BITS Pilani presentation

D. Powar
Lecturer,
BITS-Pilani, Hyderabad Campus

BITS Pilani
Hyderabad Campus

innovate achieve lead

BITS Pilani
Hyderabad Campus

CS ZG527
Cloud Computing
Lecture -02

Objectives

- Origins and Motivation
- Introduction to Cloud Computing
  - Definition
  - 3-4-5 rule
- Types of Clouds and Services
- Cloud Computing stack
- Challenges for cloud computing

Why do I study Cloud Computing??

- We don’t know what we don’t know?
- To be updated in the present computing technology
- Present IT infrastructure can’t be managed with traditional IT infrastructure

How much data we produce?

- Wayback Machine has 2 PB + 20 TB/month (2006)
- Google processes 20 PB a day (2008)
- NOAA has ~1 PB climate data (2007)
- CERN's LHC generated 15 PB a year (2008)

http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

How much data we produce? (contd..)

- 2.7 Zetabytes of data exist in the digital universe today
- Facebook stores, accesses, and analyzes 30+ Petabytes of user generated data
- Akamai analyzes 75 million events per day to better target advertisements
- Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data
- 100 terabytes of data uploaded daily to Facebook

Source: http://wikibon.org/blog/big-data-statistics/

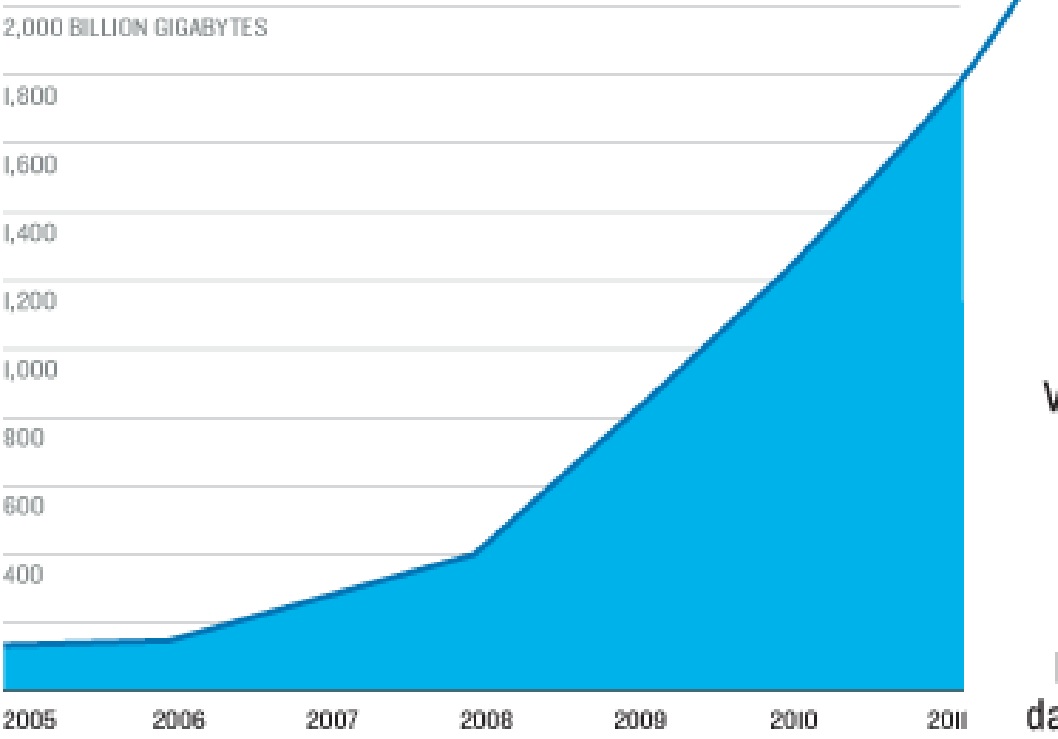
MIT Technology review

65 billion
Location-tagged payments made in the U.S. annually

154 billion
E-mails sent per day

87%
U.S. adults whose location is known via their mobile phone

Digital Information Created Each Year, Globally



2,000%
Expected increase in global data by 2020

111 Megabytes
Video and photos stored by Facebook, per user

75%
Percentage of all digital data created by consumers

Sources: IDC, Radicati Group, Facebook, TR research, Pew Internet

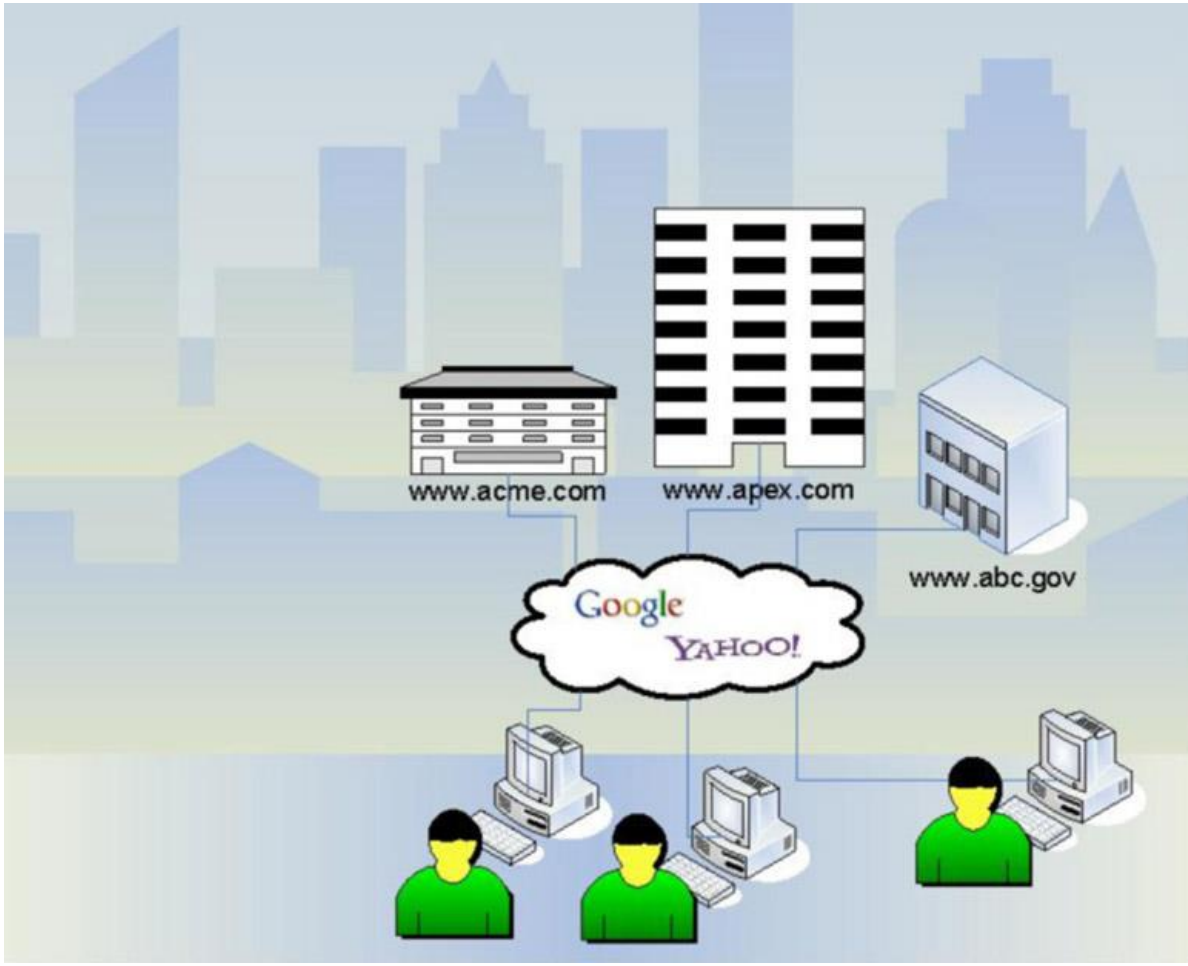
Data center

Where are we today?

Evolution of the Web

Web 1.0 (1990s):

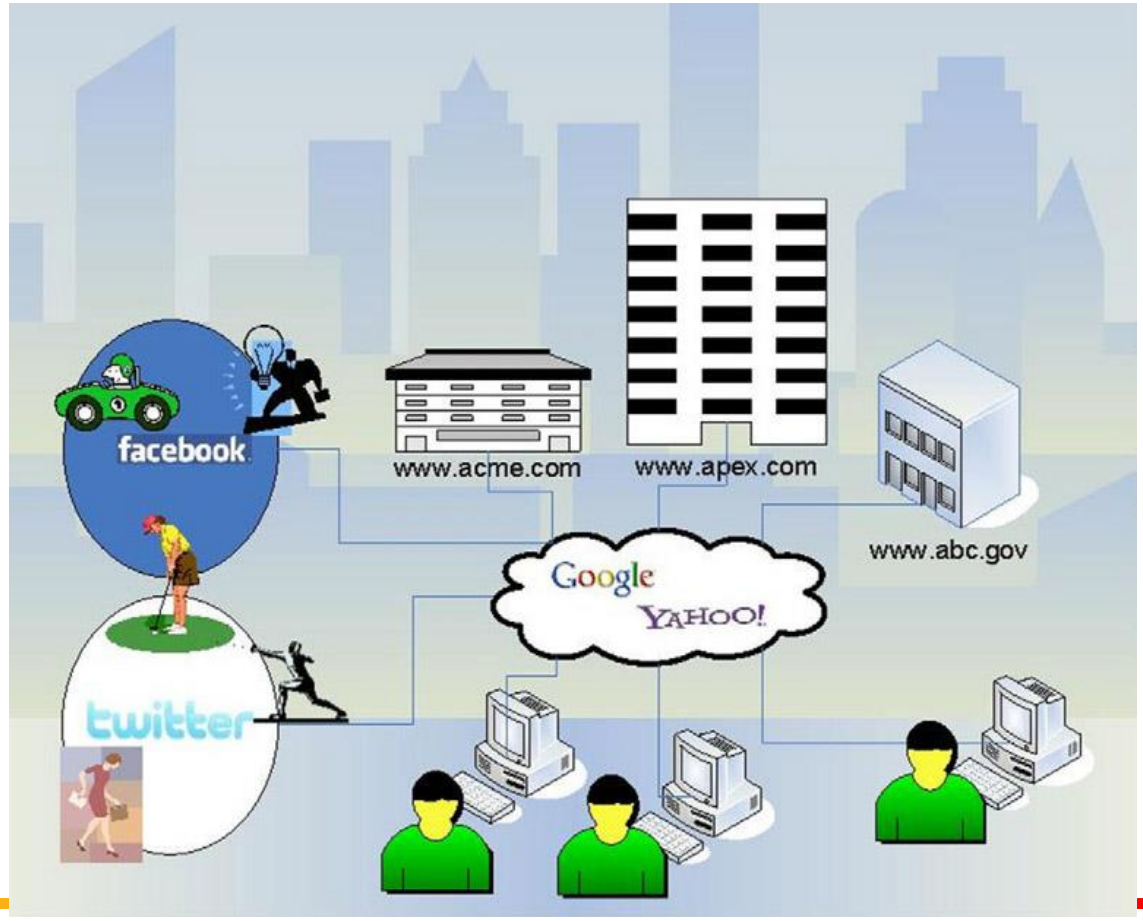
- Information flow was strictly one-way from web sites to users



Where are we today? (contd..)

Web 2.0 (2000s):

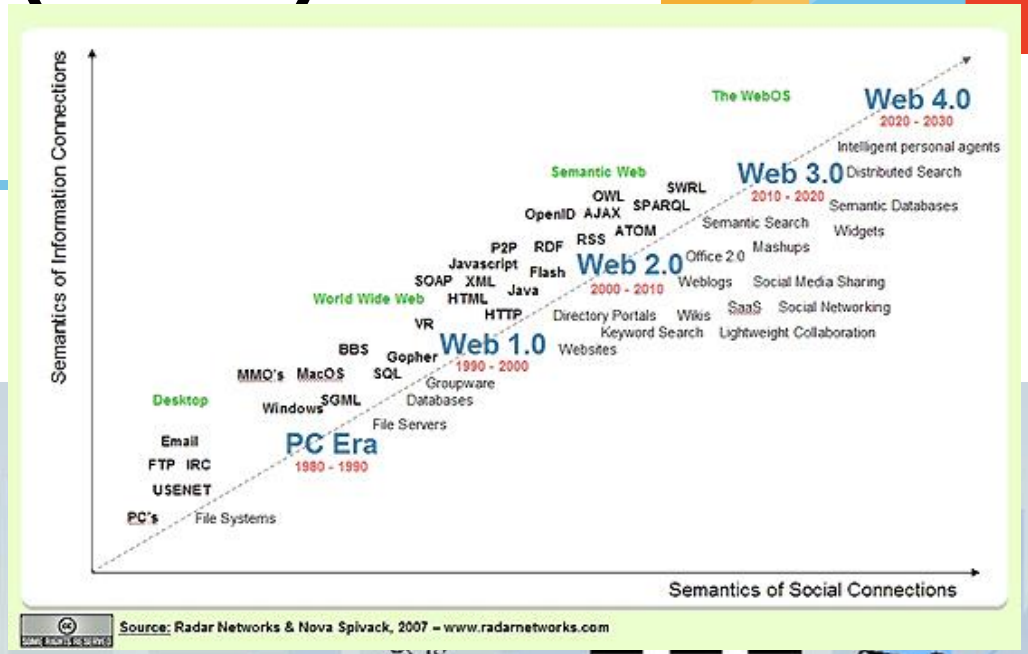
- Information flow was bi-directional, thus allowed users to upload information to the Web
- Mobile web

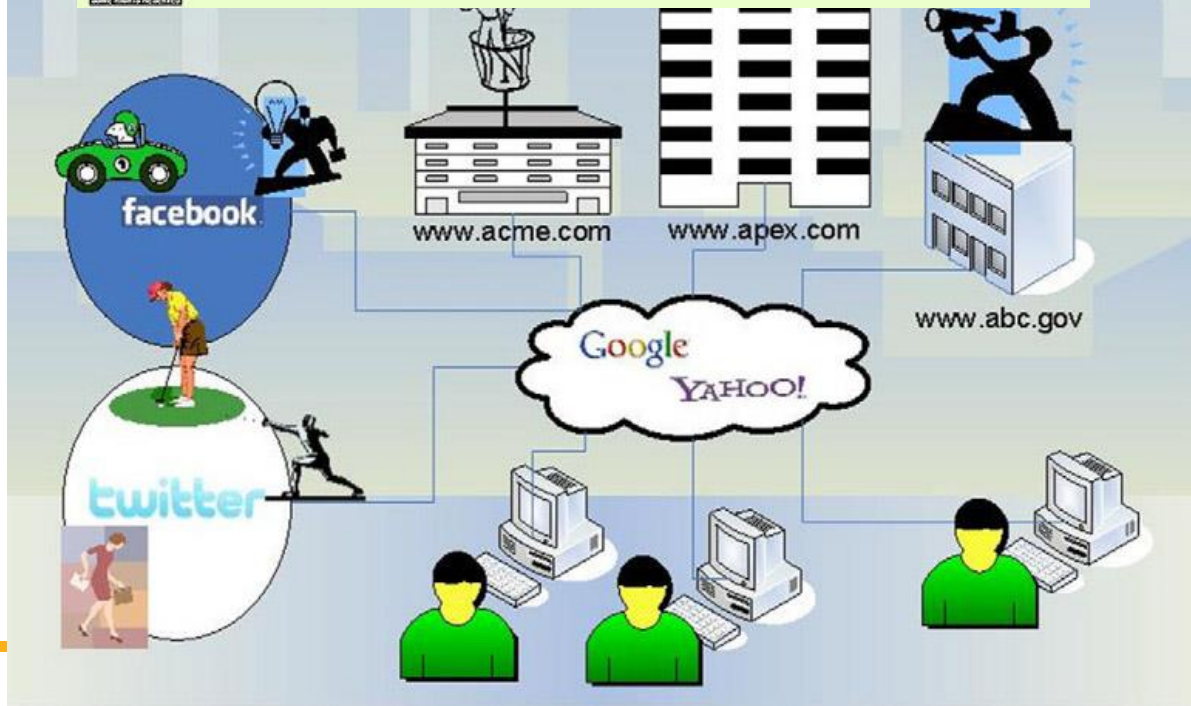


Where are we today? (contd..)

Web 3.0 (2010):

- The next generation Web
- Has been humorously called ”Cyberspace looks at You”





Do we need to improve??

- Should application developers be burdened by the tasks of ensuring that a specific server is up and running?
- Should they worry about disk space?
- Should they worry about which OS their app should support?
- The focus should be on solving the much bigger problems
- The compute infrastructure, platform, libraries and application deployment should all be automated and abstracted.
- This is where Cloud Computing plays a major role

Why could cloud computing be successful when other paradigms have failed??

- It is in a better position to exploit recent advances in software, networking, storage, and processor technologies promoted by the same companies who provide cloud services
- It is focused on enterprise computing; its adoption by industrial organizations, financial institutions, government, and so on could have a huge impact on the economy
- It consists of a homogeneous set of hardware and software resources
- It provides the illusion of infinite computing resources
- It eliminates the need for up-front financial commitment

Introduction

CLOUD COMPUTING IN A NUTSHELL

- Cloud computing is like plugging an electric appliance into an outlet
- All because of virtualization
- An umbrella term to describe a category of sophisticated on-demand computing services
- It denotes a model on which a computing infrastructure is viewed as a “cloud”, from which businesses and individuals access applications from anywhere in the world on demand
- Principle: Offering computing, storage, and software “as a service

What is Cloud Computing?

*Cloud computing is a model for enabling **ubiquitous**, 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.*

(by NIST)

"an Internet based computing paradigm that delivers on-demand software and hardware computing capability as a 'service' through virtualization where the end user is completely abstracted from the computing resources"

What is Cloud Computing? (contd..)

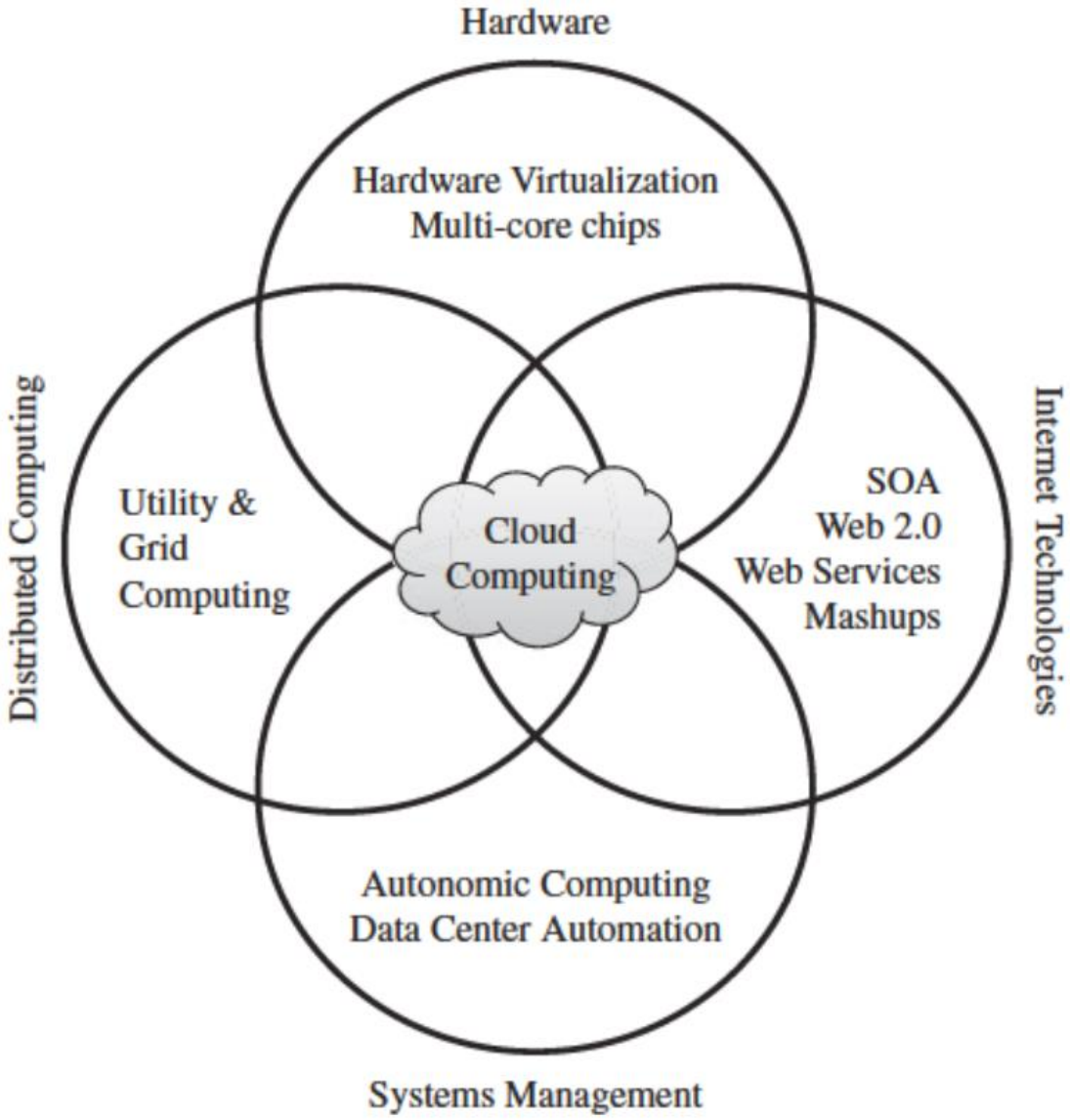
Author/Organization	Definition
The *Open Cloud Manifesto* Consortium	The ability to scale and provision computing power dynamically in a cost-efficient way and the ability of the consumer (end user, organization, or ITstaff) to make the most of that power without having to manage the underlying complexity of the technology (OpenCloudManifesto.org, 2009, p.2)
The University of California, Berkeley Reliable Adaptive Distributed Systems Laboratory	Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the Data Centers that provide those services. The services themselves have long been referred to as Software as a Service (SaaS), so we use that term. The Data Center hardware and software is what we call a Cloud (Armbrust, et al., 2009)
Gartner	A style of computing where massively scalable IT-related capabilities are provided "as a service" using Internet technologies to connect multiple external customers (Gartner, 2008)
Michael Brown	A data-processing infrastructure in which the application software—and often the data itself—is stored permanently not on your PC but rather a remote server that's connected to the Internet (Brown, 2009)
Jaeger, Lin, Grimes, and Simmons	An emerging model of computing where machines in large data centers can be dynamically provisioned, configured, and reconfigured to deliver services in a scalable manner, for needs ranging from scientific research to video sharing to e–mail (Jaeger, et al., 2009)

Note: **Most significant switch in the IT world since the advent of the Internet**

ROOTS OF CLOUD COMPUTING

Hardware	virtualization, multi-core chips
Internet technologies	Web services, service-oriented architectures, Web 2.0
Distributed computing	clusters, grids
Systems management	Autonomic computing, data center automation

Convergence of technology fields that significantly advanced and contributed to the advent of cloud computing



3-4-5 Rule

???

3 : Services

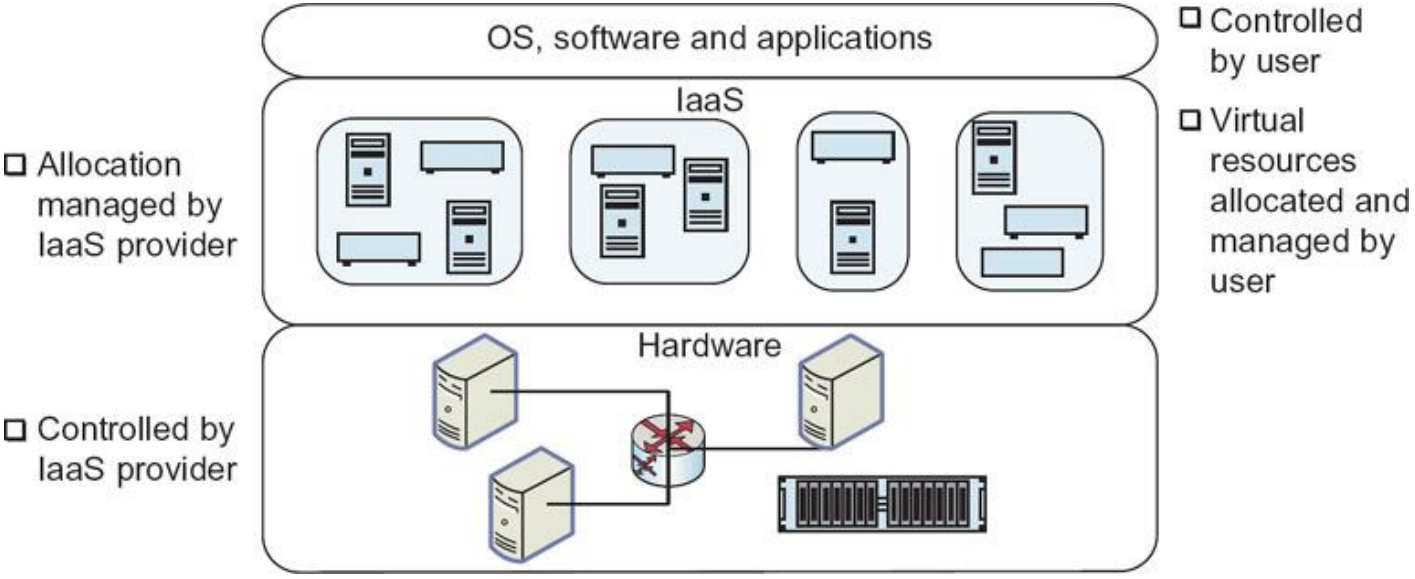
4 : Deployment Models

5 : Characteristics

3 : Services

1. IaaS (Infrastructure as a Service):

The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls)



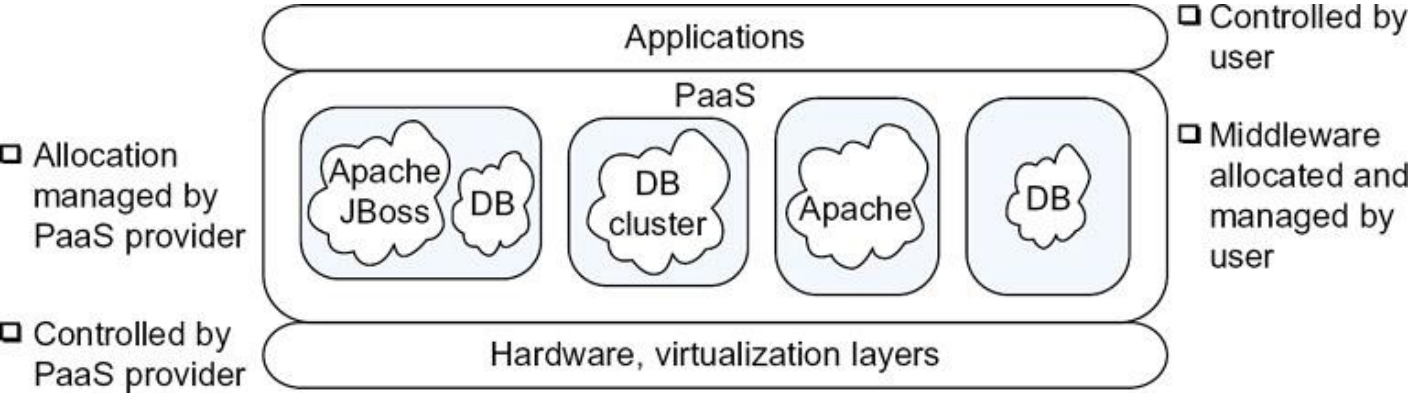
Public IaaS:
Amazon EC2, Rackspace, Rightscale, etc.

Private IaaS:
HP, IBM, Microsoft, OpenStack, Eucalyptus, etc.

3 : Services

2. PaaS (Platform as a Service):

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations

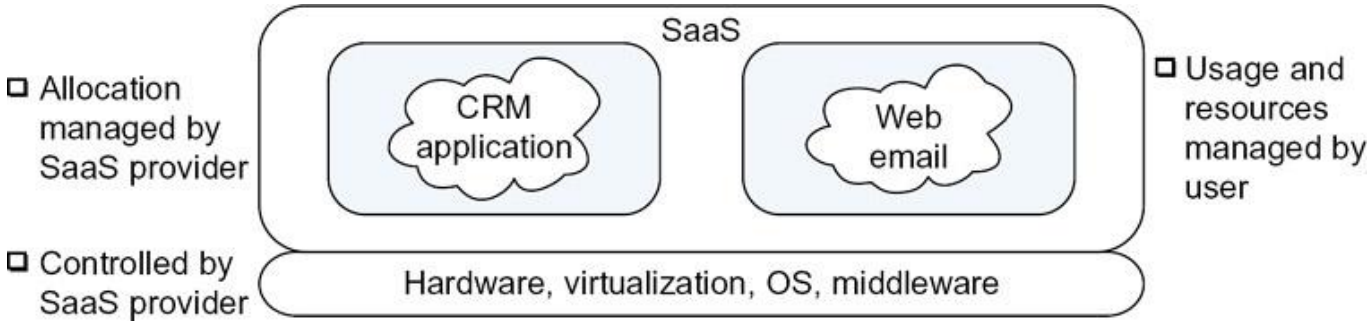


Windows Azure, Google App Engine, Hadoop, etc. are some well-known PaaS platforms

3 : Services

3. SaaS (Software as a Service):

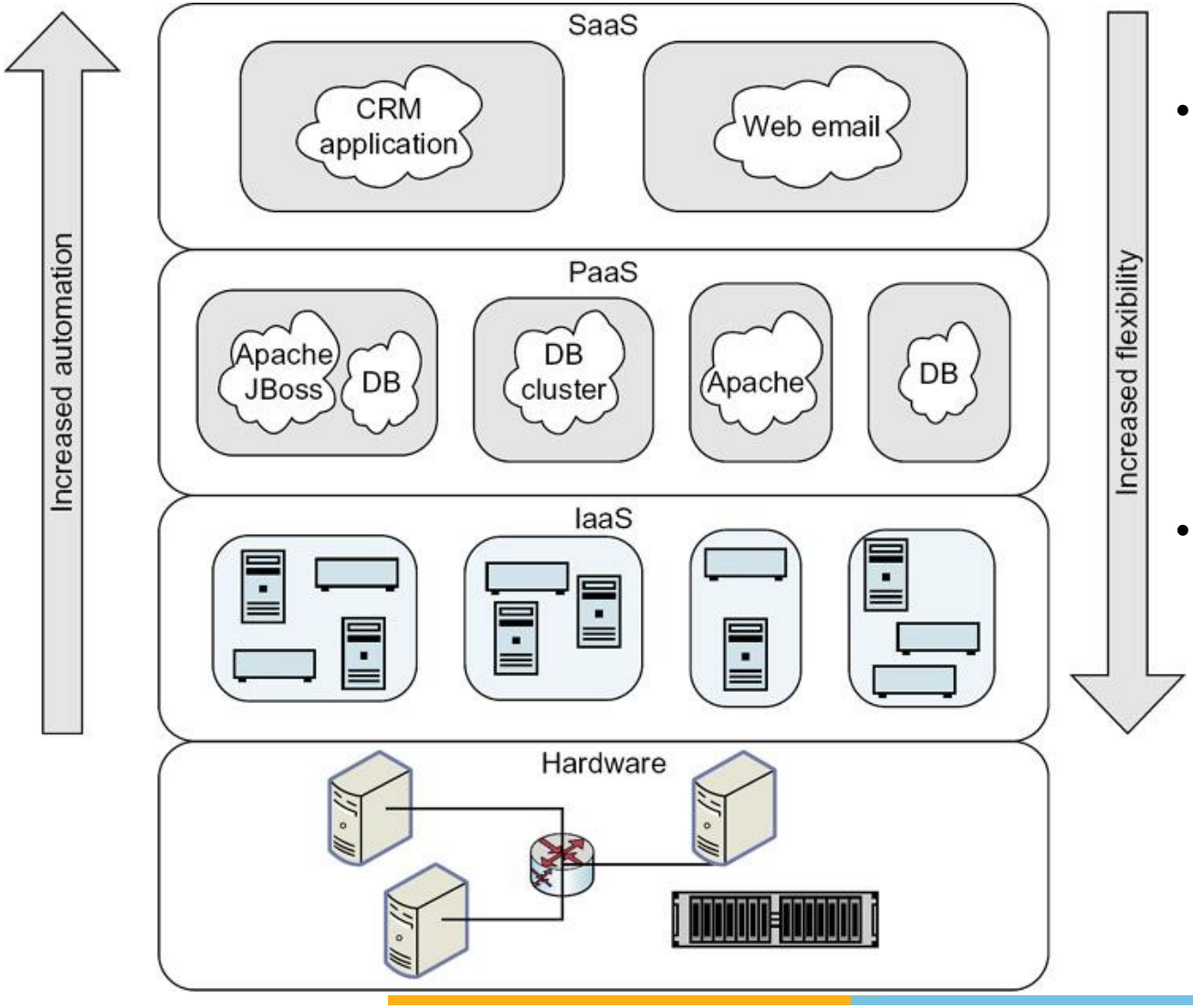
The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings



Moto: "No Software"

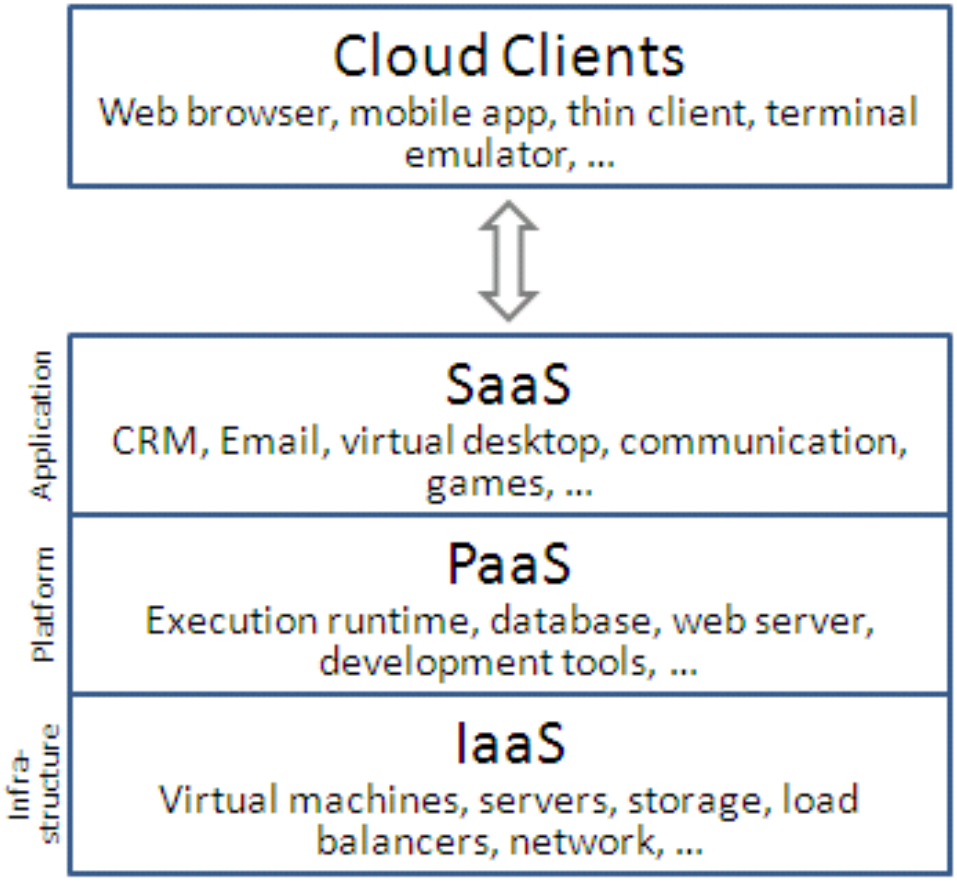
Ex: Salesforce.com for CRM, Google Docs for document sharing, and web email systems like Gmail, Hotmail, and Yahoo! Mail

Cloud Service Models



- While the features offered by the 3 service types may be different, there is a common set of technological challenges that all cloud architectures face.
- These include computation scaling, storage scaling, multi-tenancy, availability, and security.

Cloud Service Models (contd..)



Source: wiki

Note: Cloud Service Model is also called as **'SPI** Model' by NIST

Cloud computing services (contd..)

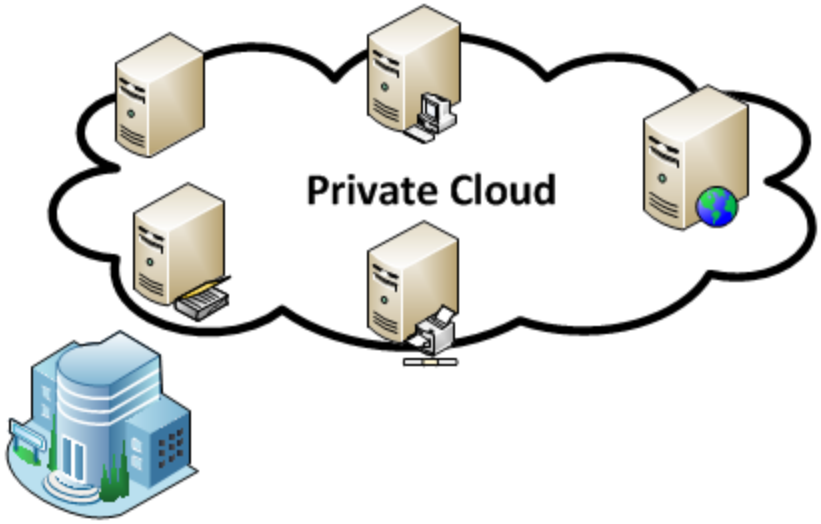
- Few categories of services heard as on today of **cloud computing** technology:
  - Storage-as-a-service
  - Database-as-a-service
  - Information-as-a-service
  - Process-as-a-service
  - Application-as-a-service (SaaS)
  - Integration-as-a-service
  - Security-as-a-service
  - Management/governance-as-a-service
  - Testing-as-a-service
  - XaaS
  - Etc.

4: Cloud Deployment Models

- Private Cloud
- Community Cloud
- Public Cloud
- Hybrid Cloud

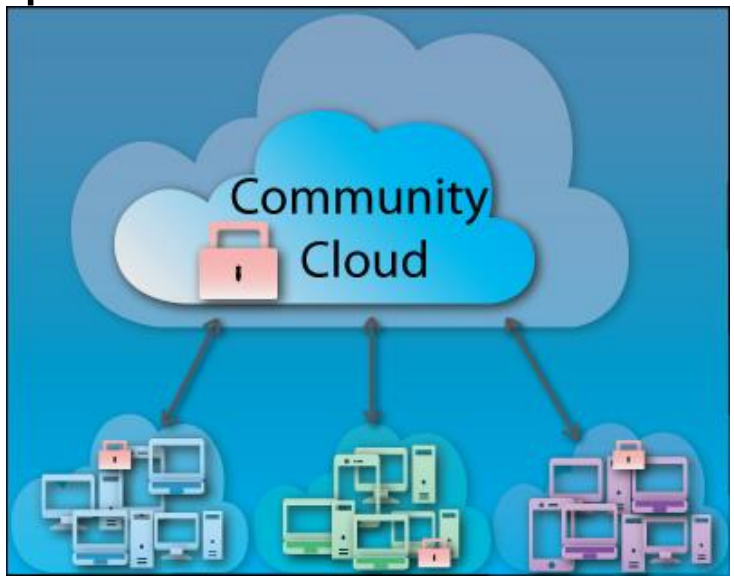
Private Cloud

- The cloud infrastructure is operated solely for a single organization. It may be managed by the organization or a third party, and may exist on-premises or off-premises
- Also called as “Enterprise Cloud”



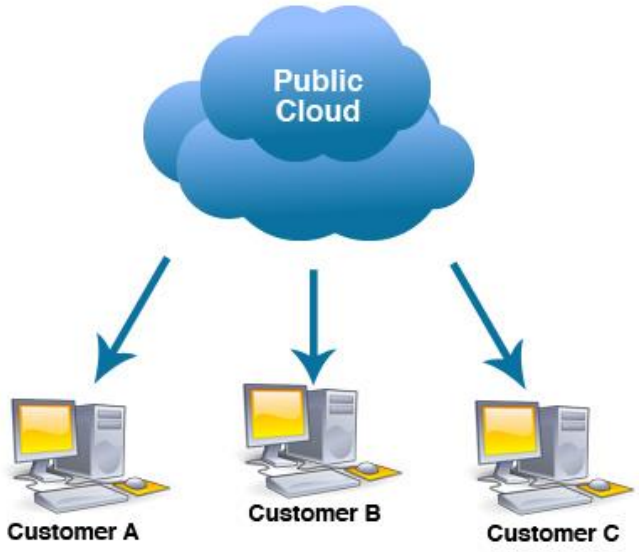
Community Cloud

- The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, or compliance considerations).
- It may be managed by the organizations or a third party and may exist on-premises or off-premises.



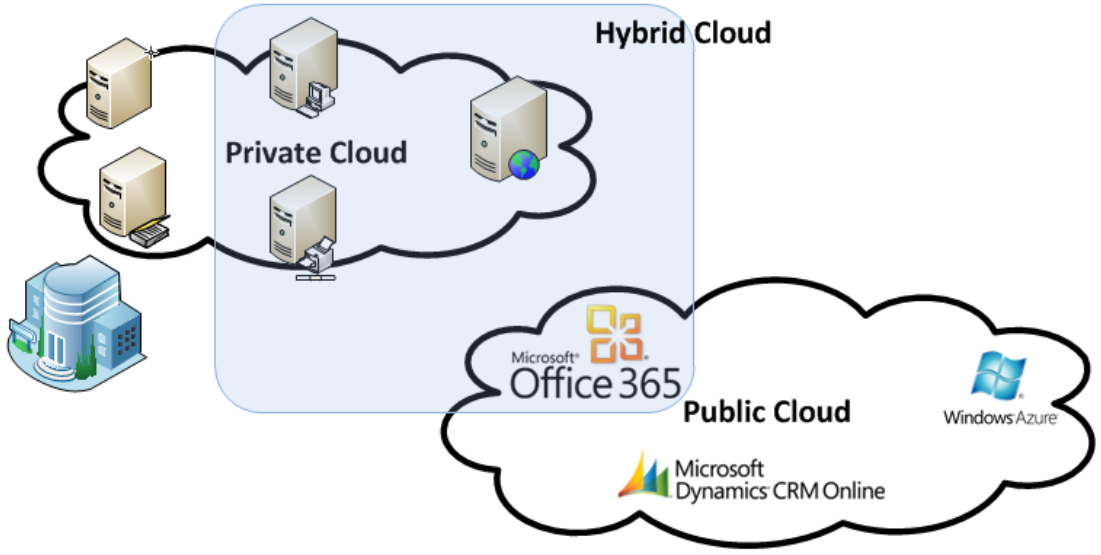
Public Cloud

- The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services
- Also called as “Internet Cloud”



Hybrid Cloud

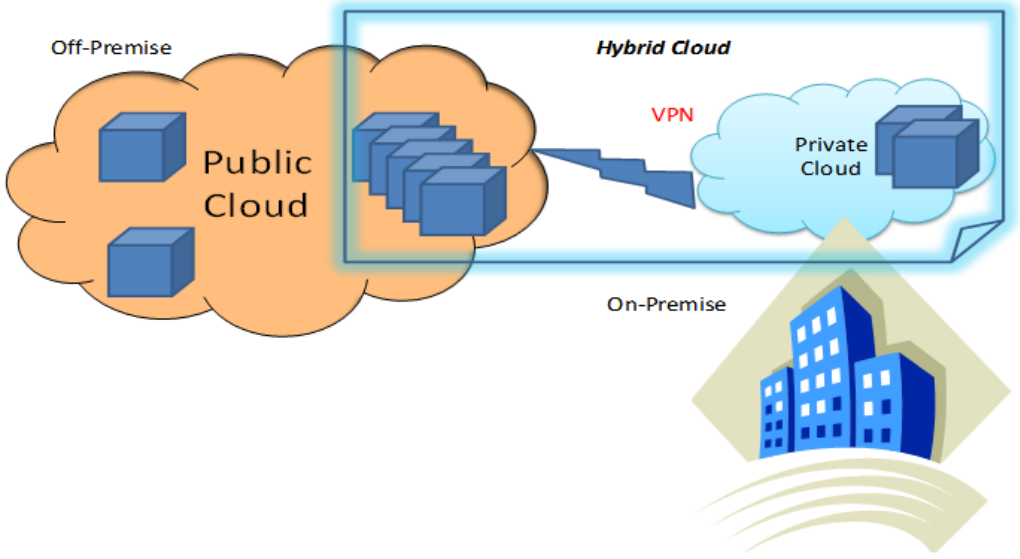
- The cloud infrastructure is a composition of two or more clouds (private, community, or public)
- Leasing public cloud services when private cloud capacity is insufficient
- Also called as “Mixed Cloud”



Hybrid Cloud (contd..)

Cloud-bursting?

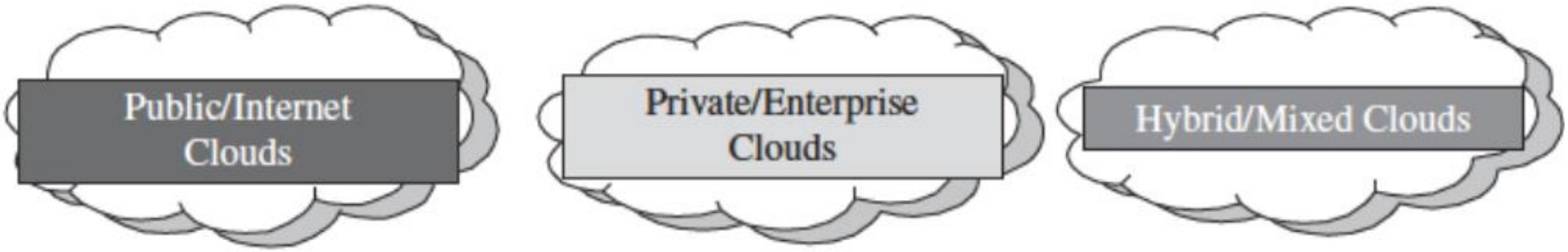
- A hybrid cloud takes shape when a private cloud is supplemented with computing capacity from public clouds
- The approach of temporarily renting capacity to handle spikes in load is known as "Cloud-bursting"



Summary of deployment models

	Type	Properties
1.	Private cloud	• Outsource or own • Lease or buy • Separate or virtual data center
2.	Community cloud	• Private cloud for a set of users with specific demands • Several stakeholders
3.	Public cloud	• Mega scaleable infrastructure • Available for all
4.	Hybrid cloud	• Combination of two clouds • Usually private for sensitive data and strategic applications

Deployment Models (contd..) (*Types of clouds*)



3rd party, multi-tenant Cloud infrastructure & services:

* available on subscription basis (pay as you go)

Cloud computing model run within a company's own Data Center/ infrastructure for internal and/or partners use.

Mixed usage of private and public Clouds: Leasing public cloud services when private cloud capacity is insufficient

Ref: Text book1

5: Characteristics

On-demand self-service:

- Computing resources can be provisioned on-demand by the users, without requiring interaction with the CSP
- The process of provisioning resources is automated

Broad network access:

- Any computing capabilities are available over the network. Many different devices are allowed access through standardized mechanisms

Resource pooling:

- Multiple users can access clouds that serve other consumers according to demand
- Cloud services need to share resources between users and clients in order to reduce costs
- All because of multi-tenancy and virtualization

5: Characteristics (contd..)

Rapid Elasticity:

- Computing resources can be provisioned rapidly and elastically based on need
- Two types of scaling
  - ❖ Horizontal scaling (scaling out)
  - ❖ Vertical scaling (scaling up)

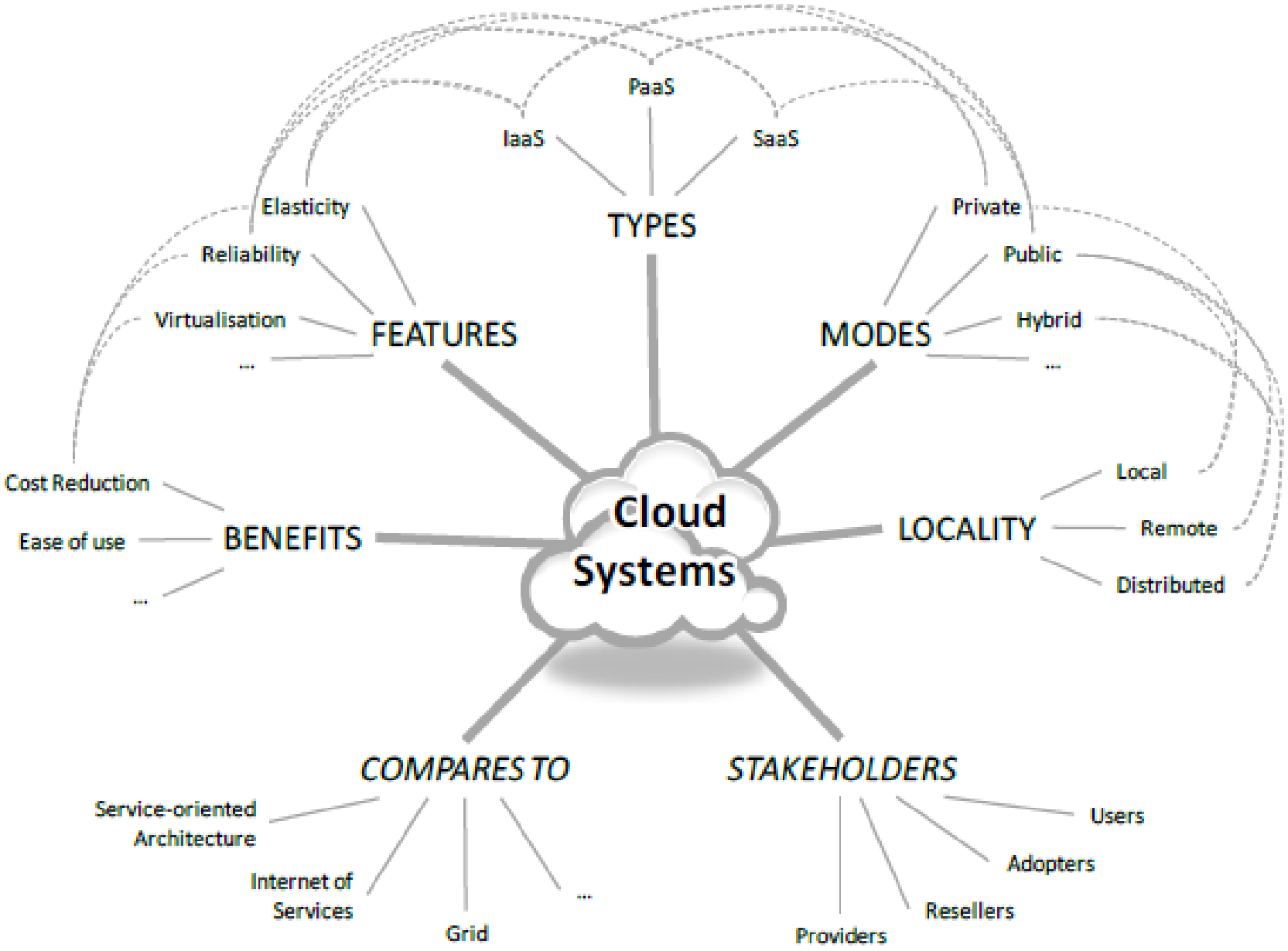
Metered or measured service:

- One of the compelling business use cases for cloud computing is the ability to "pay as you go", where the consumer pays only for the resources that are actually used by his applications
- The usage of the cloud resources is measured
  - CPU cycles, Physical memory, Storage space, No. of network I/O requests, etc.

The Cloud computing stack

Service Class	Main Access & Management Tool	Service content
 SaaS	Web Browser	**Cloud Applications** Social networks, Office suites, CRM, Video processing
 PaaS	Cloud Development Environment	**Cloud Platform** Programming languages, Frameworks, Mashups editors, Structured data
 IaaS	Virtual Infrastructure Manager	**Cloud Infrastructure** Compute Servers, Data Storage, Firewall, Load Balancer

Main Aspects of a Cloud System



wiki

Challenges for cloud computing

- Availability of service: What happens when the service provider cannot deliver?
- Vendor lock-in: Diversity of services, data organization, user interfaces available at different service providers limit user mobility; once a customer is hooked to one provider it is hard to move to another. Standardization efforts at NIST!
- Data confidentiality and auditability: a serious problem
- Data transfer bottleneck: many applications are data-intensive

Challenges for cloud computing (contd..)

- Performance unpredictability: one of the consequences of resource sharing

- Elasticity: the ability to scale up and down quickly.
  - New algorithms for controlling resource allocation and workload placement are necessary. Autonomic computing based on self-organization and self-management seems to be a promising avenue

Summary:

- Origins and Motivation
- Introduction to Cloud Computing
  - Definition
  - 3-4-5 rule
- Types of Clouds and Services
- Cloud Computing stack
- Challenges for cloud computing