Unit-I
Overview To Virtualization

Abhirup Khanna
Department of Virtualization

Overview To Virtualization

- 1.1 Traditional IT Infrastructure
- 1.2 Benefits of Virtualization
- 1.3 Characteristics of Virtualized Environments
- 1.4. Comparison of virtualized and non-virtualized systems
- 1.5. History of Virtualization

Virtualization

- Builds an abstraction of the existing system.
- A **generic logic** definition of the system is built into software.
- Useful to replicate the behavior of the **cloned system** over multiple instances.
- These multiple instances can then be made to run on the same hardware system it clones, and can share resources through a **hypervisor** that multiplexes the requests to the single system real hardware instance.

Virtualization facilitates

- **Elimination of** most of the inflexibilities inherent in the hardware system.
- **Better manageability** leading to a better utilization of the system.

Traditional IT Infrastructure

- Traditionally, organizations have relied on **physical infrastructures** to manage data and information.
- With the explosion of information and related data, the **physical footprint** required to manage the information base **has grown** tremendously.

Shortcomings of Physical Infrastructure

Asset Management

- Increasingly difficult to keep track of physical server assets.
- Adhoc increase in servers, datacenters, storage dispersed in diverse geographical locations. [to manage huge information base increase]

A typical organization runs at-least 2-3 datacenters to meet their IT computing and availability needs.

- The **heterogeneity** of the infrastructure also makes it difficult to run any common manageability applications to track the physical assets.

Tracking Utilization:

- unmanageability of physical systems leading to difficulty to track the utilization on each system.
- As a result, a datacenter either ended up with systems that are simply powered on but put to no good use, or systems that are running at maximum load and low performance, in spite of available free capacity in the datacenter.

Contd…

Security and Compliance:

- Security lapses due to inefficient datacenter management and monitoring
- Not able to achieve complex compliance requirements.
- Lost cost over ensuring security/compliance in the infrastructure.

Contd…

- **Provisioning**
- In traditional IT datacenter takes on an average 6-8 weeks.
- Involve huge time and complexity in creating rack-space, setting up the physical server and storage, installation of OS and patches, setting up the applications and allied components.
- Additional time investment to test the setup after configuration.
- A large part of testing is manual to test things that were configured manually.
- Good scope for errors at various stages of configuration and testing.
  - Also:
  - **Staff for Administration:**
  - Exponential increase in staff requirements for managing the ever-growing and complex physical infrastructure.

Provisioning (contd..)

- **Sizing**
- Any new requirement, the physical infrastructure has to be resized well in advance.
- In future, if the requirement goes down, the surplus infrastructure remains largely unutilized for a long time, before it is marked for another purpose.
- This is largely due to the inflexibility inherent in the physical infrastructure.
- **Optimization**
- Lack of any granular mechanism to monitor the complex infrastructure for utilization and performance, leading to difficulty to optimize on an ongoing basis.
- Too high optimization effort as against justification for any cost-savings.

Benefits of Virtualization

- Virtualization aims to mitigate most of the issues of traditional IT infrastructure by building a more **flexible** and **manageable** IT infrastructure.

Benefits of Virtualization: 3 Categories

- **Utilization:**
- Virtualization builds an abstraction of hardware system resources in software.
- Each OS/Application can be run in its own isolated environment called the **virtual machine**.
- Sharing of system resources at a much granular level and brings up the utilization of the physical system.
- Consolidates applications running on separate systems onto one system leading to save datacenter space, power and cooling.

More…..

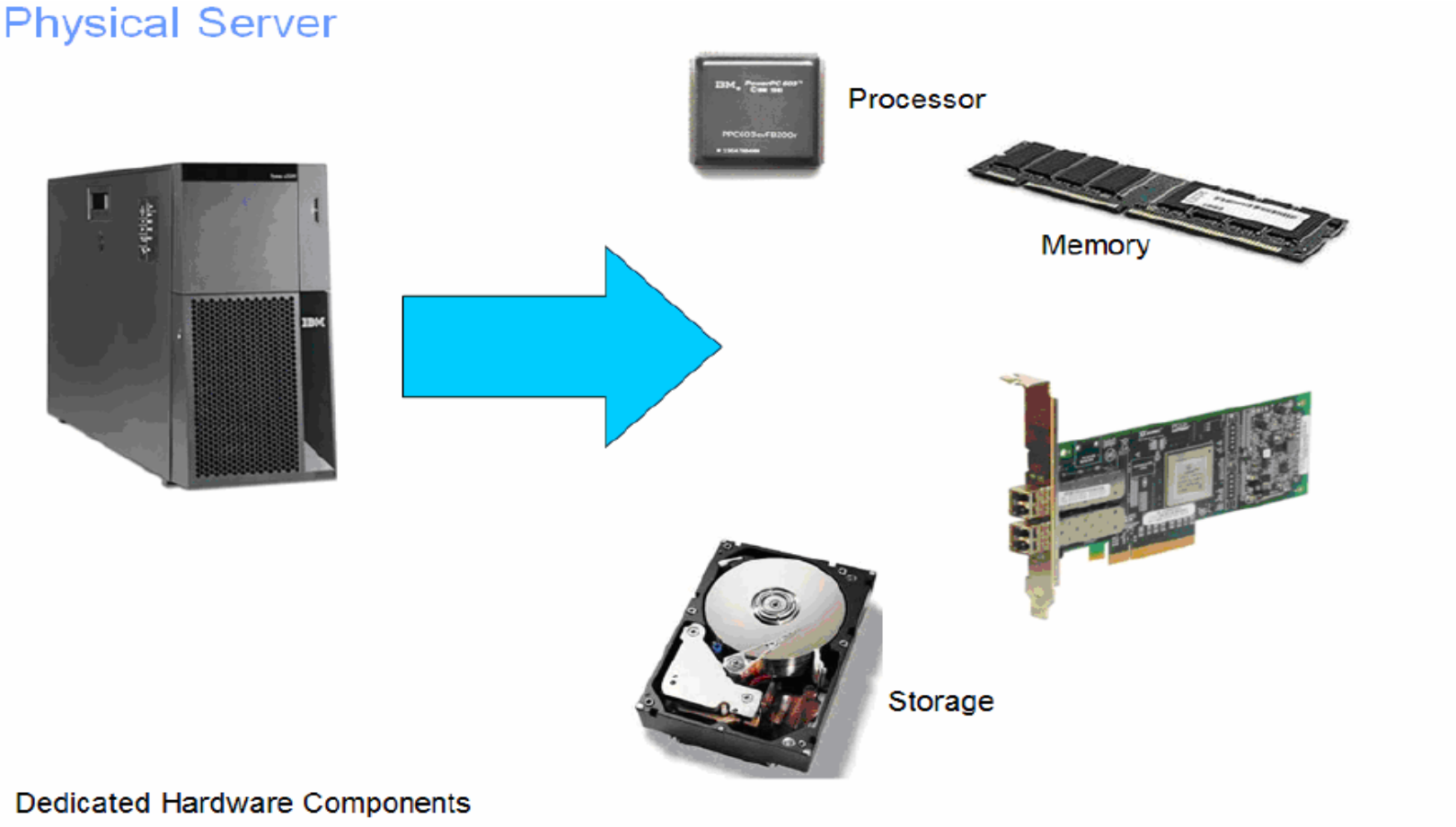
- **Security**:
- Each application runs in its own virtual machine, there is a strict isolation of system resources and sharing only happens in a controlled manner, thus allowing for higher security in the system
- **Manageability:**
- It is possible to move the virtual machine and abstracted system components around in the infrastructure based on load distribution and save a snapshot of virtual machine to be restored later.
- Patched VMs images can be maintained in a VM library and can be cloned and activated on demand.
- This saves a lot of time spent in system administration.

A tabular comparison of traditional IT infrastructures with virtualized infrastructures is as follows:

Parameter	Traditional IT	Virtualization
Utilization	0-20%	Typically 60-70%
Provisioning	Typically takes 6-8 weeks	1 day
Monitoring	Usage of monitoring tools. However, need manual intervention to take care of any hardware failures	Comparative ease in monitoring using automated tools. However, need manual intervention to take care of any failures
Sizing	Sizing needs to be completed before deployment. Re-sizing involves procuring new hardware and planned downtimes	Easier to resize. However, manual intervention required to resize
Staff for Administration	Require larger number of Full Time employees to manage the infrastructure	Reduced number of Full Time employees
Cost	Upfront costs involved in outright purchase of hardware	Initial hardware cost reduced due to sharing of hardware assets and increased utilization. There is a typical reduction of 40% in hardware
Optimization	Difficult to do as there is no easy way to monitor and load balance across machines	Easy to share resources and re-balance loads on the virtual machines on the same host. However, re-balancing across physical hosts require advanced features and planned downtime

Implementing Virtualization

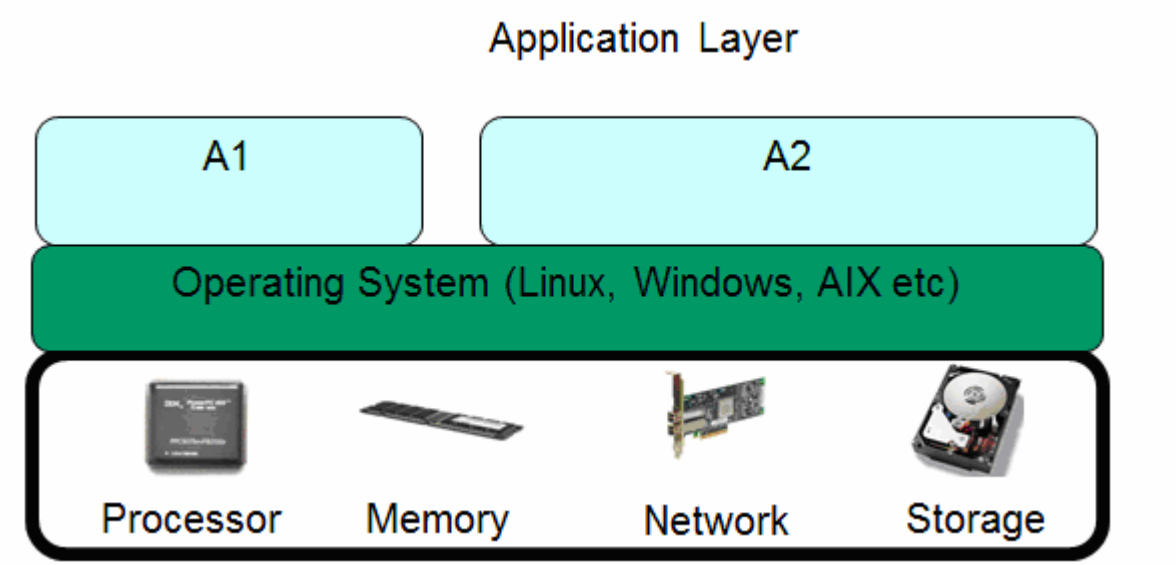
A physical server is typically composed of four major physical components –



Physical Components…….

- Four hardware components are required – CPU, Memory, Network and Storage.
- In addition, the application requires support from the Operating System and allied components.
- **View Figure (Previous Slide):**
- The components marked in black box are the hardware components.
- The Operating System resides over the hardware.
- The applications are stacked over the operating system and use services provided by the OS.

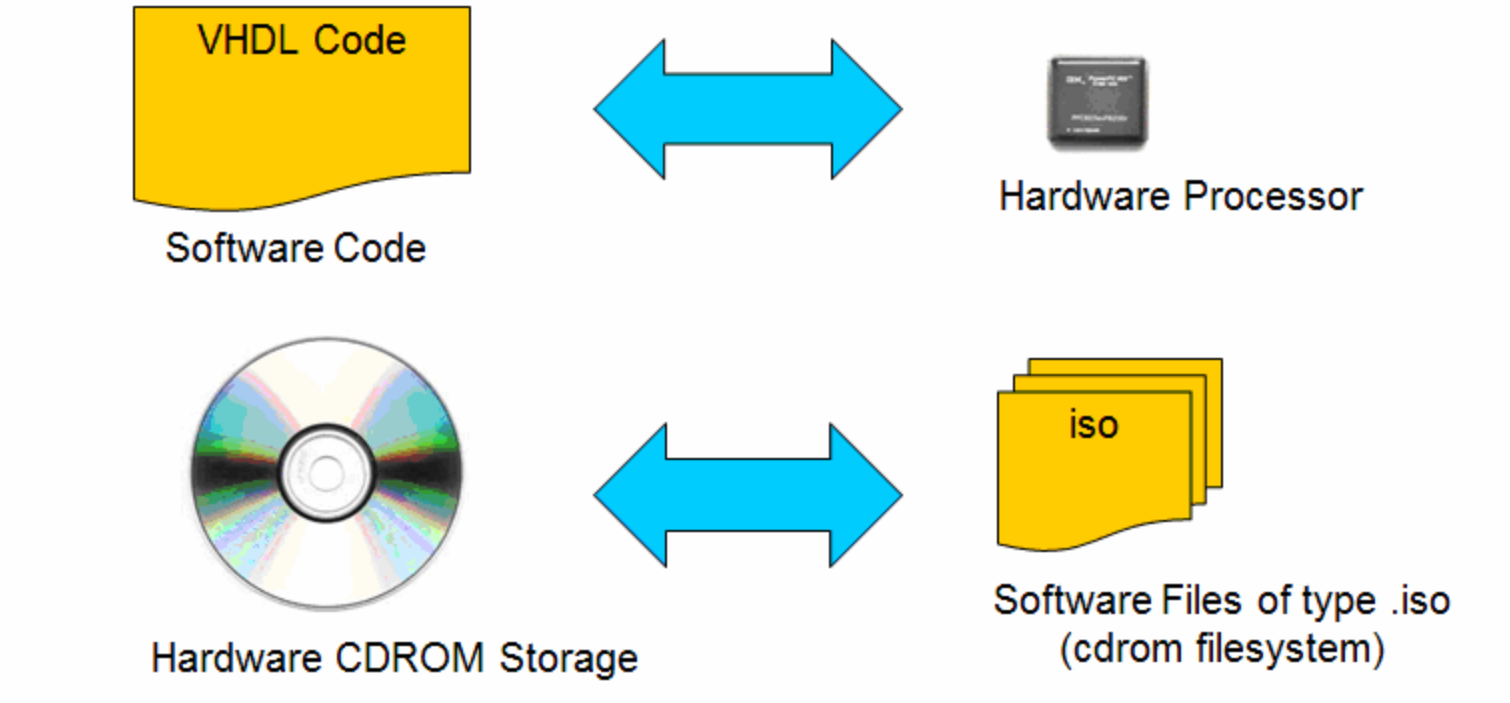
Typical Server Track



Introduction to Virtualization→ Principal of Hardware-Software Equivalence

- Any logic written in software can be easily converted to a hardware equivalent.
- Any existing hardware can be easily converted to software.

The examples shown below illustrate the logical equivalence:



Example elaborated…

- The VHDL code (software) is logically equivalent to a hardware processor.
- The VHDL code can be synthesized and fabricated into a physical processor.
- On the other hand, given a physical processor, it is possible to represent the logic function in an equivalent VHDL code.

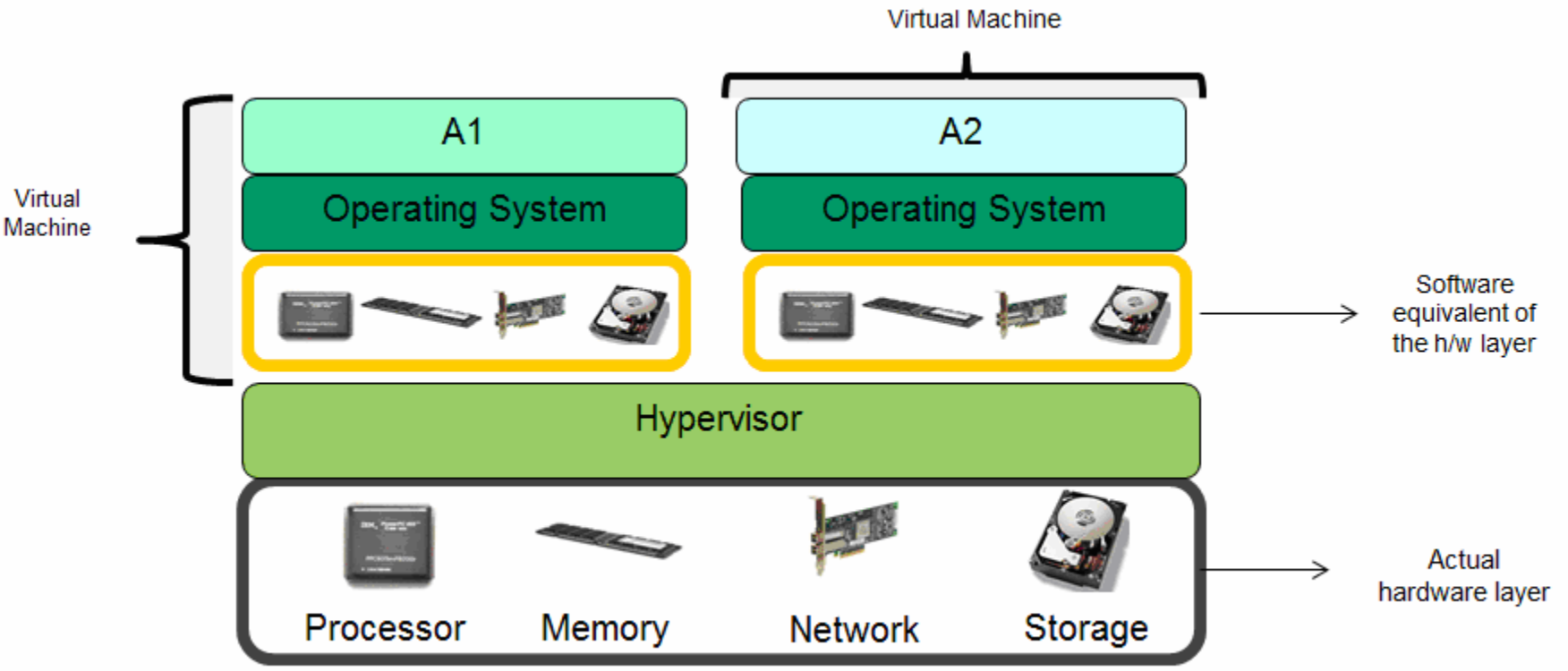
- A CDROM storage that can be converted into an iso9660 software image.
- It is also possible to convert/burn the image back to a CDROM.
- Logically these two are equivalent.

Contd…

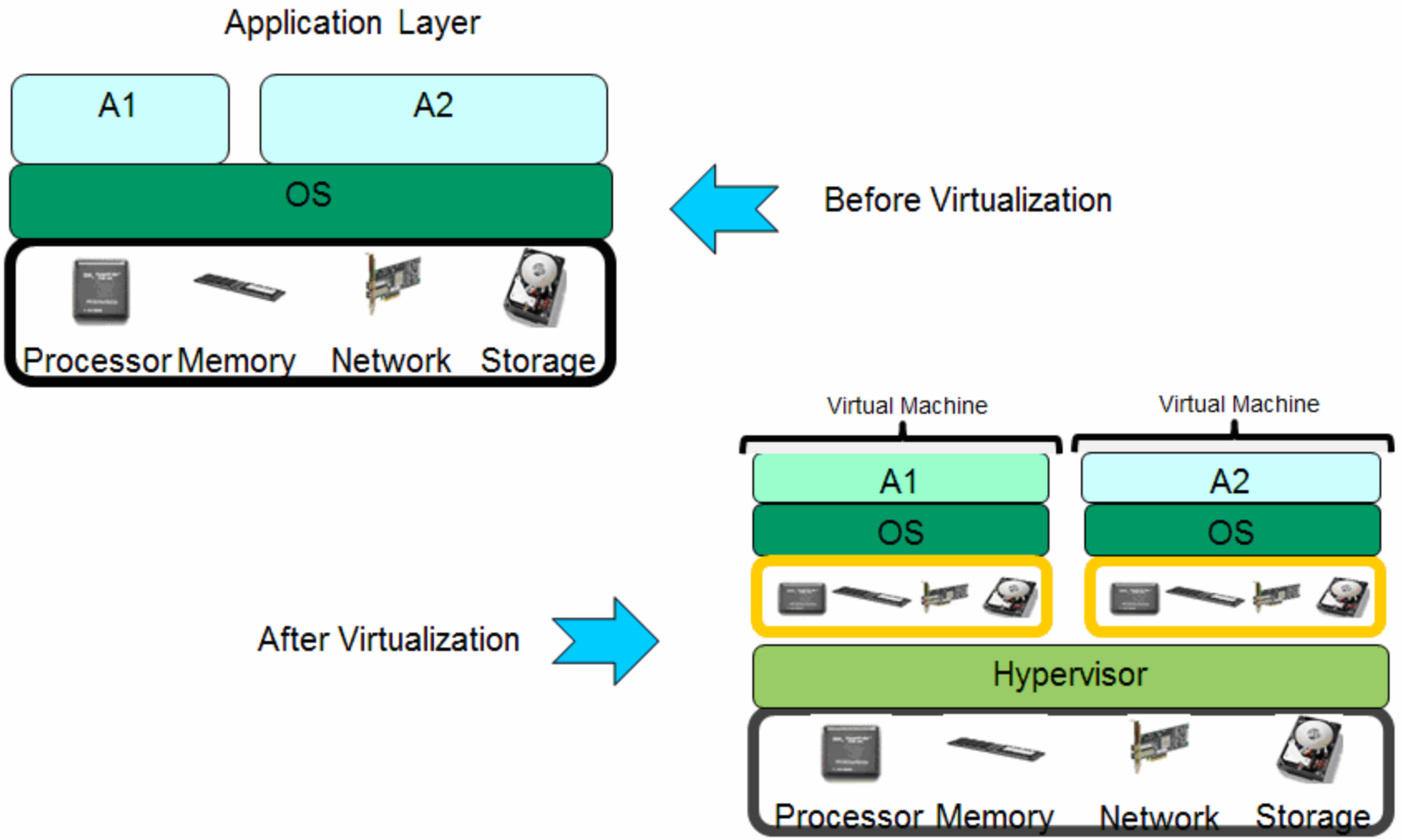
- The hardware-software logical conversion is generally a tradeoff between performance and
- flexibility.
- VHDL code is easier to modify and simulate than a physical processor.
- A physical processor once fabricated cannot be changed.
- However, the performance of simulation using VHDL is much slower than a physical processor.

Contd…

- Based on the principle of hardware-software equivalence, it is possible to logically convert the system hardware components required to run an OS/application, to their software equivalent or a virtual machine (orange box in the diagram).
- The virtual machine can then be replicated to create multiple virtual machines each running its own instance of the OS and the application.
- A hypervisor layer (marked in light-green) is required to multiplex the actual real hardware resources among the virtual machines.
- The hypervisor is responsible for allocating memory, CPU resources, network and storage to each virtual machine.


Virtual Machine
A1
A2
Operating System
Operating System
Virtual Machine
Software equivalent of the h/w layer
Hypervisor
Processor
Memory
Network
Storage
Actual hardware layer

Pre and Post Virtualization



History……

- Virtualization started out as a concept of **time-sharing** in the 1960s.
- The machines were large and there was **need to split up the machines into multiple portions that can be used for different purposes simultaneously.**
- This resulted in evolution of newer innovations related to **hardware sharing, paging techniques and multiprogramming**.

Advent of Time Sharing Systems

- Earlier systems in the 1960s allowed only **one user/job** at a time on the system.
- **Major disadvantage** to larger computers that were technically **powerful enough to execute multiple jobs** at a time, **but not capable due to hardware/software constraints** of the system.
- Simultaneous time-sharing of the system was seen as a solution to this.
- However, modifying legacy batch systems to accommodate multiple simultaneous users on the system made the operating system too complex to maintain and support.

Contd..

- The first supercomputer to take advantage of the concepts of shared physical hardware was the **Atlas Computer.**
- The computer was developed at **Manchester University**.
- **The computer allowed for separation of the supervisory (OS) components from the user components**.
- The supervisory code monitored and managed system resources (CPU, memory and IO).
- The supervisory component responded to special instructions that enabled it to provision and monitor the user computing environment.

Virtual Memory System of Atlas..

- A very **rudimentary virtual memory system** was also introduced by Atlas in the form of **one level store and paging** techniques.
- This created an abstraction of how memory gets used by the user programs and made managing memory less dependent on hardware.
- IBM at the time was looking for a time-sharing implementation on their Mainframe systems.
- The **IBM Engineering Team at Cambridge, Massachusetts** came up with a novel approach to solve the problem.
- They provided each user with a **virtual machine (VM)** with an operating system.

Virtual Machine…..

- The operating system running in the VM **still supported one user**, so it does not have to be complex.
- In fact the existing operating system could be used without any changes.
- The real ingenuity was in how the **VMs on the Mainframes multiplexed** these hardware resources among themselves.

IBM Mainframe Virtualization

- The earliest pioneer of modern virtualization technology was IBM.
- IBM invented virtualization more than 40 years ago.
- IBM started with virtualization in the 1960s with the M44/44X project.
- This was developed at the IBM Thomas J. Watson Research Center in Yorktown, NY.
- The foundation of this technology was an IBM 7044 (M44) scientific computer.

VMM

- The new operating system had a special component called the **VMM (Virtual Machine Monitor)** which was capable of running many virtual machines, with larger virtual memory running on virtual copies of the hardware.
- Each virtual machine ran an unmodified copy of the operating system with good performance.
- The virtual machine was capable of running UNIX and other operating systems.
- One of the unique features of the IBM virtualization technology was that **the virtualization is part of the system's firmware.**
- **The hypervisor sits in the firmware layer** and provides an integration between the hardware and the VMs running an OS.

IBM PowerVM Virtualization

- In addition to Mainframes, IBM currently provides virtualization on **Power Servers – the midrange UNIX systems.**
- The Power Servers are capable of advanced virtualization mechanisms some of the notable features are : **micro-partitioning, Advanced memory sharing, Live partition mobility, Virtual IO Server for IO Virtualization.**
- The first Power Server incorporating Advanced Power Virtualization (APV) was shipped in 2004.
- APV was rebranded to IBM PowerVM in 2008.

First Open-Source Hypervisor..

- Year 2003, marked the release of the **first open source hypervisor for x86 machines** called **Xen Hypervisor.**
- The company XenSource that developed the hypervisor was later acquired by Citrix.
- Citrix is currently one of the major virtualization solution provider in the x86 market.

Further developments

- In 2006/2007, Virtual Iron released Virtual-Iron, an x86 bare-metal hypervisor for enterprise customers.
- In the recent years, a new technique to provide virtualization on Linux/x86 has been developed.
- This project is called the “KVM” Project.
- The source code for kvm is part of the main linux kernel tree.
- KVM relies on the support from x86 hardware (including the Intel VT or AMD-V support) for better performance.
- In absence of hardware support, Qemu is used in emulating the required hardware components.
- KVM is now part of the standard linux kernel distributed along with a standard Linux distribution – Redhat, Novell, Ubuntu, Debian etc.

Hardware Support for x86 virtualization

- Early x86 processors had no built-in support for Virtualization.
- The virtualization was achieved using a software-only hypervisor using complex techniques to multiplex resources among virtual machines.
- The performance of these systems was reasonable but not suitable for putting in production.
- The virtual machines were primarily used in the test and development teams or in places where performance was not a qualifying criteria.
- **Hardware Assisted Virtualization** began to take shape from 2005, which implemented some of the commonly used functions into x86 hardware.

More…..

- Intel introduced Intel VT-x and AMD introduced AMD-V to support virtualization in hardware.
- Pentium 4 (Model 662 and 672) were the first Intel processors to support VT-x.
- AMD Athlon processors starting from Athlon 64 ("Orleans") supported AMD-V, the technology to support Virtualization in hardware.
- As time progressed, in addition to **hardware-assist for CPU virtualization, advanced hardware techniques to virtualize memory, and IO were introduced through different chipsets**.
- A basic principle behind these techniques was to provide a shadow copy of the hardware to each virtual machine and to have a bypass implemented in the hypervisor to access these copies directly.

Impact of Virtualization

- Making infrastructure assets more traceable, manageable and amenable to change.
- Decrease in associated cost to manage an IT infrastructure.
- Sustenance of growth of IT.

Two Major Headings: Cost & Manageability…

- **Cost Impact**
- Traditional physical infrastructures consume lot of power to keep IT running. However, a large percentage of the infrastructure remains unutilized for a myriad of reasons
- **Security/Segregation of critical applications**:
- The applications are segregated on different physical servers for security/criticality reasons.
- If only these applications can be closely packed on a single server we can bring up the utilization levels on that physical server. In most cases, a physical server runs with a utilization of no more than 20%, thus about 80% of capacity on the server remains unutilized.
- **Limited or No Monitoring:** Traditional physical systems are not monitored for performance or utilization on an ongoing basis. Monitoring/Audit of physical systems is generally done only on a periodic basis.

Contd…

- **Ineffective Asset Management:**
- Lack of automated tools to monitor removal/addition of physical components in the infrastructure.
- Even addition/removal of servers sometimes goes unrecorded when done on an adhoc basis.
- The asset list frequently goes out of date to make any conclusive assessment on the available capacity at a later date.
- **Provisioning Turnaround:**
- The time required to provision a server may range from a day to a month based on the request.
- The provisioning may require creating rack-space for new machines, provisioning for networking and storage, installation of OS and complex patch updates which take time ranging in days to months.
- In the process, valuable machine time gets wasted, that could have been used for running an application.
- All these reasons contribute to ineffective utilization of the physical servers thus resulting in higher cost to maintain and run a datacenter.

Contd….

- Virtualization on the other hand eliminates most of the drawbacks of the traditional IT infrastructure by building flexibility into the infrastructure.
- Since each application is running into its own virtual machine on a physical server, it becomes possible to **closely pack virtual machines thus bringing up the utilization.**
- Monitoring is relatively easier in virtualized environments since monitoring **could be done from the hypervisor with inbuilt statistics.**
- Typically, 60-70% of the infrastructure can be virtualized thus resulting in reduction of about 60-70% of physical servers. This enables **better asset management and monitoring.**

Contd…

- Provisioning with virtualization is faster as it is now possible to capture and store existing configuration of the virtual **machine in the form of images and image libraries**.
- These images can then be restored back to a fully functional virtual machine on demand, without repeating the OS and application configuration.
- A virtualized environment can also be **maintained by reduced staff** due to the ease of managing the reduced physical footprint.
- Virtualization saves cost on three fronts: **People, Process and Technology.**

Manageability Impact

- A virtualized infrastructure brings more visibility into the working of an IT datacenter.
- It is possible to track servers on an utilization basis, move workload around to load balance machines in real-time, save on power and cooling.
- An important aspect of virtualization is the visibility it brings to the IT infrastructure by enabling better manageability in the system.
- An improved visibility helps to plan the infrastructure better, avert any disasters, optimize and fine-tune backups, better accounting to plan for new hardware requirements and overall growth.
- Virtualization is a vast field.
- It is now possible to virtualize all aspects of an IT infrastructure ranging from Desktops to Enterprise Servers.
- The decision to virtualize is solely based on the customer requirements.

Some of the areas where virtualization is not recommended:

- Legacy software/systems that are not designed for virtualization
- Resource Intensive Applications that make assumptions on specific system characteristics to operate and perform (Proprietary Applications)
- Real-Time Applications where the timings are critical.
- We place these applications as close to the hardware as possible to guarantee turnaround times.
- Other large applic.ations such as databases