Cloud
- Big Data Analytics in the Cloud
- Virtualization for Cloud
- Cloud Security
- Cloud Application Development Tools and Frameworks
- Cloud Application Development
- Cloud Deployment and Management
- Cloud Project Development