# Cloud Computing

Professor Yeali S. Sun
(孫雅麗)
National Taiwan University
Spring 2011

# 有夢最美、希望相隨 … ☺


amazon
web services™
EC2
Google
Microsoft
資安雲
教育雲
NTU Cloud
觔斗雲


2

# Cloud Computing：Concept (概念)、Technology (科技)、Value (價值)

- Evolution vs. Innovation
- Paradigm
  - new, users, providers, management
- Business model
  - Feasible, sustainable, profitable

# Cloud Computing: Value Creation

- Vendors
- Enterprises (business process, customers)
- *Entrepreneurship (**imagination, innovation, realization …**)*

- Don't be just cost reduction but

  *CREATE VALUES*!

  – Cloud computing is not just a collection of servers/platform/software for cost reduction.
  – Value creation is important and it is *business-specific*.

# Challenges

- Free from fear (trust, security, … )
- Free from worry (full of opportunity … IT Service as a Service)
- Only new (needed) technology ⇔ service/paradigm/business model *innovations* can continue driving the sector grow and prosper. *-> Opportunities!*

# Outline

- What is Cloud Computing?
  - Compare cloud with conventional computing
  - The five essential characteristics of Cloud Computing
- What are the top technical and non-technical obstacles and opportunities of Cloud Computing?

# What is Cloud Computing?

# Why is Cloud Computing so hot？

- 政府四大智慧型產業 之一
- US General Service Administration, GSA announced that government agencies would be able to access cloud service offerings through apps.gov

- …

# Cloud Computing: Better, faster, cheaper

- The definition of cloud computing is quite murky.
  - In 2009, *The Cloud Computing Journal* published an article that had 21 definitions of cloud computing. (Twenty-One Experts Define Cloud Computing)
  - 「人雲亦雲」、「眾說紛雲」、「各自表述」
- In 2011, a common goal is crystal clear.
  - To provide an order of magnitude improvement in the **cost effective**, **dynamic provisioning** of IT services.
  - "better, faster, cheaper“
- Challenge: Whether if cloud computing solutions are better in higher quality is a subject of future demonstration.

# Cloud Computing: technology perspective

- Cloud Computing aims to deliver **on demand IT resources on a pay per use basis.**
- Cloud Computing aims to be **global** and to provide such services to the **masses**, e.g.,
  - end user that hosts its personal documents on the Internet;
  - enterprises outsourcing their entire IT infrastructure to external data centers.

# Terminology

- **The hardware** – Infrastructure as a Service (IaaS)
- **Systems software, library/utility** – Platform as a Service (PaaS)
- **The services** – Software as a Service (SaaS).

# What is Cloud Computing?

- 美國政府「國家標準與技術研究院」（National Institute of Standards and Technology,（NIST)）

「雲端運算的概念是提供使用者一個**方便、可隨時利用網路界接**到**一大群共享的運算資源**（包含網路、伺服器、儲存設備、應用軟體與服務）。這些資源可以按照使用者的**需求**，很快地以**最小的管理成本**提供給需求者使用。」

# Cloud Computing: a Model

NIST

Cloud computing is a **model** for enabling **convenient**, **on-demand network access to a shared pool of configurable computing resources** (e.g., networks, servers, storage, applications, and services) that can be **rapidly provisioned** and released with minimal management effort or service provider interaction.

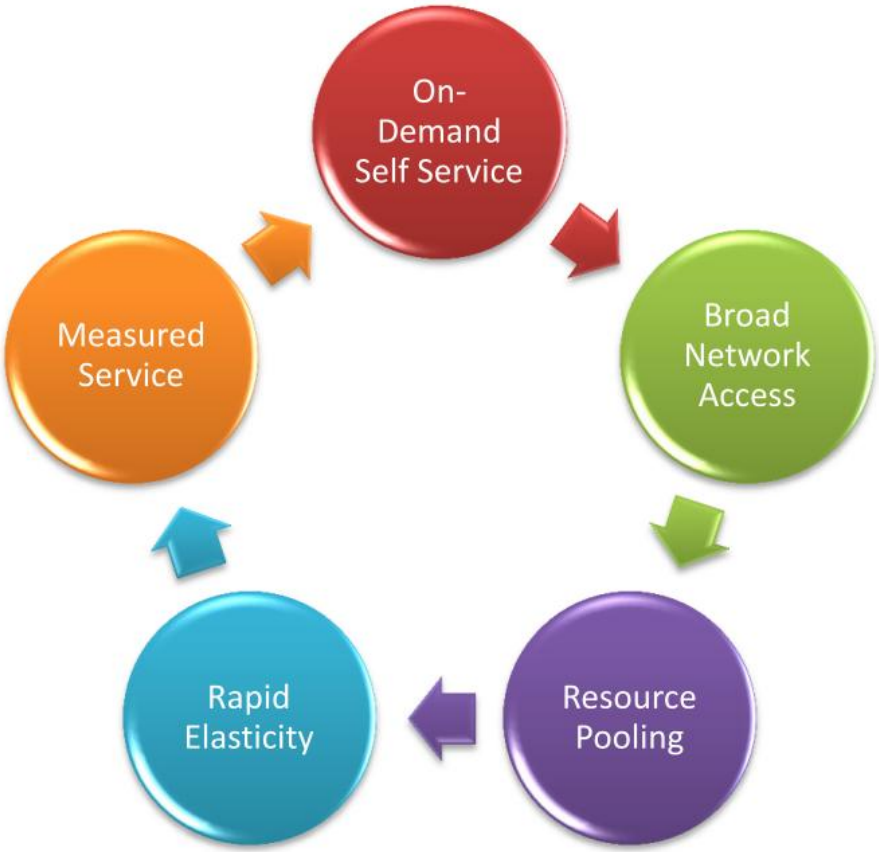
This cloud model promotes availability and is composed of

- ✓ **five** essential **characteristics** (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured Service);
- ✓ three service models (Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), Cloud Infrastructure as a Service (IaaS)); and,
- ✓ four deployment models (Private cloud, Community cloud, Public cloud, Hybrid cloud).

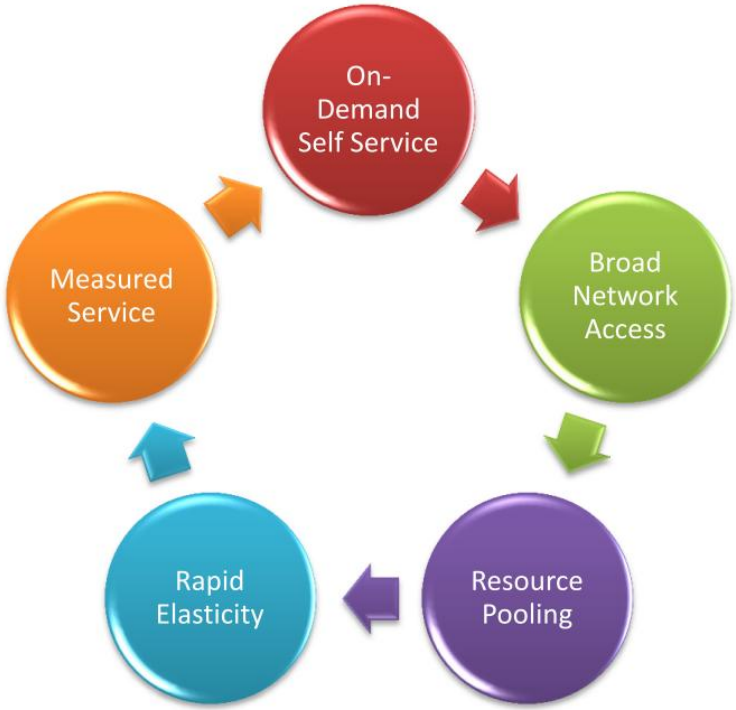
Key enabling technologies include: (1) fast wide-area networks, (2) powerful, inexpensive server computers, and (3) high-performance virtualization for commodity hardware.

# Cloud Computing: Five essential characteristics

- *On-demand self-service*
- *Broad network access*
- *Resource pooling*
- *Rapid elasticity or scaling*
- *Measured Service, pay-as-you-go*



# *On-demand Self-service*

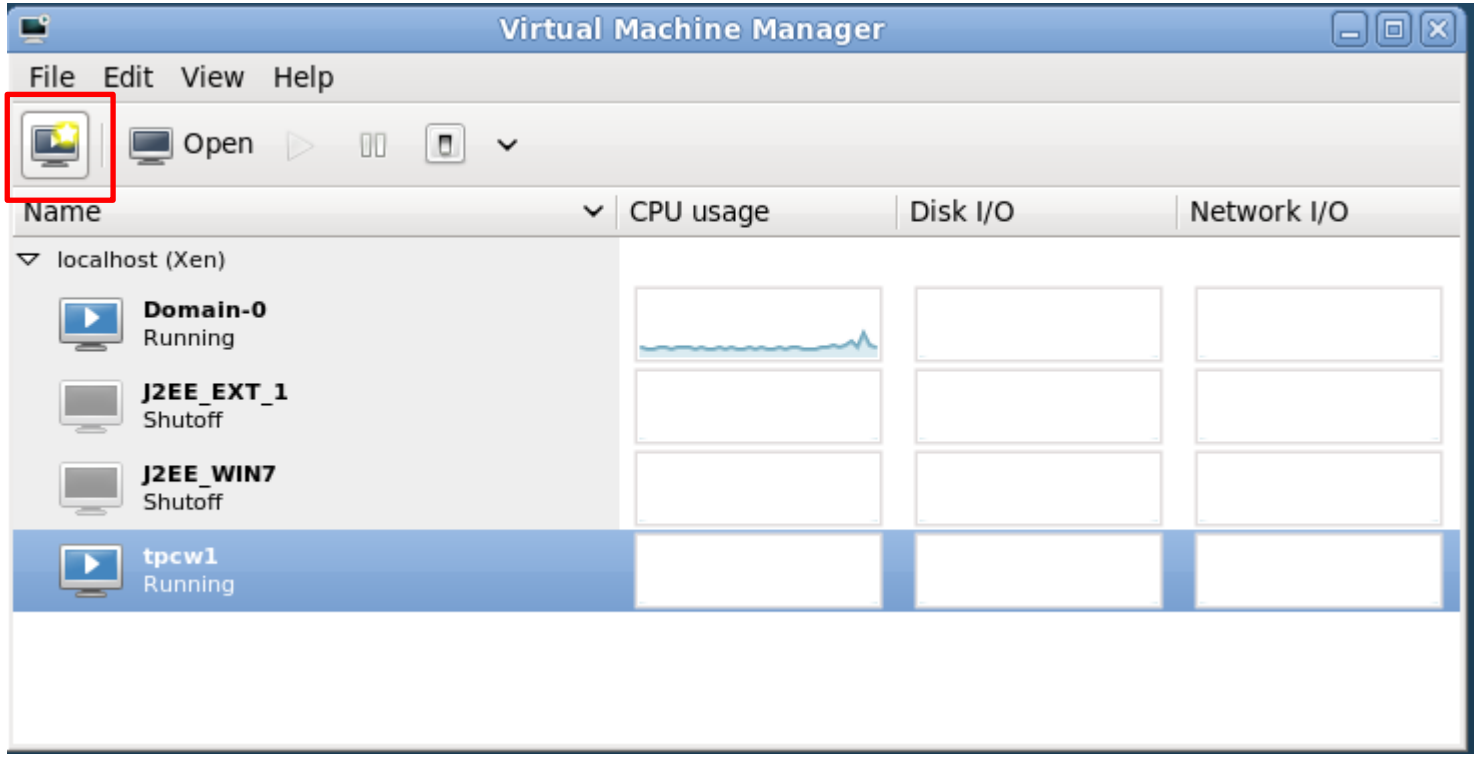


# NTU Cloud VM Service

- Demo:
  - how a customer can use online web-based self service to rent a computer (or virtual machine)
- Example:
  - a real-time on-demand VM provision using Xen 4.0

# Create a Virtual Machine (VM) (1/7)

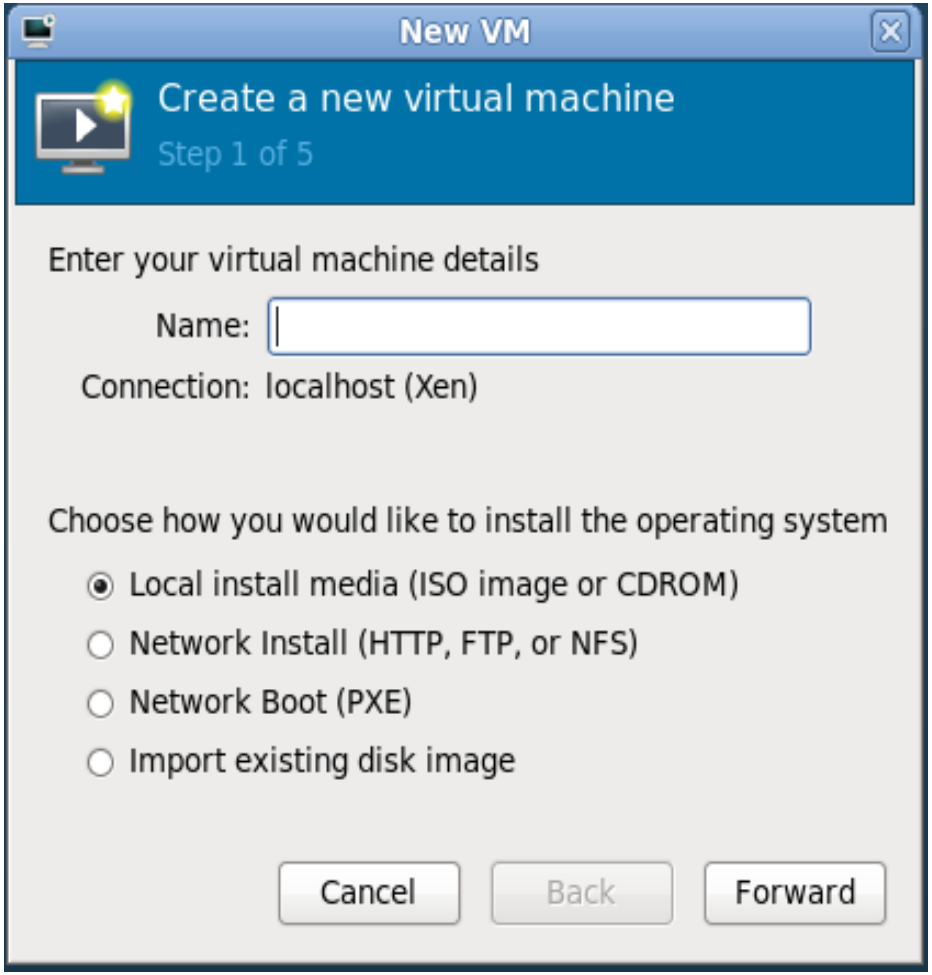
- Xen 4.0
- 打開虛擬機器管理員
- 點擊該圖示便可建立一台新的VM



# Create a VM (2/7)

**步驟 1. 輸入虛擬機的名稱及選擇安裝方式**

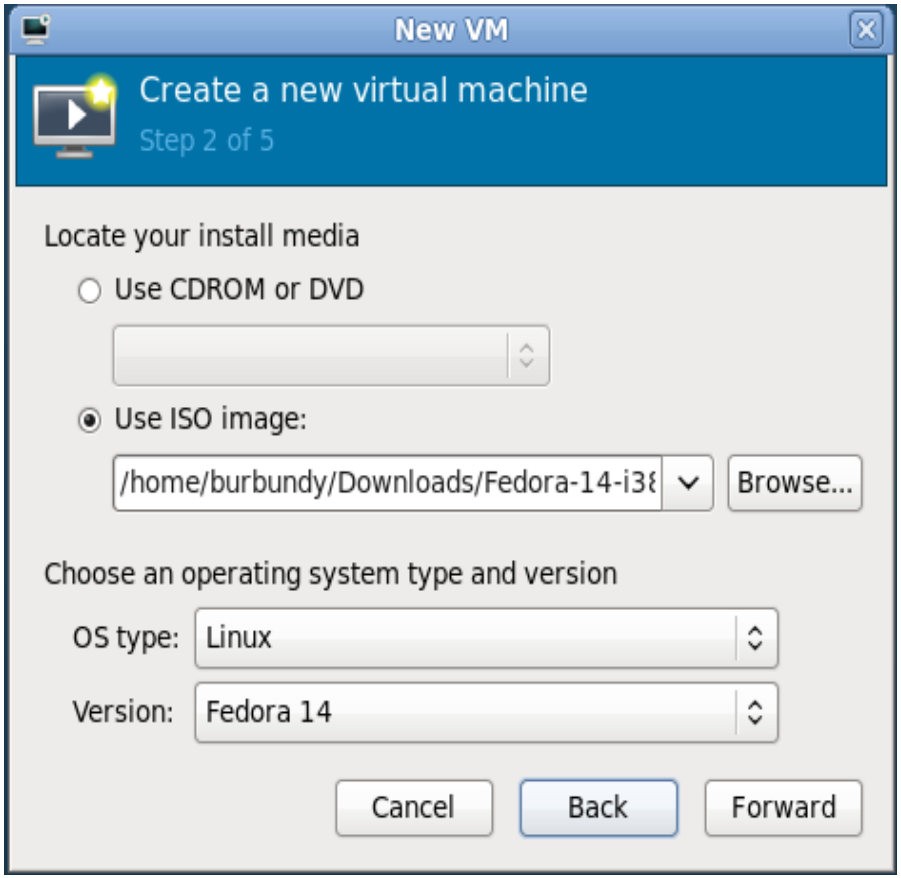
- 四種安裝方式
  - 透過本機的ISO檔或實體光碟
  - 透過網路安裝
  - 透過PXE網路開機安裝 http://www.webopedia.com/TERM/P/PXE.html
  - 直接匯入既有的image檔



# Create a VM (3/7)

## 步驟2. 設定安裝來源

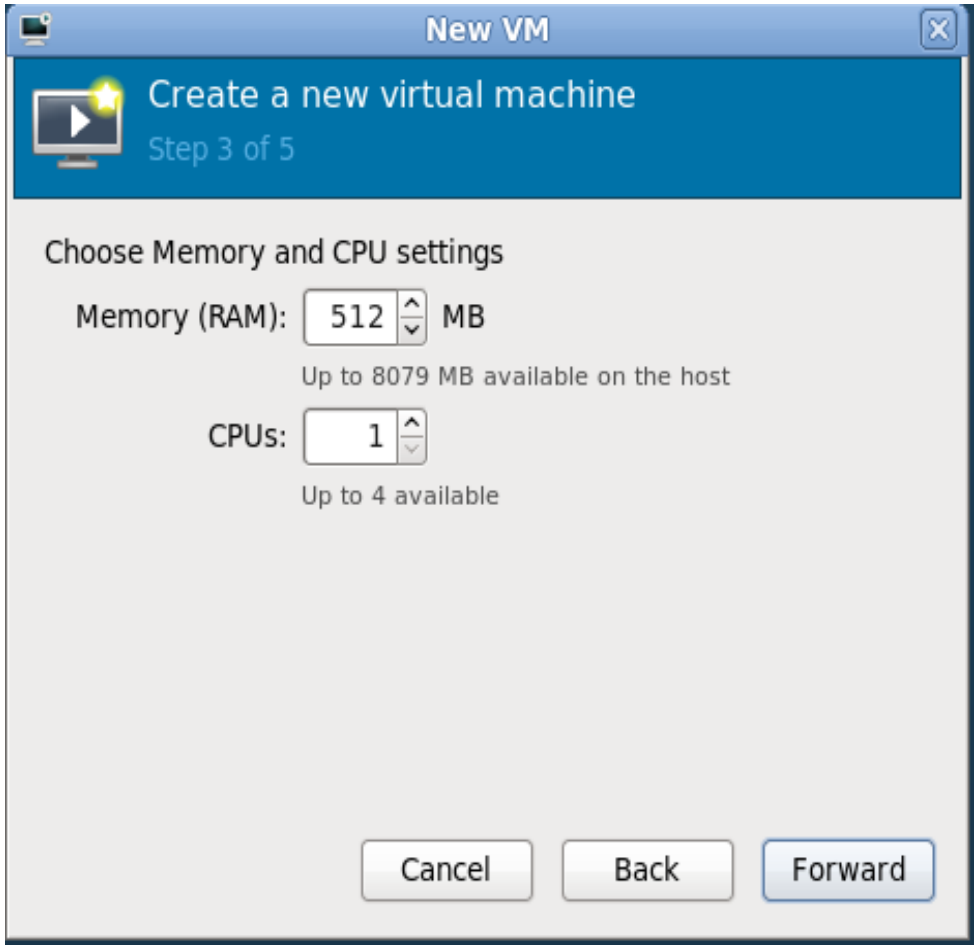
- Example: 本機iso檔



# Create a VM (4/7)

**步驟3. 設定Memory大小及CPU (Core) 數量**

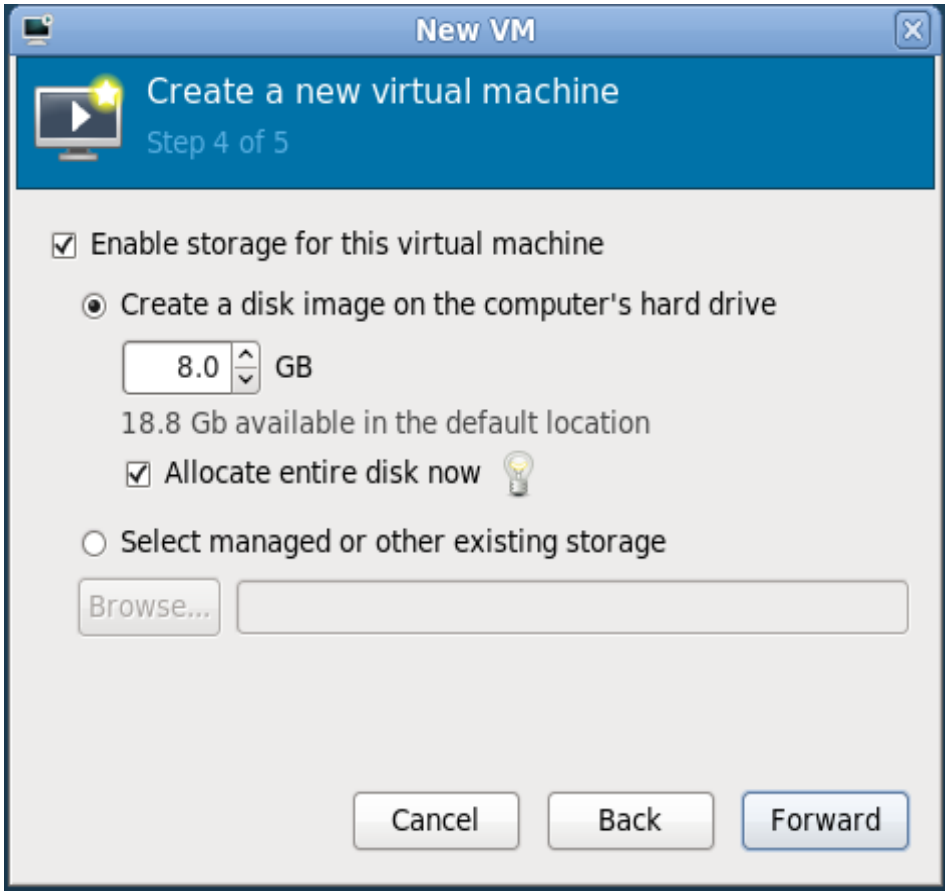
- 根據官方公布 Xen 4.0可以支援多達128個 physical core 的 Host machine, memory最多可以支援到1TB
- 而單一台VM最多也可以分配到128個VCPU及1TB的RAM.



# Create a VM (5/7)

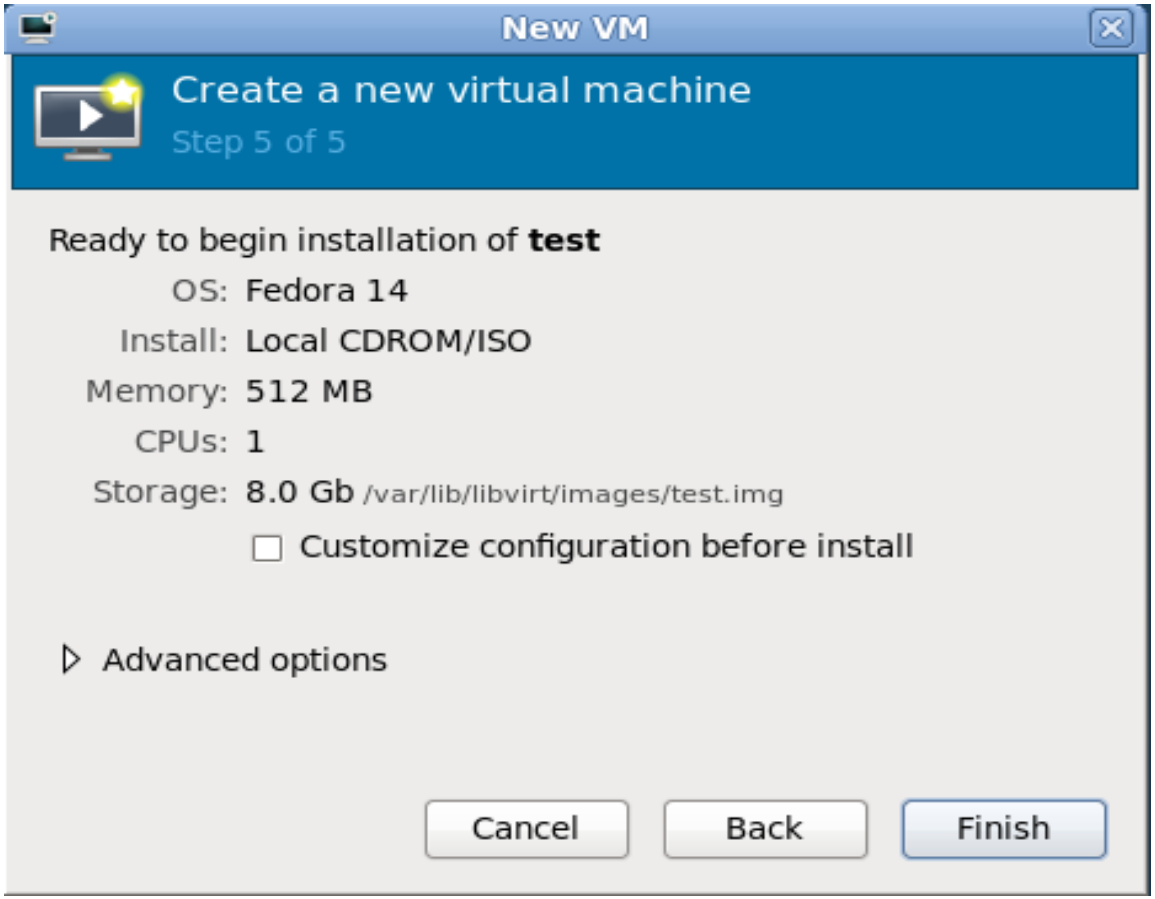
**步驟4. 設定該虛擬主機的虛擬硬碟硬碟位置及大小**

- 18.8Gb是提醒使用者在預設路徑下的硬碟空間剩下多少可以用來作為虛擬硬碟
- Allocate entire disk now 如果不勾選將不會限制硬碟大小(可能會造成空間不足的問題)



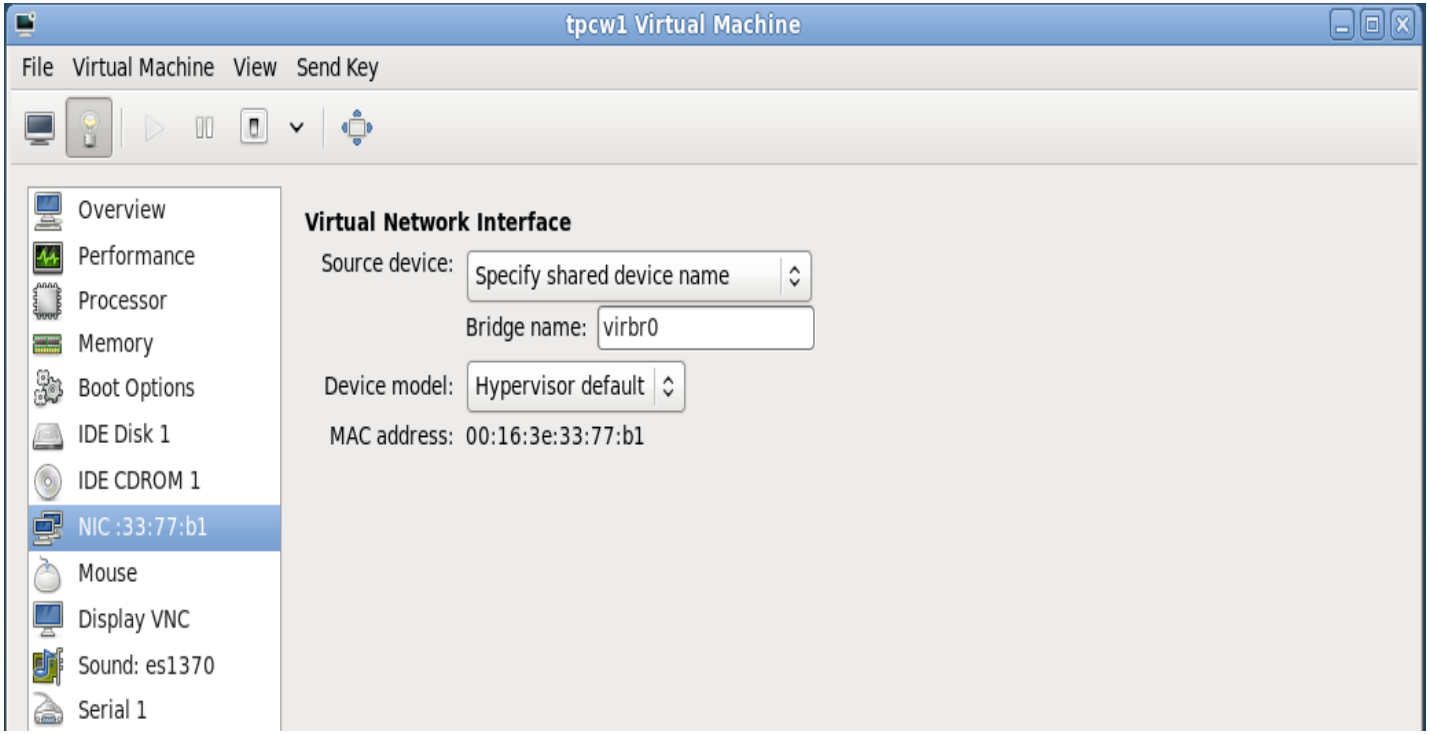
# Create a VM (6/7)

Done!
(in one minute)



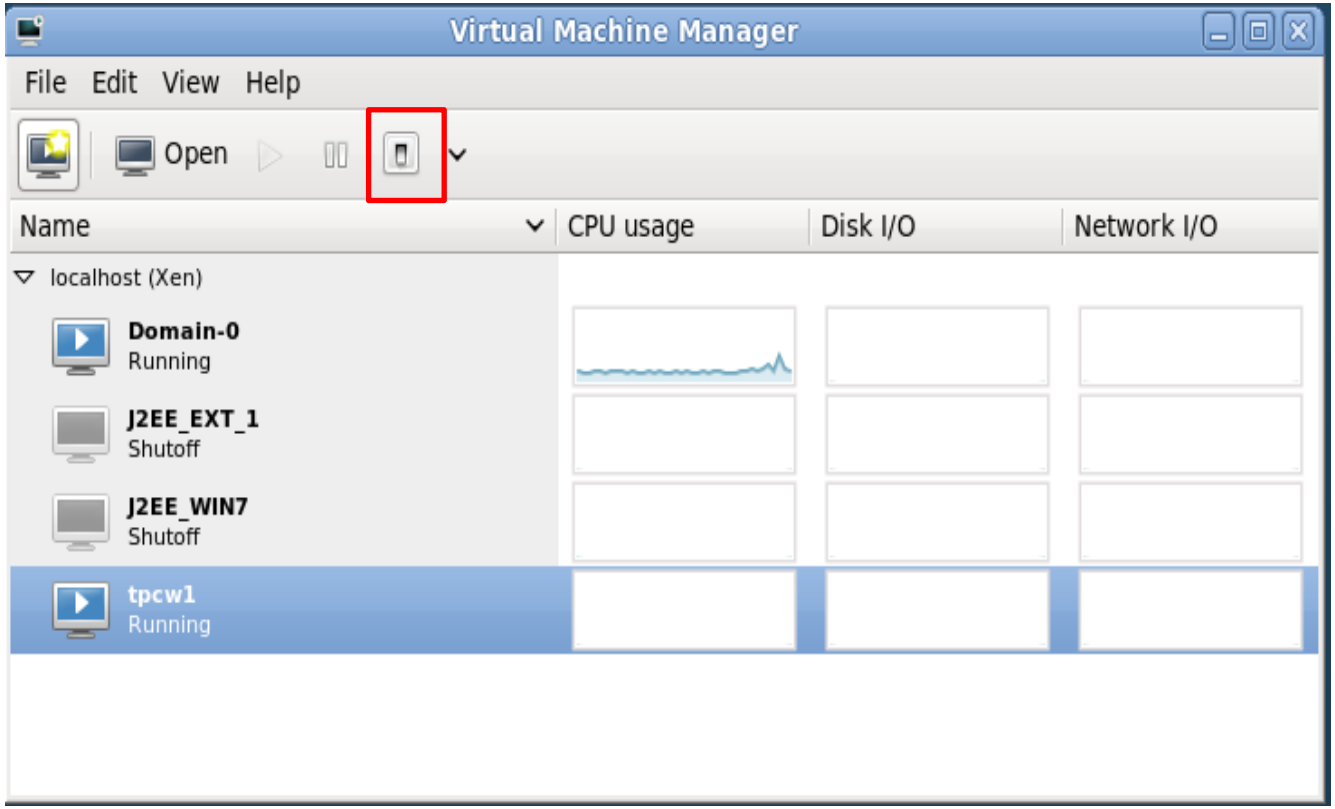
# Create a VM (7/7)

- Configure virtual network interface
  - Default MAC address assigned by hypervisor
  - IP address configured by guest OS

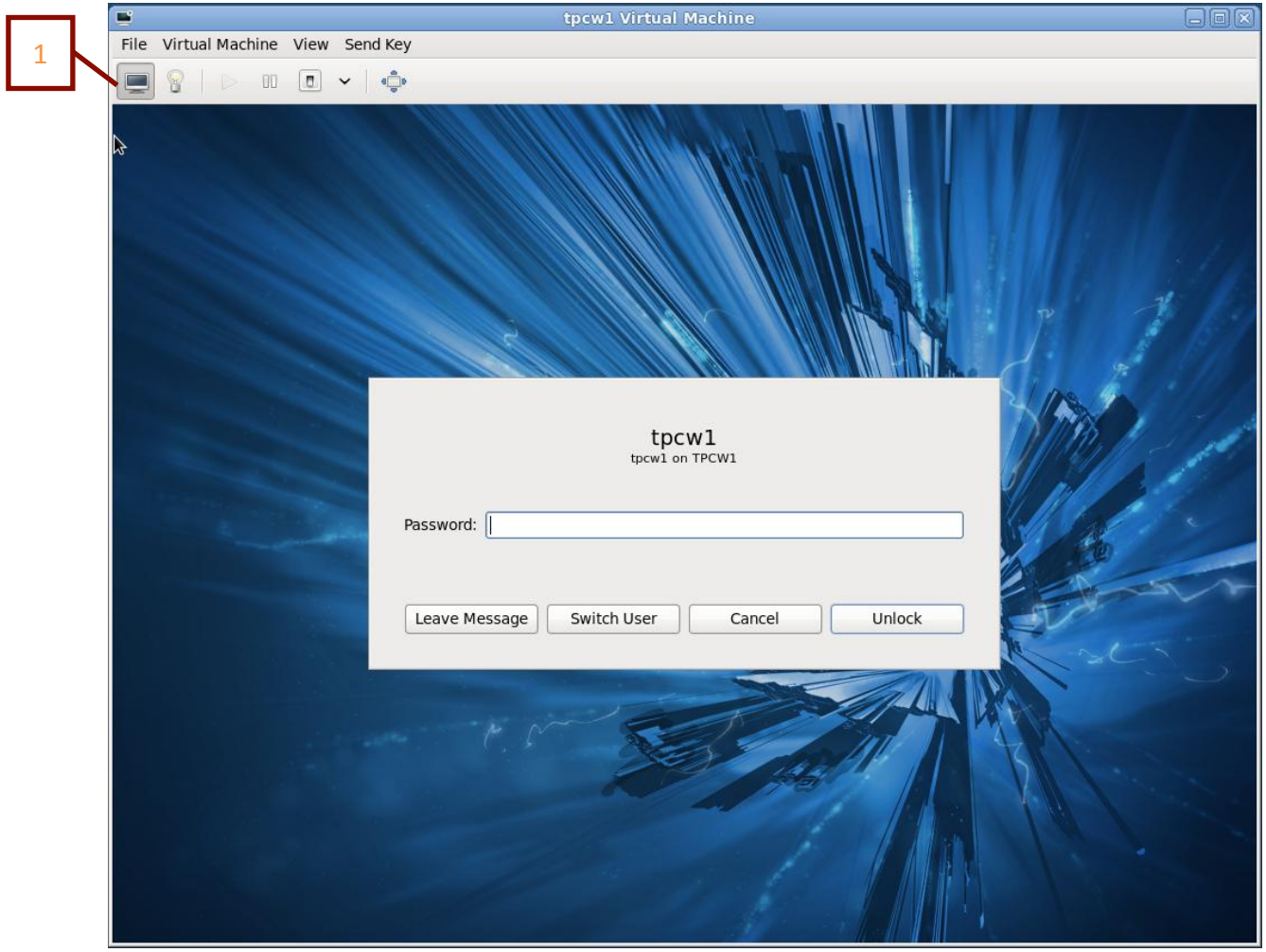


# VM Management

- 點擊想要操作的VM
- 開啟VM的操作畫面
- 開機、執行、暫停、關機、強制關機、保存等動作



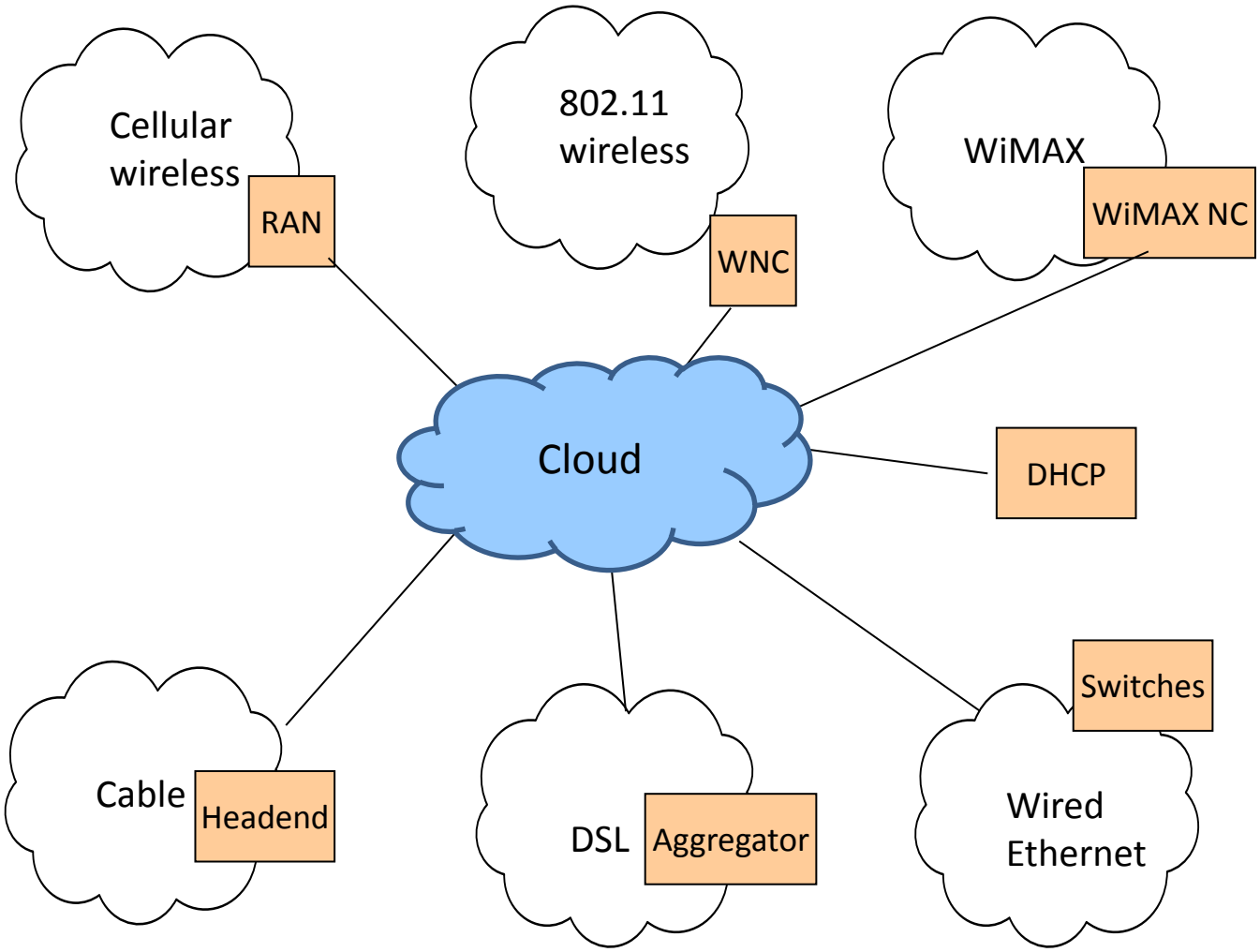
# Activate a Virtual Machine



# *Broad Network Access*

# Networking Challenges in Cloud

- Network structure
- Access network, core network
  - Bandwidth
  - Response time
  - WAN optimization



# What's the Internet: basics

Internet: interconnection of networks

PC

server

wireless laptop

cellular handheld

- millions of connected computing devices: *hosts = end systems*
  - running *network apps*

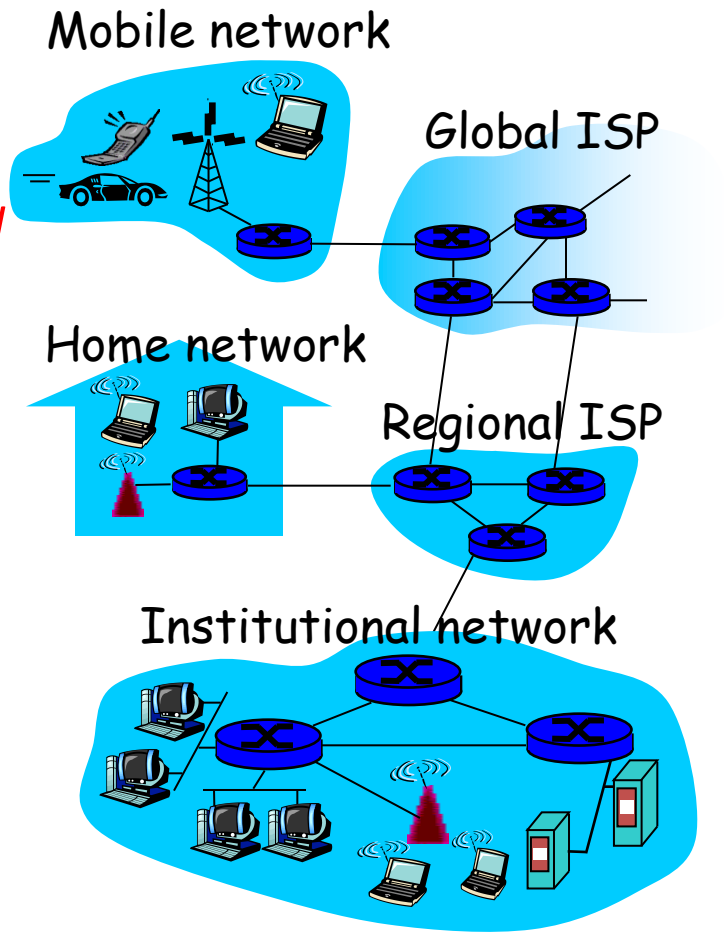
access points

wired links

- *communication links*
  - fiber, copper, radio, satellite
  - transmission rate = *bandwidth (bps)*

router

- *routers:* forward packets (chunks of data)



# *Resource Pooling*

# Resource Pooling and Sharing

- 雲端服務的一個重要關鍵目標是經由「資源整合」與「有效利用」，提供使用者一個更方便、經濟的運算、儲存、應用與內容資源的服務。
- 經由「資源共享、共用」、「需求多少、使用多少」的方式，達到「節省能源」、「節省成本」的目的。
- What value is resource sharing brought to the user?
  - Lower cost through resource sharing
  - The number of physical servers that would be needed with and without virtualization is reduced
- An abstraction or illusion of *infinite* pool of resources
- *Scalable* resource acquisition/performance (*real-time, automatic, on-demand, adaptive, pay-as-you-go, pay-as-you-grow!*)
- Any time, any where, any device

# Types of Resources

- Computing
- Memory
- Disk
- Network

# Bioinformatics Application

- Requires lots of memory because Gene computation.

# Resource Pooling and Sharing

- Complicated mechanisms and techniques are necessary to manage resources among different groups of needs and demands.

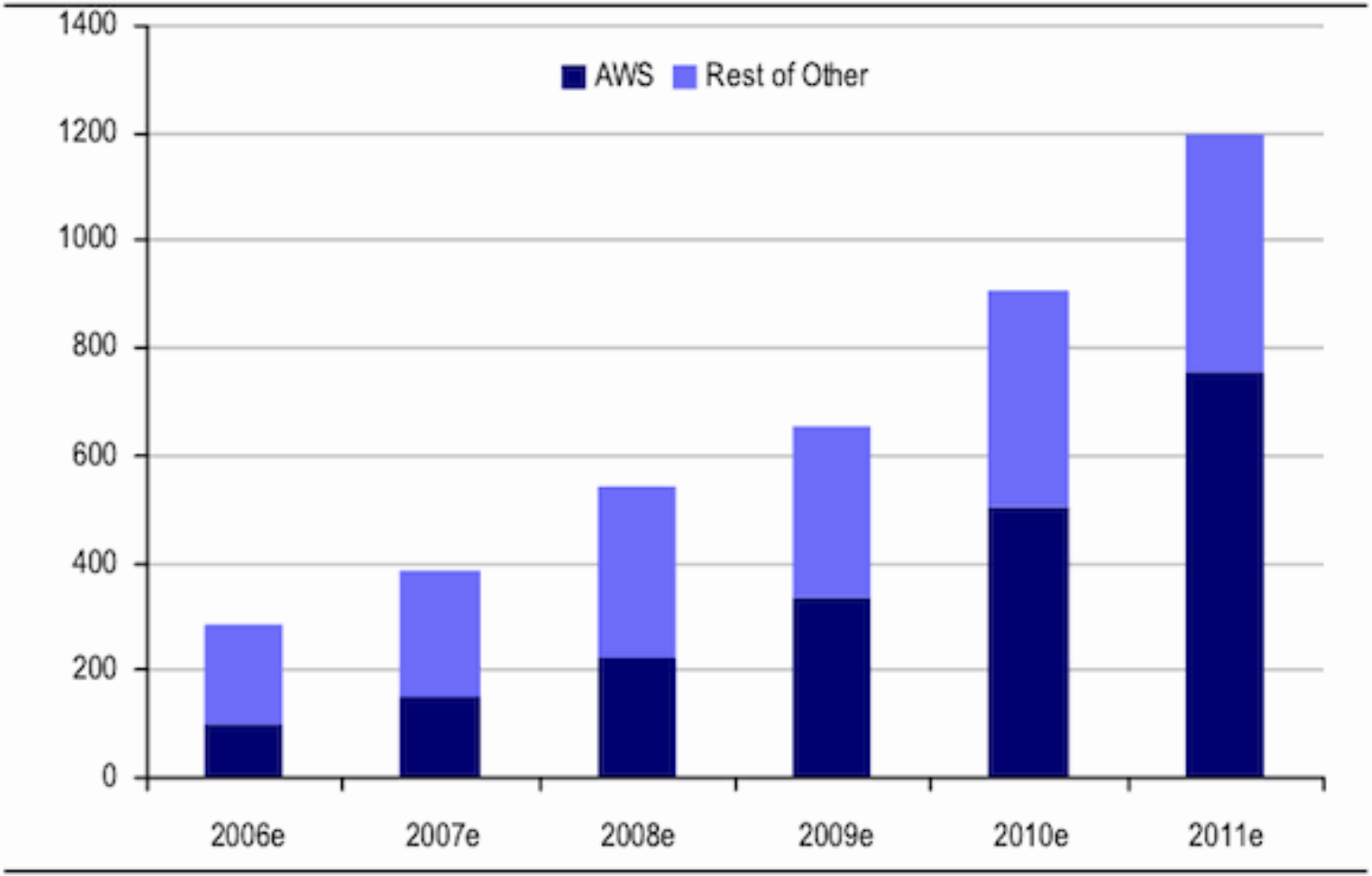
- What value is resource sharing brought to the user?
  - **Lower cost** through resource sharing
  - The number of physical servers that would be needed with and without virtualization is **reduced**

- The hero phase (James Staten, a vice president and principal analyst for Forrester Research)
  - because IT can reduce the time and cost of adding new resources!

# How Amazon got into AWS – cloud computing business

- AWS (Amazon Web Services)
- Initially it was to **better utilize** its expensive storage and computing hardware.
- It ended up turning it into a billable service.
- It also opened up *new vistas for entrepreneurs* and *made corporations rethink how they build and use their computing and storage resources*.

## Chart 3: 2006-2011 AWS Revenue Model ($MM)




Source: UBS

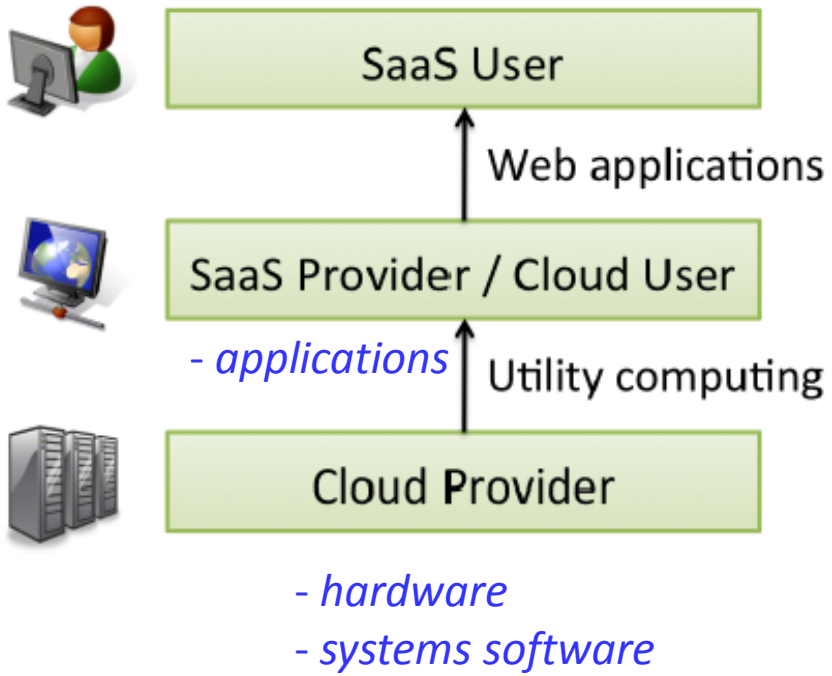
# It is more than renting out IT infrastructure …

- AWS essentially ‘rents out’ IT infrastructure to companies that seek to outsource IT needs such as
  - Application Hosting, Web Hosting, High Performance Computing, Storage, E-Commerce, and more.

# *Rapid Elasticity*

# UC Berkeley's views

- Cloud Computing - the long-held dream of computing as a **utility**
- Cloud Computing refers to both the *applications delivered as* ***services*** *over the Internet* and the *hardware and systems software that provide those services*.
  - The services themselves have been referred to as Software as a Service (SaaS).
- The datacenter hardware and software is called a *Cloud*.
  - IaaS and PaaS
- The service being sold is Utility Computing.




• Cloud Computing is the sum of SaaS and Utility Computing.
• Focus on SaaS Providers (Cloud Users) and Cloud Providers.

# From hardware point of view: three new aspects

- The illusion of infinite computing resources available on demand
  - **No** need for Cloud users to plan far ahead for provisioning.
- The elimination of an up-front commitment by Cloud users
  - allow to start small and increase hardware resources **only when an increase in need**.
- The ability to pay for use of computing resources on a **short-term** basis as needed
  - e.g., processors by the hour and storage by the day

# 雲端服務的議題:

## Service Level Agreement

- 對使用者而言，合約必須書名所承諾的服務品質(Quality of Service)，服務提供者必須要能依據合約確實履行.
  - 以 IaaS為例，服務品質包含了CPU、記憶體與儲存空間…等IT資源的使用
  - 過去AT&T提供專線連線服務，會提供使用者一台設備即時監控，讓使用者放心所購買的連線服務在連線頻寬與可靠性 (reliability)、穩定性有確實達到。

- 在雲端服務資源共享、共用的架構下，雲端服務是否如所簽訂的SLA中的規範被執行，讓使用者可以安心的使用雲端服務是很重要。

# 雲端服務的議題:

## 效能監測與資源分配管理

- 如何將SLA的內容對應到雲端系統內部的資源，並且對資源進行排程管理以符合SLA的需求
- 提供使用者可以監控其所租賃的雲端服務內容的機制，讓使用者可以隨時瞭解服務內容品質
- 當雲端服務需要更高的資源需求時，系統可以自動的即時反應其增加的需求(scale up)；或是不需的多餘的資源也能適當地減少(scale down)，即具有可擴充服務的能力(scalability)
- 可規模化的資源使用監控、分配、控制與管理機制，對CSP而言非常關鍵的資源控制管理工具。

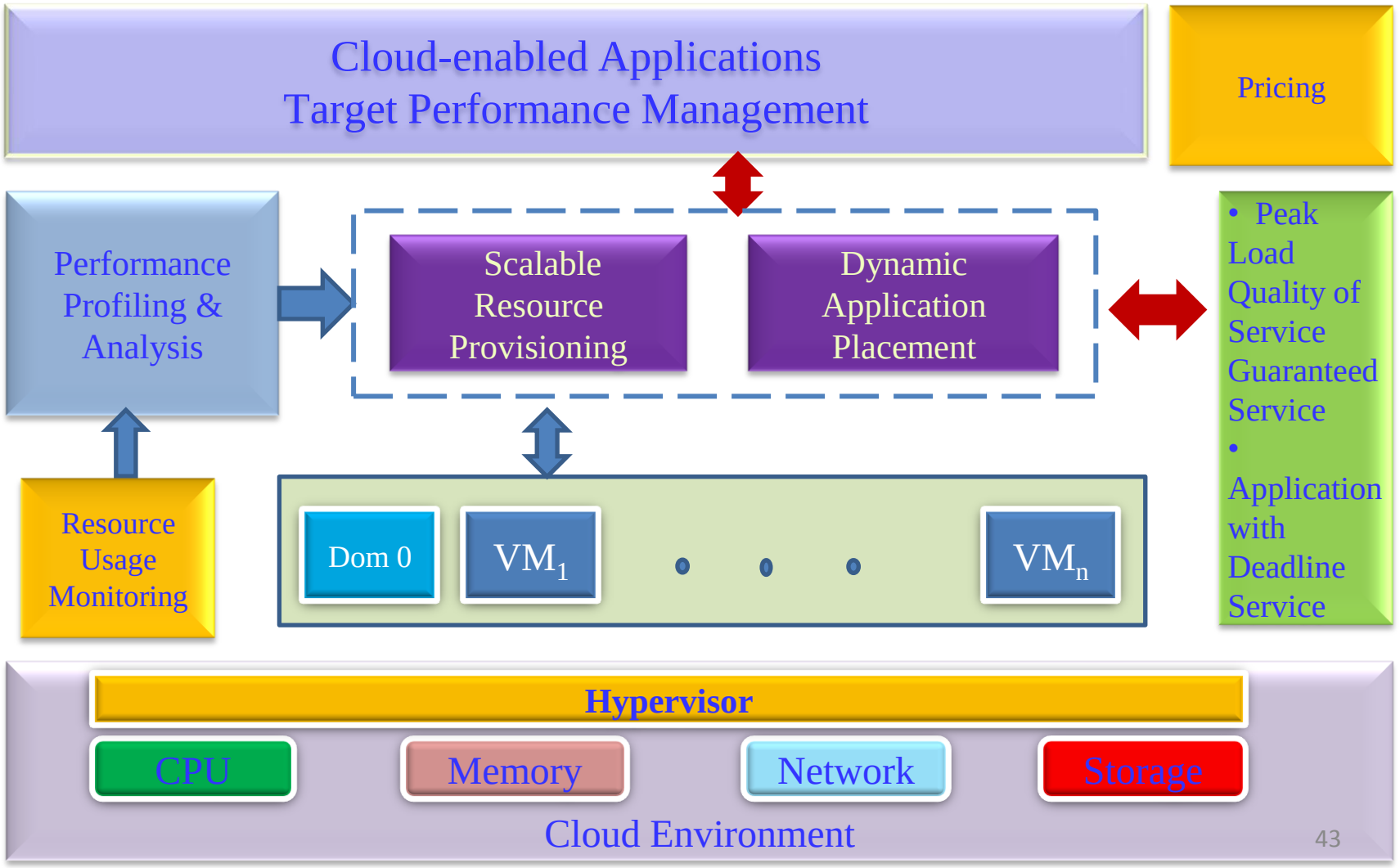
✓ 目的是確保雲端服務具有**穩定**與**即時**的擴充性與效能的保證。

✓ Stability and agility

# *Measured Service*

- Metering
- Billing
- Pricing

# Measured Services



# Cloud Computing: a Model

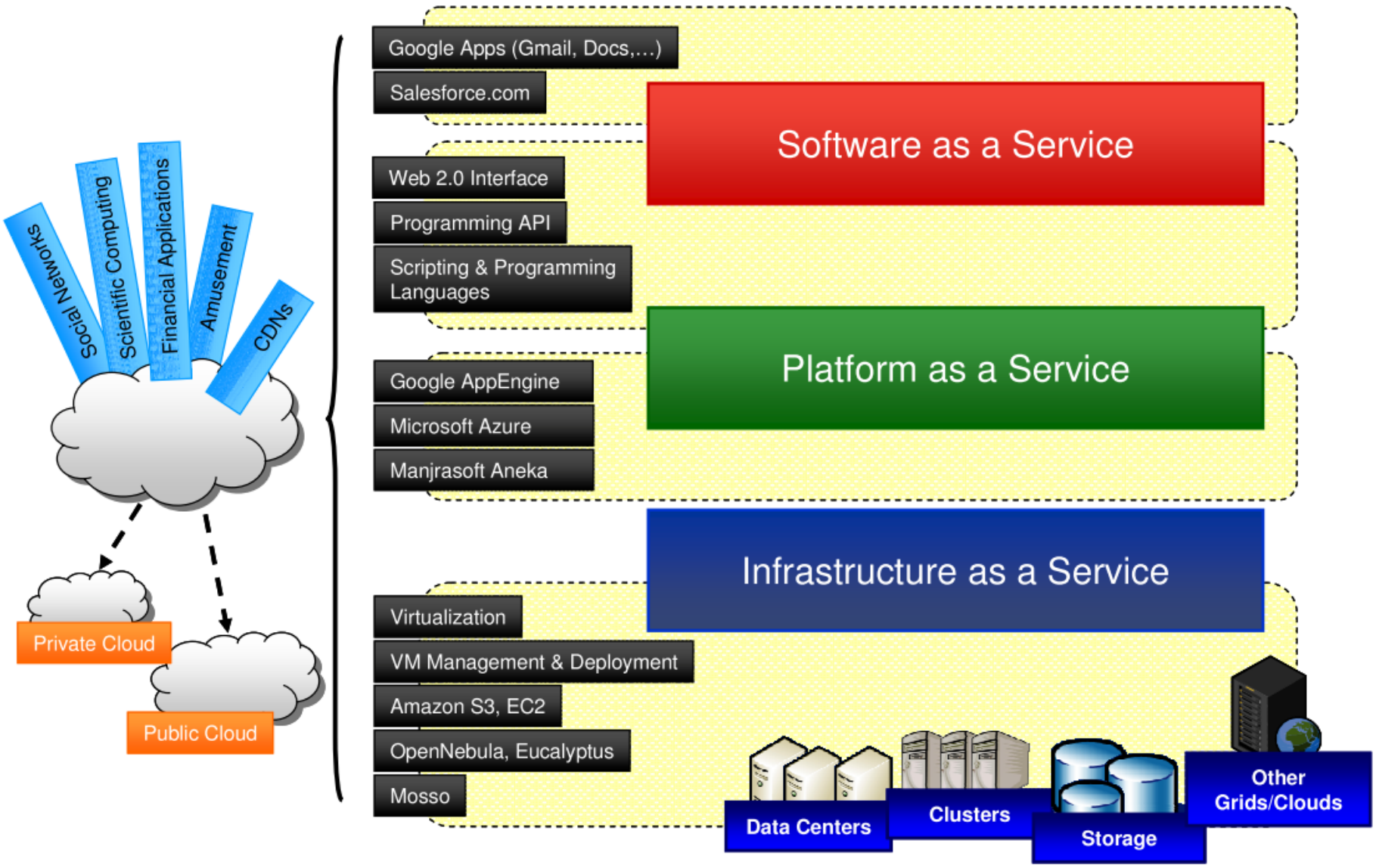
NIST

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model promotes availability and is composed of

- ✓ five essential characteristics (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured Service);
- ✓ **three service models (Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), Cloud Infrastructure as a Service (IaaS));** and**,**
- ✓ four deployment models (Private cloud, Community cloud, Public cloud, Hybrid cloud).

Key enabling technologies include: (1) fast wide-area networks, (2) powerful, inexpensive server computers, and (3) high-performance virtualization for commodity hardware.


Social Networks
Scientific Computing
Financial Applications
Amusement
CDNs
Private Cloud
Public Cloud
Google Apps (Gmail, Docs,…)
Salesforce.com
Software as a Service
Web 2.0 Interface
Programming API
Scripting & Programming Languages
Google AppEngine
Microsoft Azure
Manjrasoft Aneka
Platform as a Service
Infrastructure as a Service
Virtualization
VM Management & Deployment
Amazon S3, EC2
OpenNebula, Eucalyptus
Mosso
Data Centers
Clusters
Storage
Other Grids/Clouds

# Cloud Computing: a Model

NIST

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

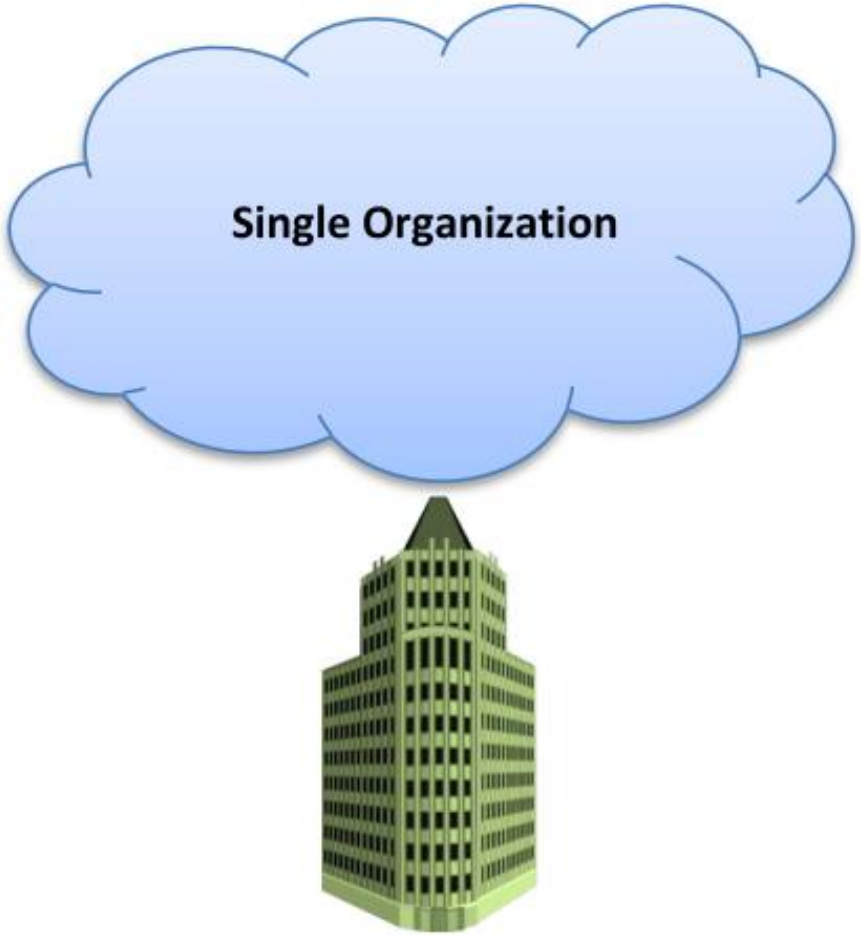
This cloud model promotes availability and is composed of

- ✓ five essential characteristics (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured Service);
- ✓ three service models (Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), Cloud Infrastructure as a Service (IaaS)); and,
- ✓ **four deployment models (Private cloud, Community cloud, Public cloud, Hybrid cloud).**

Key enabling technologies include: (1) fast wide-area networks, (2) powerful, inexpensive server computers, and (3) high-performance virtualization for commodity hardware.

*From Enterprise's Perspective …*

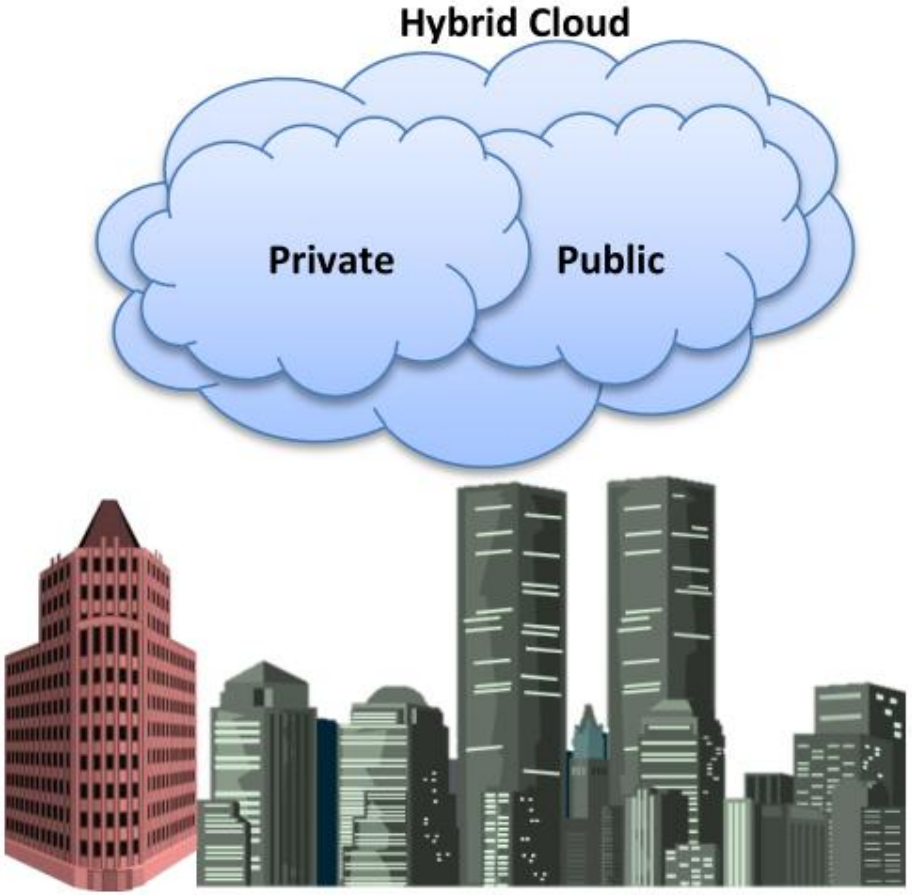
**Private Cloud**




- Virtualization
- Rapid provisioning and re-provisioning of capacity to respond to fluctuating capacity needs
- Public cloud services accomplish this with large server farms, flexibility and pay-as-you-go nature


- Leveraging external public clouds' resources/services
- What to keep and what to let go, particularly in terms of physical infrastructure.
- What to stay and what should go?
- What are needed to bring within enterprise's firewall?

# What systems should stay and what should go?



# Community, Collaborative Cloud

# NIST's notes on Cloud Computing

- Still an *evolving* paradigm
  - Its definitions, use cases, underlying technologies, issues, risks, and benefits will continue to be refined in a spirited debate by the public and private sectors.
  - These definitions, attributes, and characteristics will evolve and change over time.
- The cloud computing industry represents a large ecosystem of many models, vendors, and market niches.
  - This definition attempts to encompass all of the *various cloud approaches*.

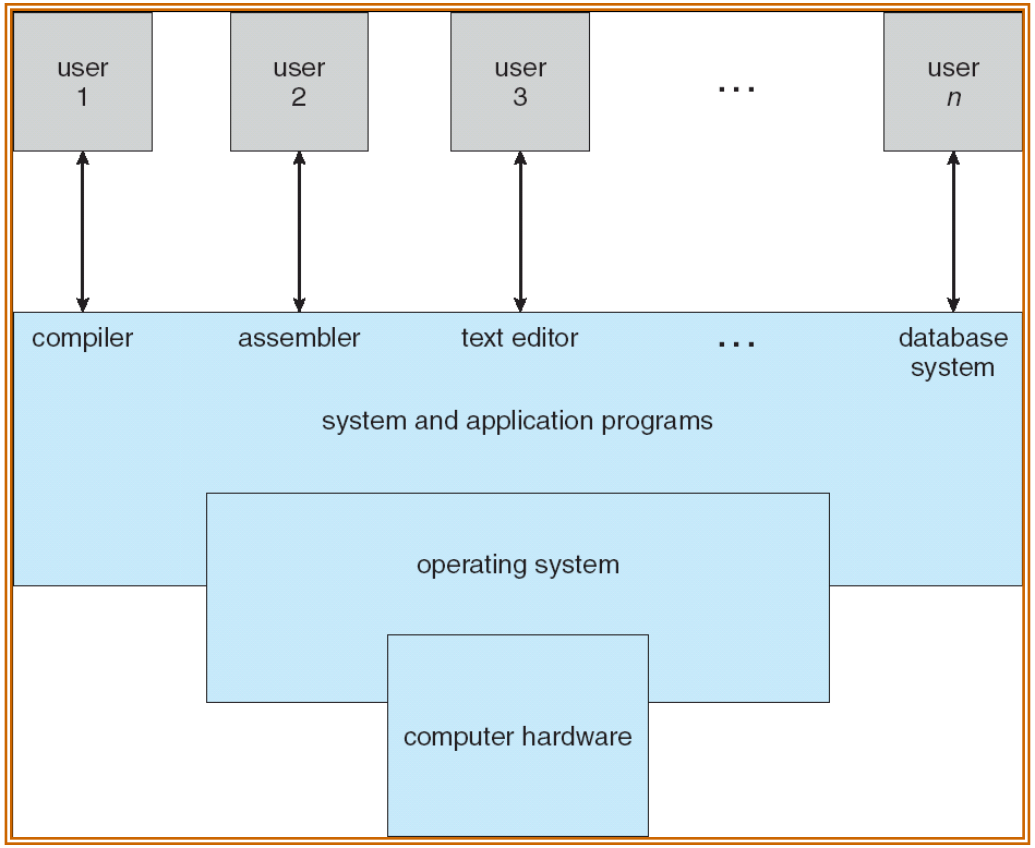
# Resource Sharing vs. Virtualization

# What is an Operating System?

- A ***program*** that acts as an intermediary between a ***user*** of a computer and the ***computer hardware***.
- ***Various*** operating system goals:
  - ***Mainframe*** operating systems: to optimize utilization of hardware.
  - ***PC*** operating systems: to support complex games, business applications …
  - ***Handheld computers***: to help users easily interface with the computer to execute programs.

# A Computer System: Four Components

- **Hardware** – provides basic **computing resources**.
  - CPU, memory, I/O devices.
- **Operating system** – **controls** and **coordinates** use of hardware among various applications and users
- **Application programs** – define the ways in which the system resources are used **to solve the computing problems of the users**.
  - Word processors, compilers, web browsers, database systems, video games.
- **Users** – people, machines, other computers.



# Viewpoints from Users (1/2)

- **PC**: the OS is designed for ***one user only***.
  - Resources are monopolized.
  - The goal is to maximize the work of the user.
  - The OS is generally designed for ***ease of use***, with some attention paid to performance and non paid to resource utilization.
- **Mainframe**: the OS is designed for ***multiple users*** – accessing the same computer through terminals.
  - These users share resources.
  - The OS is designed to ***maximize resource utilization*** – to assure that all available CPU time, memory …
  - terminal

# Viewpoints from Users (2/2)

- **Workstation**:
  - User**s** sit at workstation**s** connected to networks of other workstations and servers (file, compute, and print servers).
  - The OS is designed to compromise between individual usability and resource utilization.
  - Client-server
  - Thin client
- **Handheld computer**: are **standalone** units for individual users.
  - The OS is designed mostly for individual usability.
  - But performance per amount of battery life is important as well.
- **Computer with little (or no) user view**: embedded home devices.
  - The OS is designed to run without user intervention.
  - Machine-to-machine (Internet of things, IOT)
  - cyberphysics

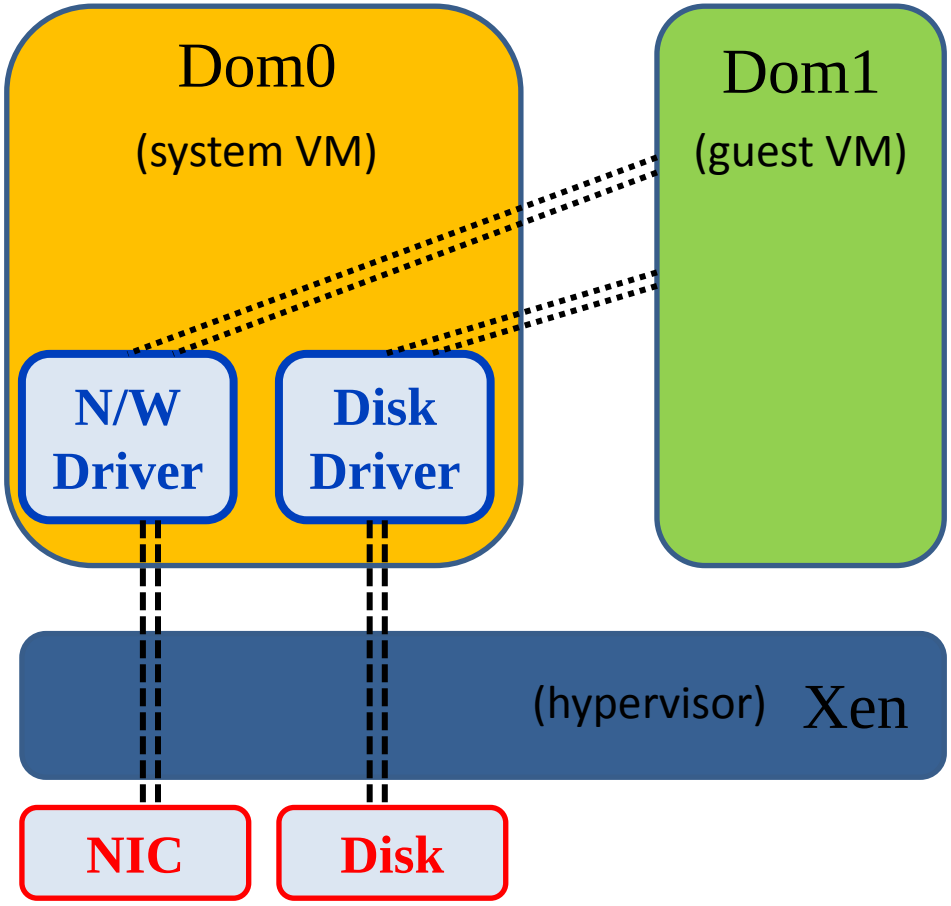
# Viewpoints from Computers

- OS for computer is the program involved with the hardware.
- OS is a ***resource allocator and manager***.
  - Manages all resources.
  - Decides between **conflicting requests** for efficient and **fair** resource use.
- OS is a ***control program***.
  - Controls execution of programs to prevent errors and improper use of the computer.

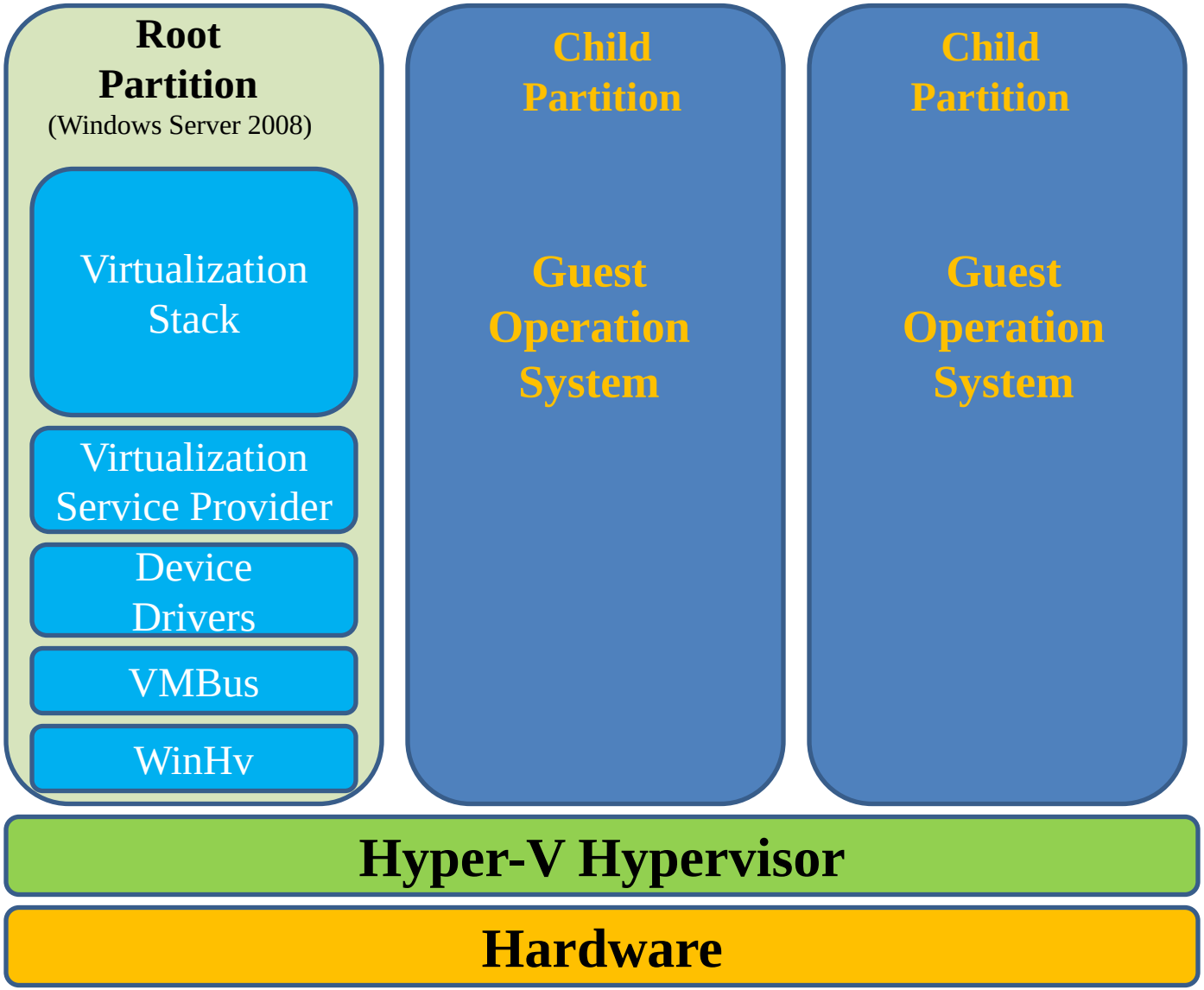
# VM Technology

- VM technology (2003)
  - Modern computers are sufficiently powerful to use virtualization.
- Solutions
  - VMware ESX, Microsoft Hyper-V, Citrix XenServer/XenClient, Xen, KVM, etc.

# Virtualization (1/2)



# Virtualization (2/2)



# Why Virtualization (1/2)

- Traditional operating system and host.
- Virtualization is a key mechanism of **scaling** the IT infrastructure and enabling enterprises to move from vertical silos of servers to **horizontal pools of resources**.
- **Server virtualization** provides the ability to slice larger, underutilized physical servers into smaller, virtual ones.
- Although virtualization has been around for more than three decades, it has found its way into the mainstream only recently, as a consequence of the recent developments in virtualization software and improved hardware support.
- A variety of solutions — both commercial and open source — are now available.

# Why Virtualization (2/2)

- It **isolates** customers from the operational and hardware complexity associated with deploying compute in a static private data center.
- Customers can control their virtual servers in the **same** way they control their physical servers.
- It achieves much **better resource utilization**, **resource planning**, and the **ability to simplify the management of the infrastructure components**.

# Virtualization: Approaches

- **Full virtualization**
  - The virtual hardware exposed is **functionally identical** to the underlying machine.
  - Benefit:
    - allowing unmodified operating systems to be hosted.
  - Drawback:
    - Hardware support

    ✓(e.g., Intel Virtualization Technology, AMD)

# Virtualization: Approaches

- Paravirtualization
  - Presenting a virtual machine abstraction that is ***similar but not identical*** to the underlying hardware.
  - Improved performance.
  - Requiring **modifications** to the guest operating system, but do **not** require changes to the application binary interface (ABI)
    - **ABI** describes the low-level interface between an application program and the operating system.
  - No modifications to guest applications

# Motivations for Enterprises

- The motivations for enterprises to adopt virtualization technologies include
  - increased flexibility;
  - server and application consolidation;
  - the ability to quickly re-purpose server capacity to better meet the needs of application workload owners;
  - to reduce overall costs of ownership.
- Virtualization services offer interfaces that
  - Support the life cycle management (e.g., create, destroy, move, size capacity) of VMs that are provided with access to shares of resource capacity (e.g., cpu, memory, input-output).
  - provide the ability to dynamically migrate VMs from one physical machine to another without interrupting application execution.

# Issues that must be addressed in virtualization (1/2)

- Virtualization plays an important role in cloud.
  - With virtualization within cloud architectures, users can provision their own space.
- VM isolation
  - It is not acceptable for the execution of one to adversely affect the performance of another.
  - Performance isolation
    - admission control when starting new virtual machines.
    - Expect that each VM to pay for the resources it requires.
  - Resource allocation/usage
  - Security isolation

# Issues that must be addressed in virtualization (2/2)

- Support different operating systems
  - To accommodate the heterogeneity of popular applications.
- The performance overhead introduced by virtualization should be small.

- Management Complexity
  - The complexity of these virtualized environments presents additional management challenges.

To be continued.