CS6006 - CLOUD COMPUTING

Module 1 - INTRODUCTION TO CLOUD COMPUTING

Presented By
Dr. S. Muthurajkumar,
Assistant Professor,
Dept. of CT, MIT Campus,
Anna University, Chennai

INTRODUCTION TO CLOUD COMPUTING

- Introduction to Cloud Computing
- Roots of Cloud Computing
- System Models for Distributed and Cloud Computing
- Layers and Types of Clouds
- Desired Features of a Cloud
- Cloud Infrastructure Management
- Infrastructure as a Service Providers
- Platform as a Service Providers
- Challenges and Risks
- Architectural Design of Compute and Storage Clouds

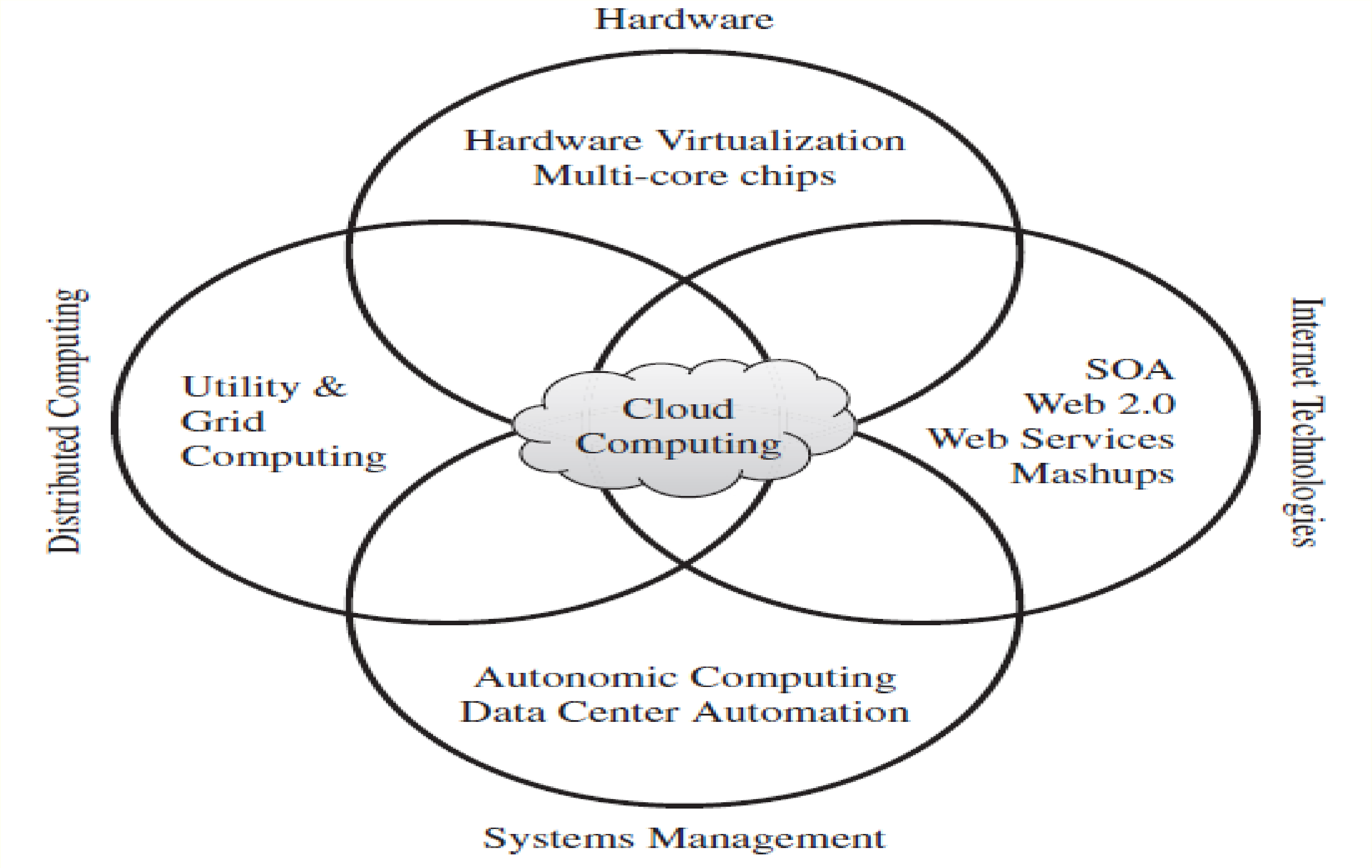
INTRODUCTION TO CLOUD COMPUTING

- Technologies such as cluster, grid, and now, cloud computing, have all aimed at allowing access to large amounts of computing power in a fully virtualized manner, by aggregating resources and offering a single system view.
- Cloud computing has been coined as an umbrella term to describe a category of sophisticated on-demand computing services initially offered by commercial providers, such as Amazon, Google, and Microsoft.
- 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.
- The main principle behind this model is offering computing, storage, and software “as a service”.

ROOTS OF CLOUD COMPUTING

- The roots of clouds computing by observing the advancement of several technologies, especially in hardware (virtualization, multi-core chips), Internet technologies (Web services, service-oriented architectures, Web 2.0), distributed computing (clusters, grids), and systems management (autonomic computing, data center automation).
  - From Mainframes to Clouds
  - SOA, Web Services, Web 2.0, and Mashups
  - Grid Computing
  - Utility Computing
  - Hardware Virtualization
  - Virtual Appliances and the Open Virtualization Format
  - Autonomic Computing

ROOTS OF CLOUD COMPUTING



Convergence of various advances leading to the advent of cloud computing

ROOTS OF CLOUD COMPUTING

- **From Mainframes to Clouds**
- Currently experiencing a switch in the IT world, from in-house generated computing power into utility-supplied computing resources delivered over the Internet as Web services.
- Computing delivered as a utility can be defined as “on demand delivery of infrastructure, applications, and business processes in a security-rich, shared, scalable, and based computer environment over the Internet for a fee”.

ROOTS OF CLOUD COMPUTING

- **From Mainframes to Clouds**
- This model brings benefits to both consumers and providers of IT services.
- Consumers can attain reduction on IT-related costs by choosing to obtain cheaper services from external providers as opposed to heavily investing on IT infrastructure and personnel hiring. The “on-demand” component of this model allows consumers to adapt their IT usage to rapidly increasing or unpredictable computing needs.
- Providers of IT services achieve better operational costs; hardware and software infrastructures are built to provide multiple solutions and serve many users, thus increasing efficiency and ultimately leading to faster return on investment (ROI) as well as lower total cost of ownership (TCO).

ROOTS OF CLOUD COMPUTING

- **SOA, Web Services, Web 2.0, and Mashups**
- The emergence of Web services (WS) open standards has significantly contributed to advances in the domain of software integration.
- Web services can glue together applications running on different messaging product platforms, enabling information from one application to be made available to others, and enabling internal applications to be made available over the Internet.
- The purpose of a SOA is to address requirements of loosely coupled, standards-based, and protocol-independent distributed computing.
- In a SOA, software resources are packaged as “services,” which are well defined, self contained modules that provide standard business functionality and are independent of the state or context of other services.

ROOTS OF CLOUD COMPUTING

- **Grid Computing**
- Grid computing enables aggregation of distributed resources and transparently access to them.
- Most production grids such as TeraGrid and seek to share compute and storage resources distributed across different administrative domains, with their main focus being speeding up a broad range of scientific applications, such as climate modeling, drug design, and protein analysis.
- A key aspect of the grid vision realization has been building standard Web service based protocols that allow distributed resources to be "discovered, accessed, allocated, monitored, accounted for, and billed for, etc., and in general managed as a single virtual system".
- The Open Grid Services Architecture (OGSA) addresses this need for standardization by defining a set of core capabilities and behaviors that address key concerns in grid systems.

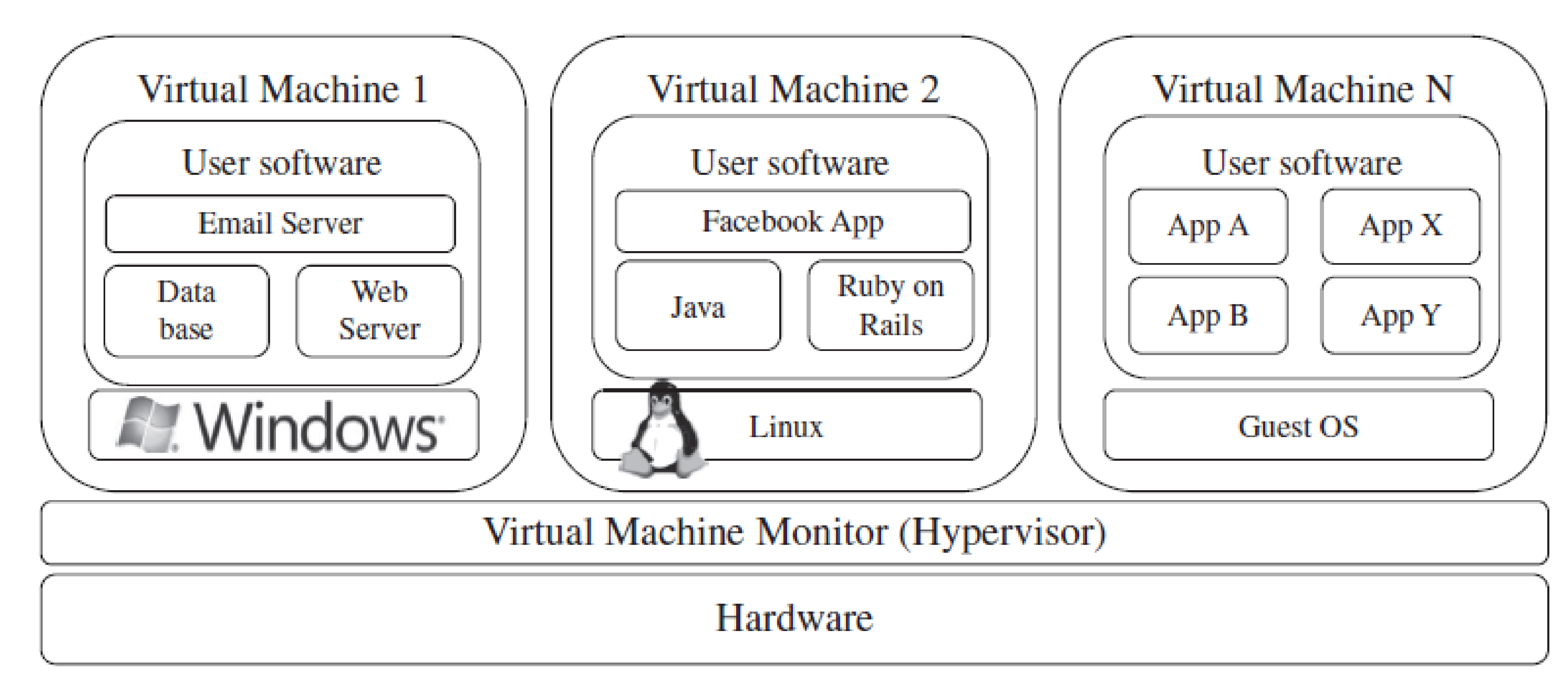
ROOTS OF CLOUD COMPUTING

- **Utility Computing**
- With increasing popularity and usage, large grid installations have faced new problems, such as excessive spikes in demand for resources coupled with strategic and adversarial behavior by users.
- Initially, grid resource management techniques did not ensure fair and equitable access to resources in many systems.
- Traditional metrics (throughput, waiting time, and slowdown) failed to capture the more subtle requirements of users.
- There were no real incentives for users to be flexible about resource requirements or job deadlines, nor provisions to accommodate users with urgent work.

ROOTS OF CLOUD COMPUTING

- **Hardware Virtualization**
- Cloud computing services are usually backed by large-scale data centers composed of thousands of computers.
- Such data centers are built to serve many users and host many disparate applications.
- For this purpose, hardware virtualization can be considered as a perfect fit to overcome most operational issues of data center building and maintenance.
- A number of VMM platforms exist that are the basis of many utility or cloud computing environments.
- The most notable ones, VMWare, Xen, and KVM.

ROOTS OF CLOUD COMPUTING



A hardware virtualized server hosting three virtual machines, each one running distinct operating system and user level software stack.

ROOTS OF CLOUD COMPUTING

- **Virtual Appliances and the Open Virtualization Format**
- An application combined with the environment needed to run it (operating system, libraries, compilers, databases, application containers, and so forth) is referred to as a “virtual appliance”.
- Packaging application environments in the shape of virtual appliances eases software customization, configuration, and patching and improves portability.
- Most commonly, an appliance is shaped as a VM disk image associated with hardware requirements, and it can be readily deployed in a hypervisor.

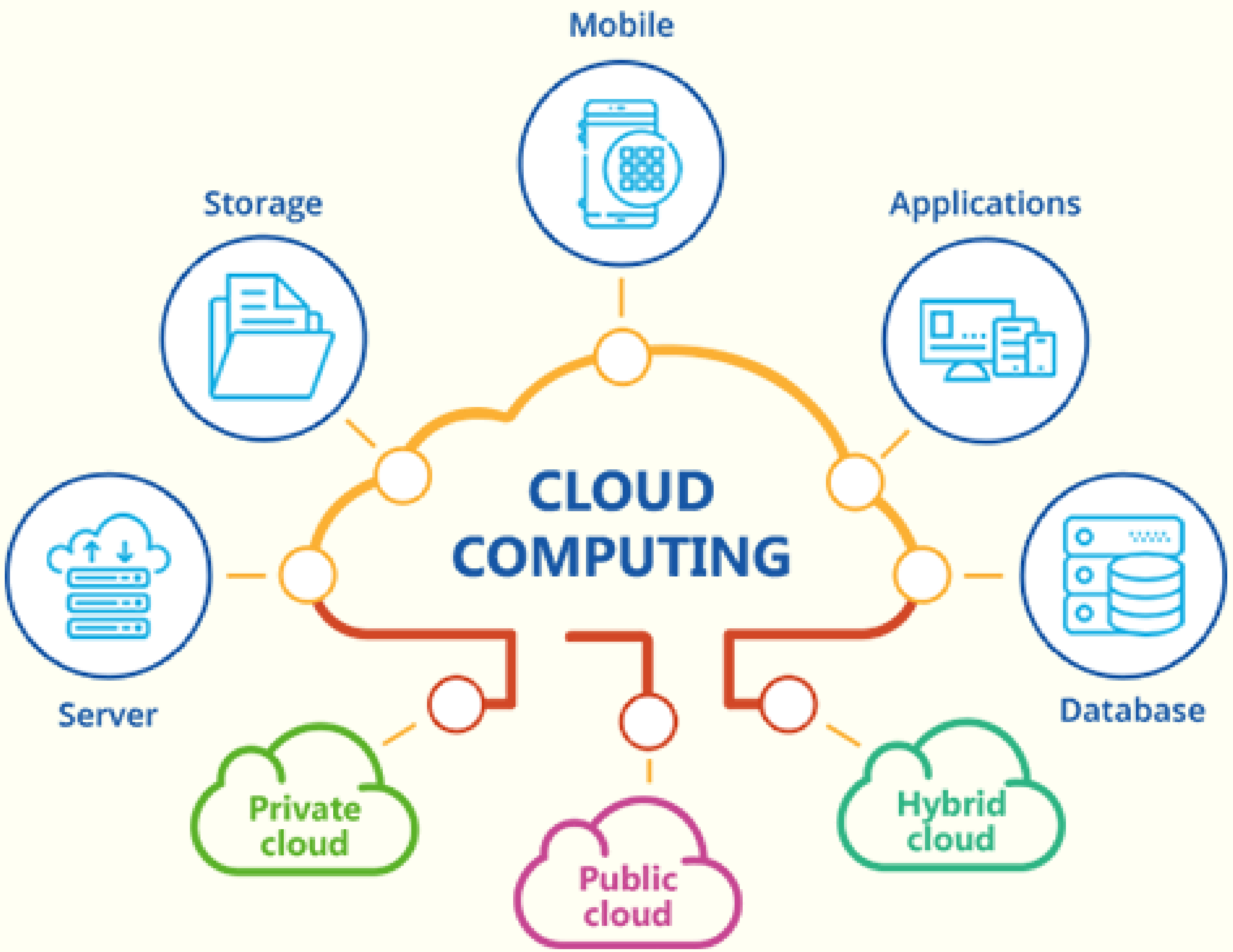
ROOTS OF CLOUD COMPUTING

- **Autonomic Computing**
- The increasing complexity of computing systems has motivated research on autonomic computing, which seeks to improve systems by decreasing human involvement in their operation.
- Autonomic, or self-managing, systems rely on monitoring probes and gauges (sensors), on an adaptation engine (autonomic manager) for computing optimizations based on monitoring data, and on effectors to carry out changes on the system.
- IBM’s Autonomic Computing Initiative has contributed to define the four properties of autonomic systems: self-configuration, self-optimization, self-healing, and self-protection.
- IBM has also suggested a reference model for autonomic control loops of autonomic managers, called MAPE-K (Monitor Analyze Plan Execute—Knowledge)

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- Distributed and cloud computing systems are built over a large number of autonomous computer nodes.
- These node machines are interconnected by SANs, LANs, or WANs in a hierarchical manner.
- With today's networking technology, a few LAN switches can easily connect hundreds of machines as a working cluster.
- A WAN can connect many local clusters to form a very large cluster of clusters.
- Massive systems are considered highly scalable, and can reach web-scale connectivity, either physically or logically.

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING



SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- Massive systems are classified into four groups:
  - Clusters
  - P2P Networks
  - Computing Grids
  - Internet clouds

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- **Clusters**
- A distributed systems cluster is a group of machines that are virtually or geographically separated and that work together to provide the same service or application to clients.
- It is possible that many of the services you run in your network today are part of a distributed systems Cluster Distributed Services:
  - Domain Naming System
  - Windows Internet Naming Service
  - Active Directory

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- **P2P Networks**
- In a P2P system, every node acts as both a client and a server, providing part of the system resources.
- Peer machines are simply client computers connected to the Internet.
- All client machines act autonomously to join or leave the system freely.
- This implies that no master-slave relationship exists among the peers. No central coordination or central database is needed.
- The system is self-organizing with distributed control.

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- **Computing Grids**
- This is the use of widely distributed computer resources to reach a common goal.
- A computing grid can be thought of as a distributed system with non-interactive workloads that involve many files.
- Grid computing is distinguished from conventional high-performance computing systems such as cluster computing in that grid computers have each node set to perform a different task/application.
- Grid computers also tend to be more heterogeneous and geographically dispersed than cluster computers.

SYSTEM MODELS FOR DISTRIBUTED AND CLOUD COMPUTING

- **Internet clouds**
- The idea is to move desktop computing to a service-oriented platform using server clusters and huge databases at data centers.
- Cloud computing leverages its low cost and simplicity to benefit both users and providers.
- Machine virtualization has enabled such cost-effectiveness.
- Cloud computing intends to satisfy many user Virtualized resources from data centers to form an Internet cloud, provisioned with hardware, software, storage, network, and services for paid users to run their applications.

LAYERS AND TYPES OF CLOUDS

- Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely:

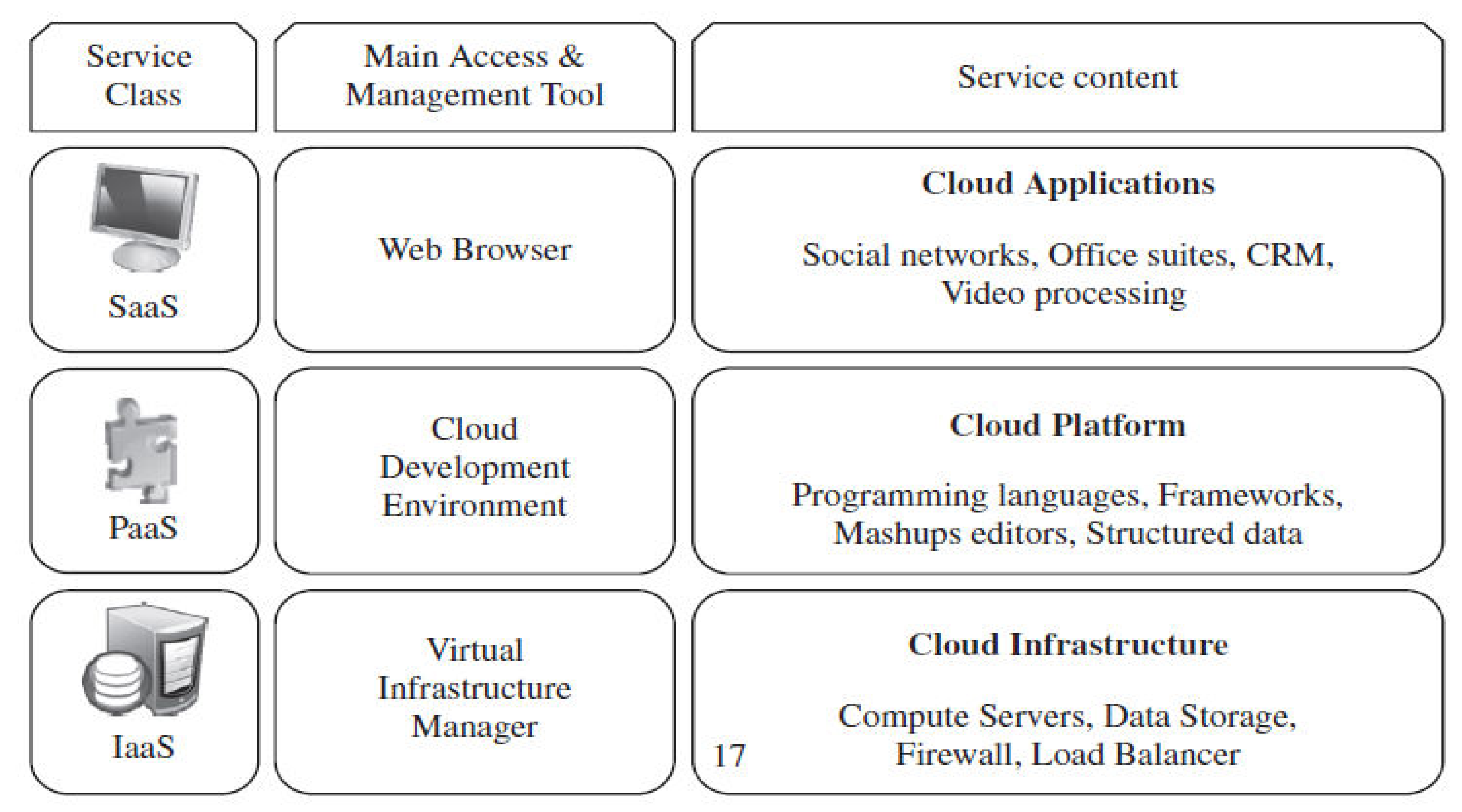
(1) Infrastructure as a Service

(2) Platform as a Service

(3) Software as a Service

(4) Deployment Models

LAYERS AND TYPES OF CLOUDS

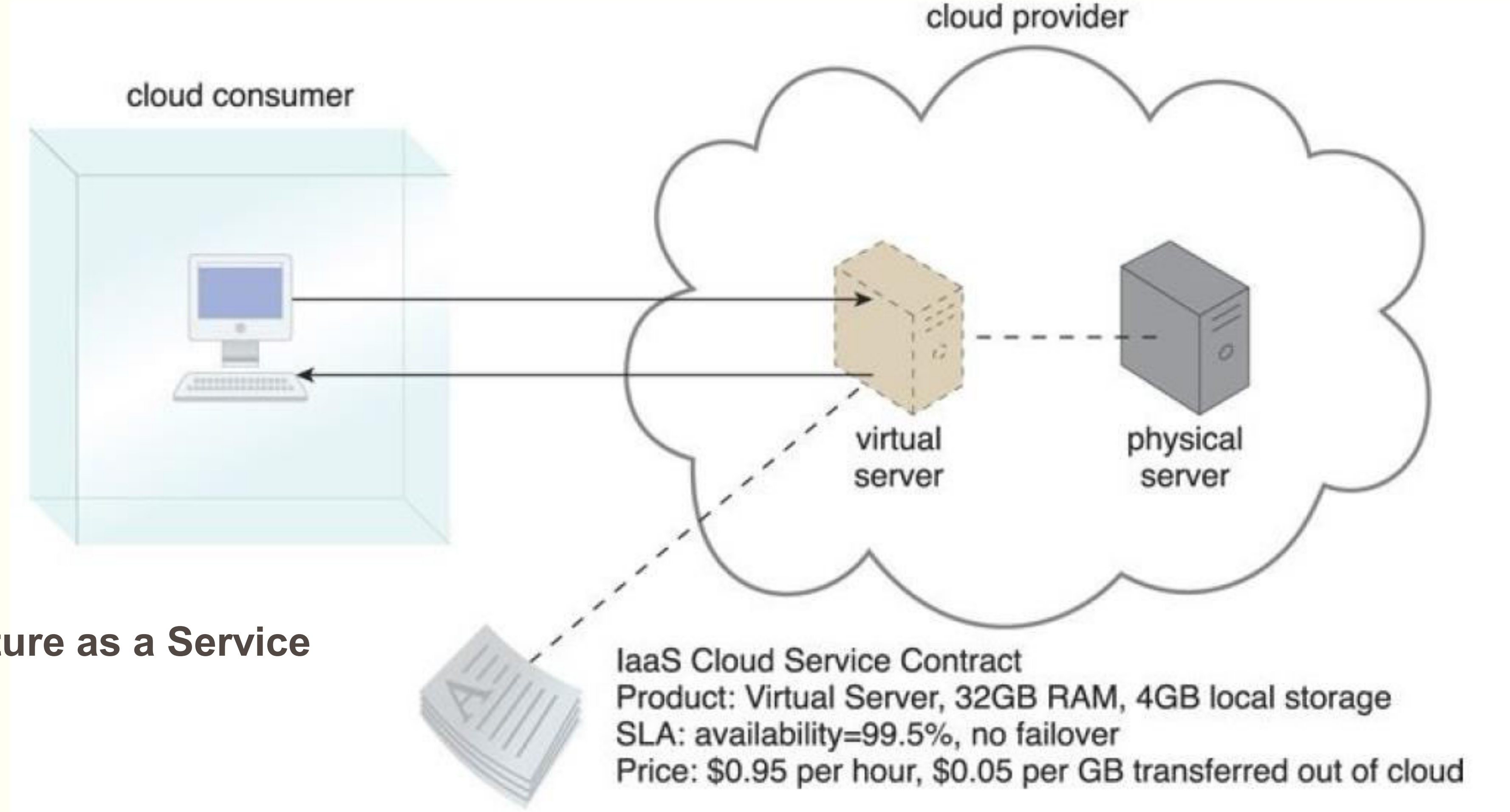


The cloud computing stack

LAYERS AND TYPES OF CLOUDS

- **Infrastructure as a Service**
- Offering virtualized resources (computation, storage, and communication) on demand is known as Infrastructure as a Service (IaaS).
- A cloud infrastructure enables on-demand provisioning of servers running several choices of operating systems and a customized software stack.
- Infrastructure services are considered to be the bottom layer of cloud computing systems.
- Amazon Web Services mainly offers IaaS, which in the case of its EC2 service means offering VMs with a software stack that can be customized similar to how an ordinary physical server would be customized.

LAYERS AND TYPES OF CLOUDS



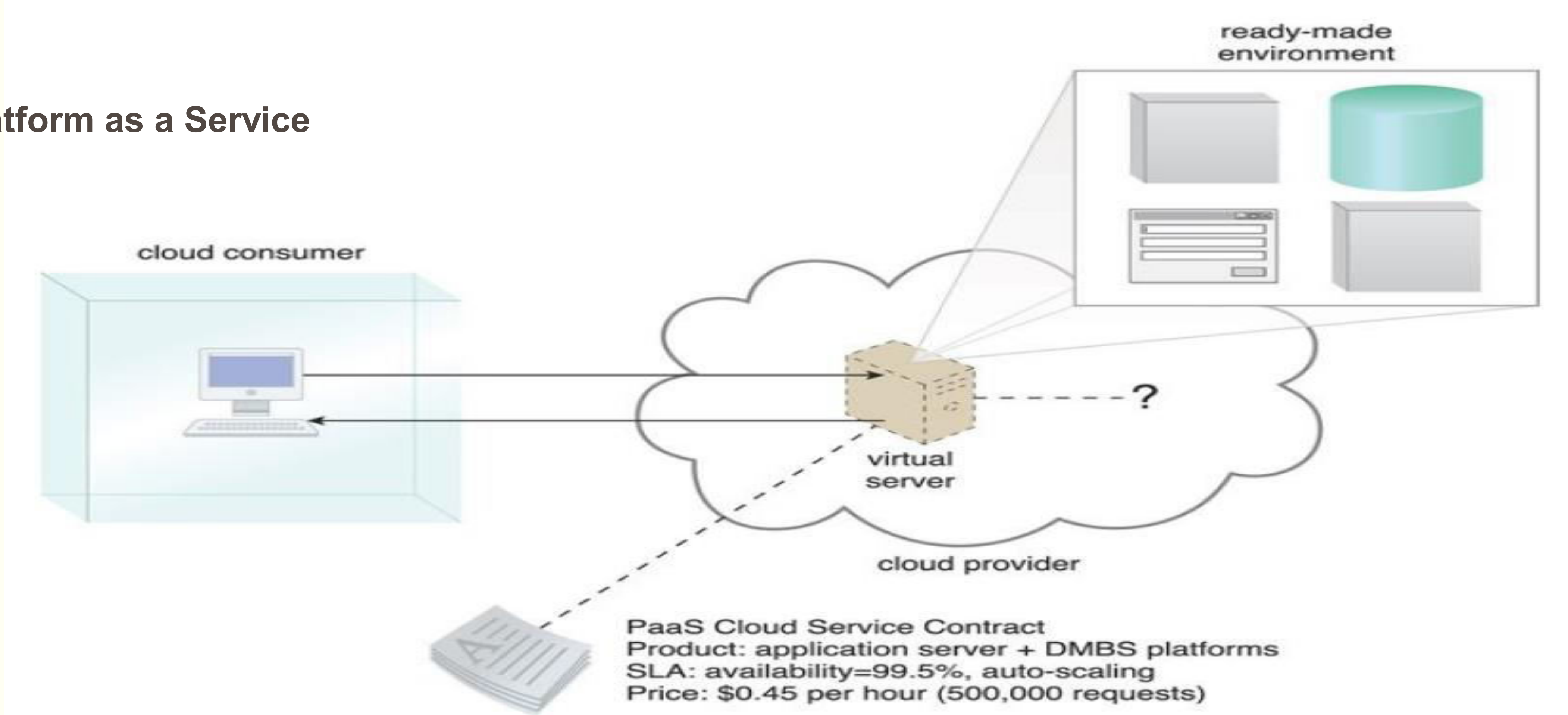
Infrastructure as a Service

LAYERS AND TYPES OF CLOUDS

- **Platform as a Service**
- In addition to infrastructure-oriented clouds that provide raw computing and storage services, another approach is to offer a higher level of abstraction to make a cloud easily programmable, known as Platform as a Service (PaaS).
- A cloud platform offers an environment on which developers create and deploy applications and do not necessarily need to know how many processors or how much memory that applications will be using.
- In addition, multiple programming models and specialized services (e.g., data access, authentication, and payments) are offered as building blocks to new applications.
- Google AppEngine, an example of Platform as a Service, offers a scalable environment for developing and hosting Web applications, which should be written in specific programming languages such as Python or Java, and use the services' own proprietary structured object data store.

LAYERS AND TYPES OF CLOUDS

Platform as a Service

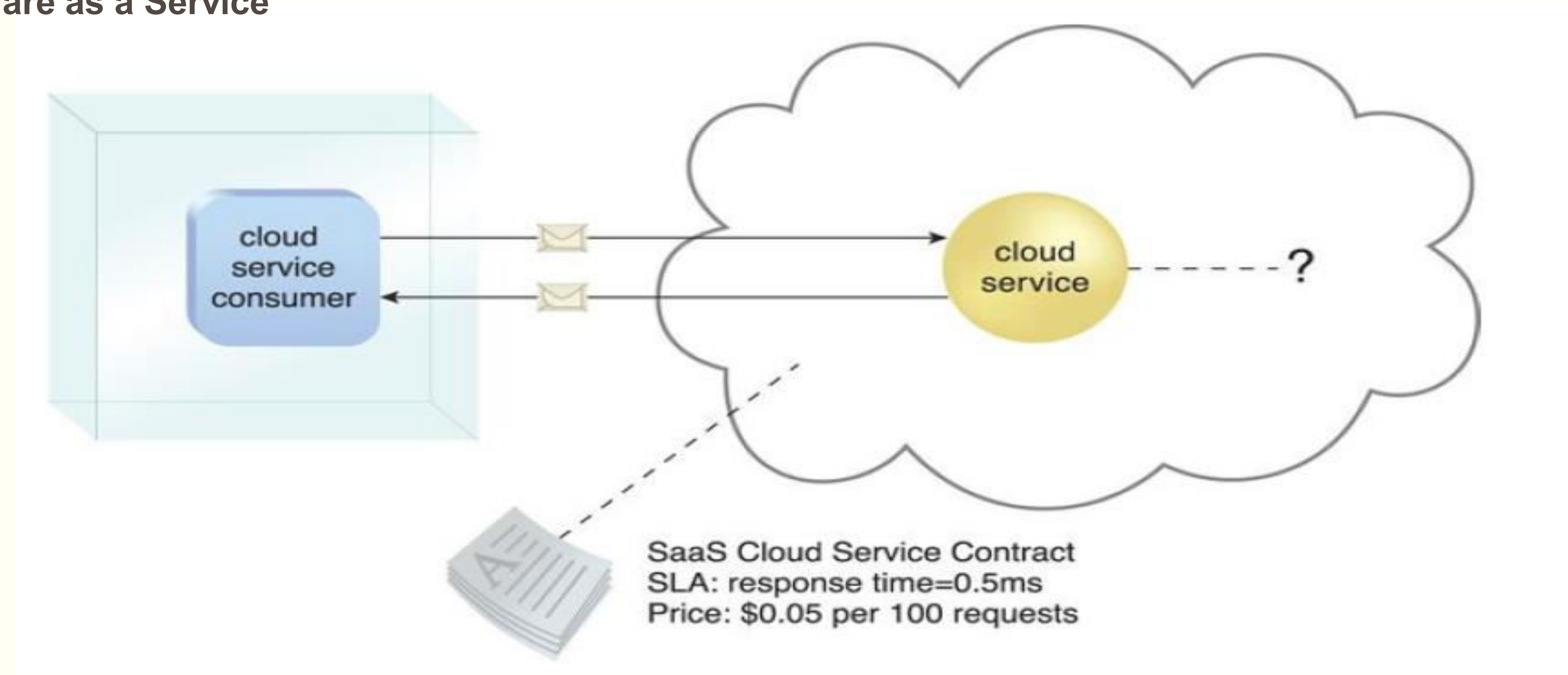


LAYERS AND TYPES OF CLOUDS

- **Software as a Service**
- Applications reside on the top of the cloud stack. Services provided by this layer can be accessed by end users through Web portals.
- Therefore, consumers are increasingly shifting from locally installed computer programs to on line software services that offer the same functionally.
- Traditional desktop applications such as word processing and spreadsheet can now be accessed as a service in the Web.
- This model of delivering applications, known as Software as a Service (SaaS), alleviates the burden of software maintenance for customers and simplifies development and testing for providers.
- Salesforce.com, which relies on the SaaS model, offers business productivity applications (CRM) that reside completely on their servers, allowing costumers to customize and access applications on demand.

LAYERS AND TYPES OF CLOUDS

Software as a Service



LAYERS AND TYPES OF CLOUDS

Cloud Delivery Model	Typical Level of Control Granted to Cloud Consumer	Typical Functionality Made Available to Cloud Consumer
SaaS	usage and usage-related configuration	access to front-end user- interface
PaaS	limited administrative	moderate level of administrative control over IT resources relevant to cloud consumer's usage of platform
IaaS	full administrative	full access to virtualized infrastructure-related IT resources and possibly, to underlying physical IT resources

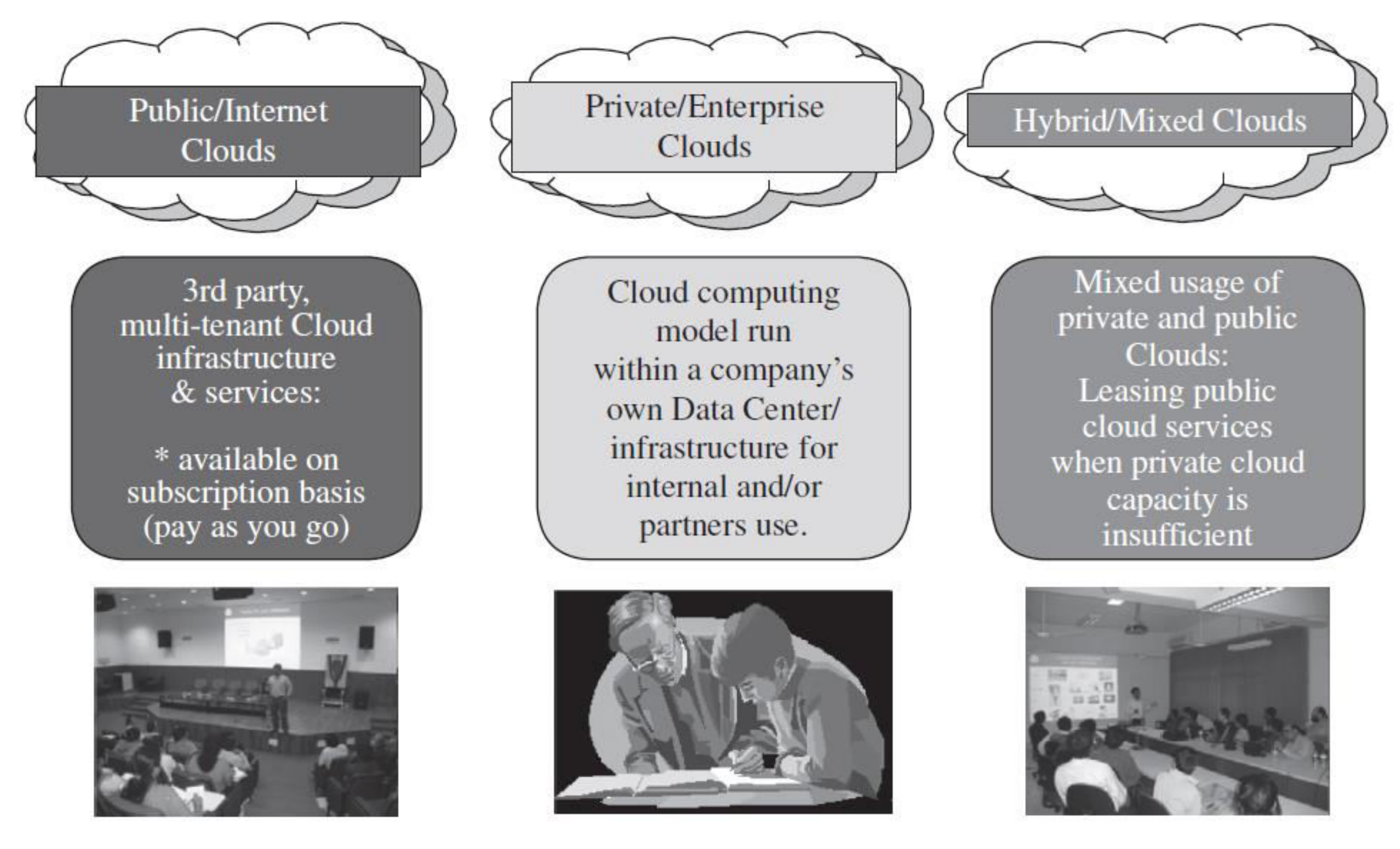
LAYERS AND TYPES OF CLOUDS

Cloud Delivery Model	Common Cloud Consumer Activities	Common Cloud Provider Activities
SaaS	uses and configures cloud service	implements, manages, and maintains cloud service Monitors usage by cloud consumers
PaaS	develops, tests, deploys, and manages cloud services and cloud-based solutions	pre-configures platform and provisions underlying infrastructure, middleware, and other needed IT resources, as necessary monitors usage by cloud consumers
IaaS	sets up and configures bare infrastructure, and installs, manage, and monitors any needed software	provisions and manages the physical processing, storage, networking, and hosting required monitors usage by cloud consumers

LAYERS AND TYPES OF CLOUDS

- **Deployment Models**
- Public
- Private
- Community
- Hybrid.

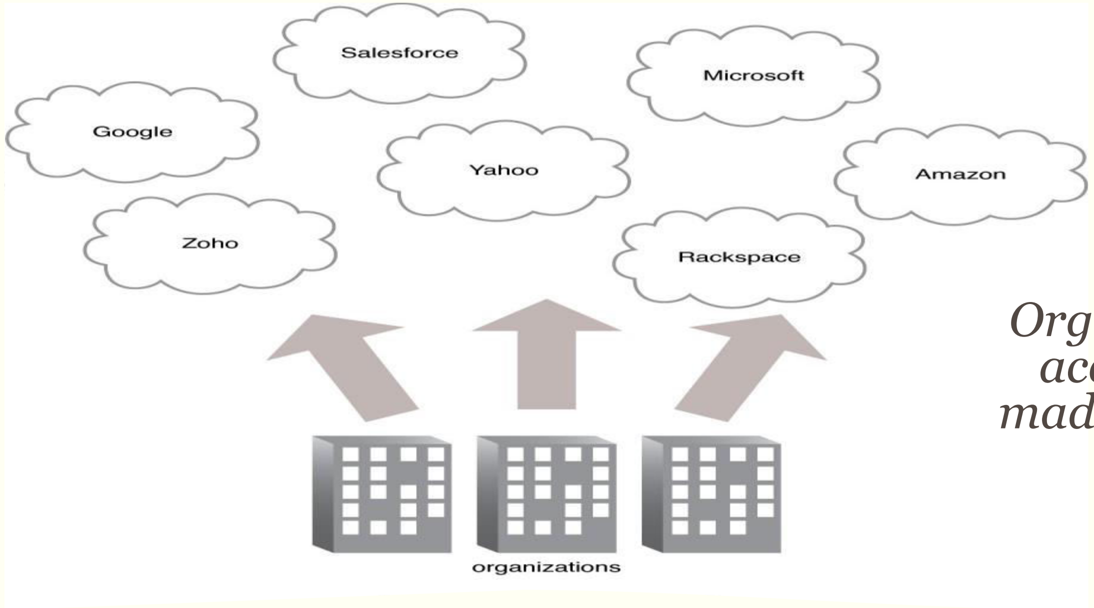
LAYERS AND TYPES OF CLOUDS



Types of clouds based on deployment models

LAYERS AND TYPES OF CLOUDS

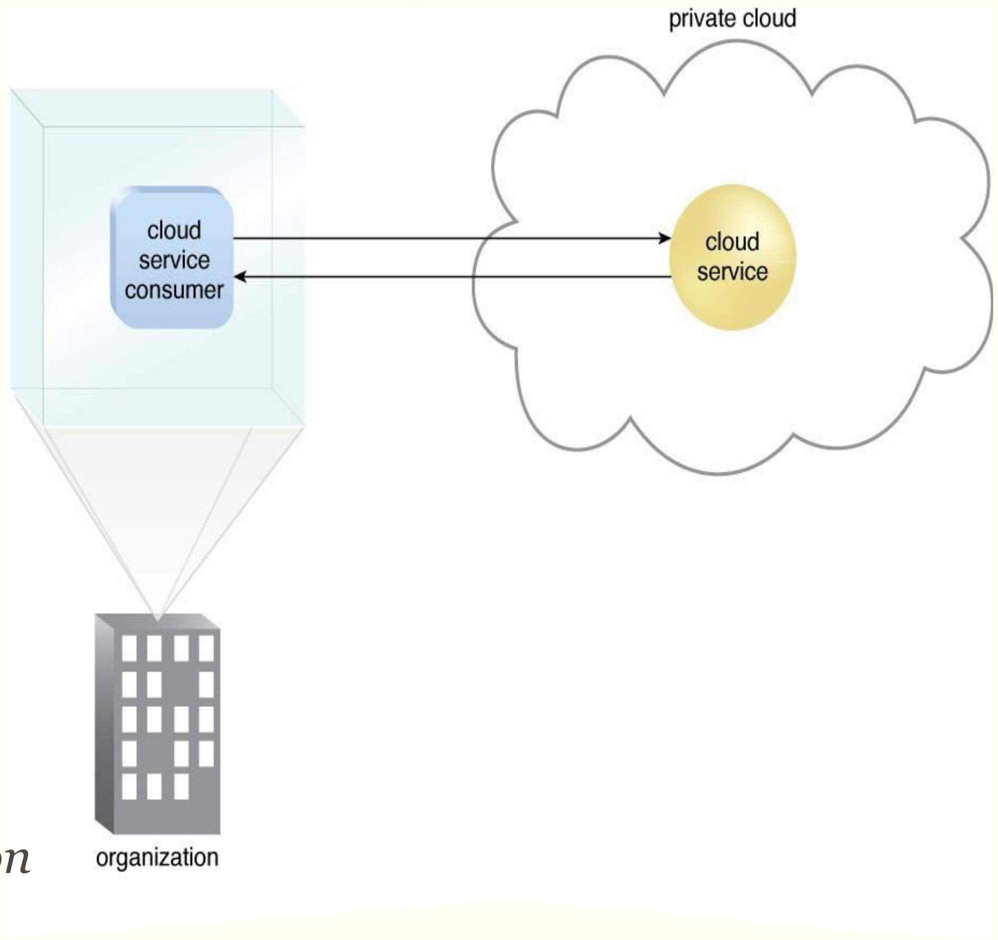
- **Public Cloud**
- public cloud as a “cloud made available in a pay-as-you-go manner to the general public” and private cloud as “internal data center of a business or other organization, not made available to the general public”.



Organizations act as cloud consumers when accessing cloud services and IT resources made available by different cloud providers.

LAYERS AND TYPES OF CLOUDS

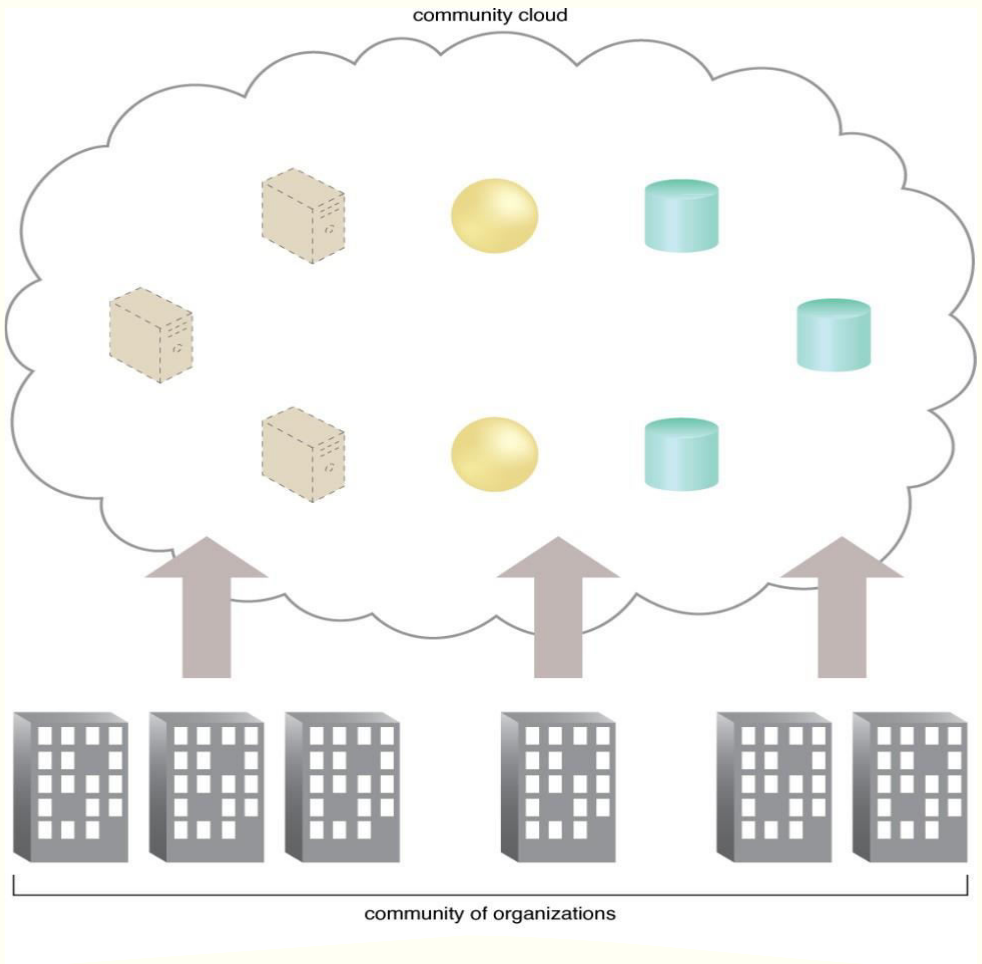
- **Private Cloud**
- Private cloud means restructuring an existing infrastructure by adding virtualization and cloud-like interfaces.
- This allows users to interact with the local data center while experiencing the same advantages of public clouds, most notably self-service interface, privileged access to virtual servers, and per-usage metering and billing.



A cloud service consumer in the organization's on-premise environment accesses a cloud service hosted on the same organization's private cloud via a virtual private network.

LAYERS AND TYPES OF CLOUDS

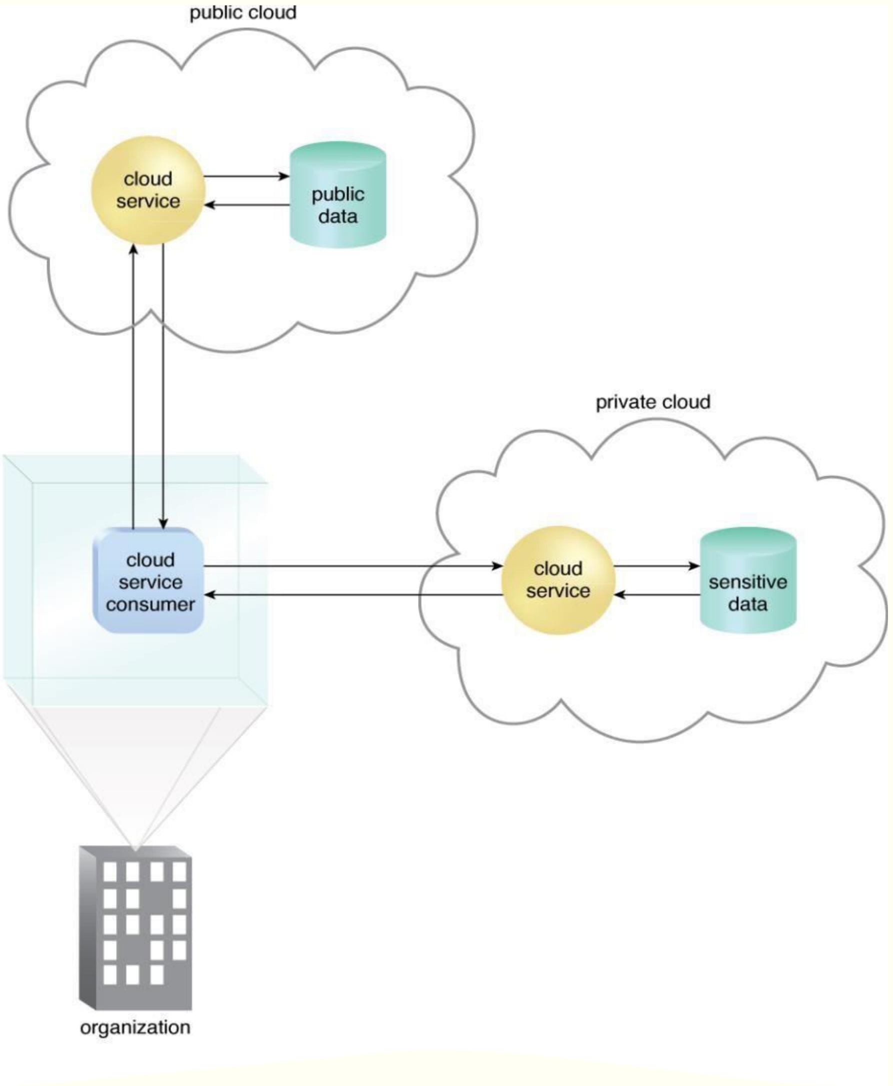
- **Community Cloud**
- A community cloud is "shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).



An example of a "community" of organizations accessing IT resources from a community cloud.

LAYERS AND TYPES OF CLOUDS

- **Hybrid Cloud**
- 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”.



An organization using a hybrid cloud architecture that utilizes both a private and public cloud.

DESIRED FEATURES OF A CLOUD

- Certain features of a cloud are essential to enable services that truly represent the cloud computing model and satisfy expectations of consumers, and cloud offerings must be

(i) Self-service

(ii) Per-usage metered and billed

(iii) Elastic

(iv) Customizable

DESIRED FEATURES OF A CLOUD

- **Self-Service**
- Consumers of cloud computing services expect on-demand, nearly instant access to resources.
- Clouds must allow self-service access so that customers can request, customize, pay, and use services without intervention of human operators.

DESIRED FEATURES OF A CLOUD

- **Per-Usage Metering and Billing**
- Cloud computing eliminates up-front commitment by users, allowing them to request and use only the necessary amount.
- Services must be priced on a short term basis (e.g., by the hour), allowing users to release (and not pay for) resources as soon as they are not needed.

DESIRED FEATURES OF A CLOUD

- **Customization**
- In a multi-tenant cloud a great disparity between user needs is often the case.
- Thus, resources rented from the cloud must be highly customizable.
- In the case of infrastructure services, customization means allowing users to deploy specialized virtual appliances and to be given privileged (root) access to the virtual servers.
- Other service classes (PaaS and SaaS) offer less flexibility and are not suitable for general-purpose computing, but still are expected to provide a certain level of customization.

CLOUD INFRASTRUCTURE MANAGEMENT

- A key challenge IaaS providers face when building a cloud infrastructure is managing physical and virtual resources, namely servers, storage, and networks, in a holistic fashion.
- The orchestration of resources must be performed in a way to rapidly and dynamically provision resources to applications.
- The software toolkit responsible for this orchestration is called a Virtual Infrastructure Manager (VIM).

CLOUD INFRASTRUCTURE MANAGEMENT

- **Features**
  - Virtualization Support
  - Self-Service, On-Demand Resource Provisioning
  - Multiple Backend Hypervisors
  - Storage Virtualization
  - Interface to Public Clouds
  - Virtual Networking
  - Dynamic Resource Allocation
  - Virtual Clusters
  - Reservation and Negotiation Mechanism
  - High Availability and Data Recovery

CLOUD INFRASTRUCTURE MANAGEMENT

- **Case Studies**
  - Apache VCL
  - AppLogic
  - Citrix Essentials
  - Enomaly ECP
  - Eucalyptus
  - Nimbus3
  - OpenNebula
  - OpenPEX
  - oVirt
  - Platform ISF - Infrastructure Sharing Facility (ISF)
  - VMWare vSphere and vCloud

INFRASTRUCTURE AS A SERVICE PROVIDERS

- Public Infrastructure as a Service providers commonly offer virtual servers containing one or more CPUs, running several choices of operating systems and a customized software stack.
- The storage space and communication facilities are often provided.
- **Features**
  - Geographic Presence
  - User Interfaces and Access to Servers
  - Advance Reservation of Capacity
  - Automatic Scaling and Load Balancing
  - Service-Level Agreement
  - Hypervisor and Operating System Choice

INFRASTRUCTURE AS A SERVICE PROVIDERS

- **Case Studies**
  - Amazon Web Services
  - Flexiscale
  - Joyent
  - GoGrid
  - Rackspace Cloud Servers

PLATFORM AS A SERVICE PROVIDERS

- Public Platform as a Service providers commonly offer a development and deployment environment that allow users to create and run their applications with little or no concern to low-level details of the platform.
- The specific programming languages and frameworks are made available in the platform, as well as other services such as persistent data storage and in memory caches.
- **Features**
  - Programming Models, Languages, and Frameworks
  - Persistence Options
- **Case Studies**
  - Aneka
  - App Engine
  - Microsoft Azure
  - Force.com
  - Heroku

CHALLENGES AND RISKS

- Despite the initial success and popularity of the cloud computing paradigm and the extensive availability of providers and tools, a significant number of challenges and risks are inherent to this new model of computing.
- Providers, developers, and end users must consider these challenges and risks to take good advantage of cloud computing.
- Issues to be faced include user privacy, data security, data lock-in, availability of service, disaster recovery, performance, scalability, energy-efficiency, and programmability.
  - Security, Privacy, and Trust
  - Data Lock-In and Standardization
  - Availability, Fault-Tolerance, and Disaster Recovery
  - Resource Management and Energy-Efficiency

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

- A computer consists of four main components, they are the input, output, storage and processing.
- They all together gather to form a computer. And again in storage device comes two branches, the primary and secondary memory.
- Primary is responsible for quick transfer of data to the processor for getting processed, and are volatile, whereas secondary is used for huge storage of data.
- The cloud storage is a new method of storing data, it is used mainly for two purposes, the first is for making a good place for data backup, and reducing the hassles of storage expansion on system.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

- Cloud computing architecture refers to the components and subcomponents required for cloud computing.
- These components typically consist of a front end platform (fat client, thin client, mobile device), back end platforms (servers, storage), a cloud based delivery, and a network (Internet, Intranet, Intercloud).
- Combined, these components make up cloud computing architecture.
- Cloud computing architectures consist of front-end platforms called clients or cloud clients.
- These clients are servers, fat (or thick) clients, thin clients, zero clients, tablets and mobile devices.
- These client platforms interact with the cloud data storage via an application (middleware), via a web browser, or through a virtual session.

REFERENCES

1. Kai Hwang, Geoffrey C Fox and Jack G Dongarra, "Distributed and Cloud Computing, From Parallel Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.
2. Barrie Sosinky,"Cloud Computing Bible", Wiley Publishing Inc,2011
3. Buyya R., Broberg J. and Goscinski A., "Cloud Computing: Principles and Paradigm", First Edition, John Wiley & Sons, 2011.
4. Rajkumar Buyya, Christian Vecchiola, S. ThamaraiSelvi,"Mastering the Cloud Computing", Morgan Kaufmann,2013
5. John W. Rittinghouse and James F. Ransome, "Cloud Computing: Implementation "Management, and Security", CRC Press, 2016.
6. David Bernstein, "Containers and Cloud: From LXC to Docker to Kubernetes", IEEE Cloud Computing, Volume: 1 , Issue: 3 , 2014.
7. VMware (white paper),"Understanding Full Virtualization, Paravirtualization, and Hardware Assist ":www.vmware.com/files/pdf/VMware_paravirtualization.pdf.

Thank You…