# INTRODUCTION TO CLOUD COMPUTING

LECTURE -1 TO LECTURE-4

# TOPICS

- Introduction to Cloud Computing: Cloud Computing in a Nutshell
- Roots of Cloud Computing,
- Layers and Types of Clouds
- Desired Features of a Cloud,
- Cloud Infrastructure Management
- IaaS,
- PaaS,
- Challenges and Risks

# WHAT IS CLOUD COMPUTING?

Cloud Computing is the **Delivery of Computing Services**—servers, storage, databases, networking, software and more—over The Internet.

Companies Offering these Computing Services are called **Cloud Providers** And Typically Charge for Cloud Computing Services Based **On Usage**, Similar to how you are billed for water or electricity at home.

---- **Pay As You Use.**

# GENERIC DEFINITIONS

- By Buyya et al.,

"Cloud is a **parallel and distributed computing** system consisting of a collection of inter-connected and virtualized computers that are **dynamically provisioned** and presented as one or more unified computing resources **based on service-level agreements (SLA)** established through negotiation between the **service provider and consumers.**"

- By Vaquero et al.,

"Clouds are a large pool of easily usable and accessible **virtualized resources** (such as hardware, development platforms and/or services). These resources can be **dynamically reconfigured** to adjust to a variable load (scale), allowing also for optimum resource utilization. This pool of resources is typically exploited by a **pay-per-use model** in which guarantees are offered by the **Infrastructure Provider by means of customized Service Level Agreements**."

# GENERIC DEFINITIONS (CONTD..)

- By National Institute of Standards and Technology (NIST)

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

- Armbrust et al.

"..data center hardware and software that provide services."

…. Many more definitions are found

# CLOUD CHARACTERISTICS

- Pay-per-use (no ongoing commitment, utility prices);
- Elastic capacity and the illusion of infinite resources;
- Self-service interface; and
- Resources that are abstracted or virtualized.

**NOTE:**

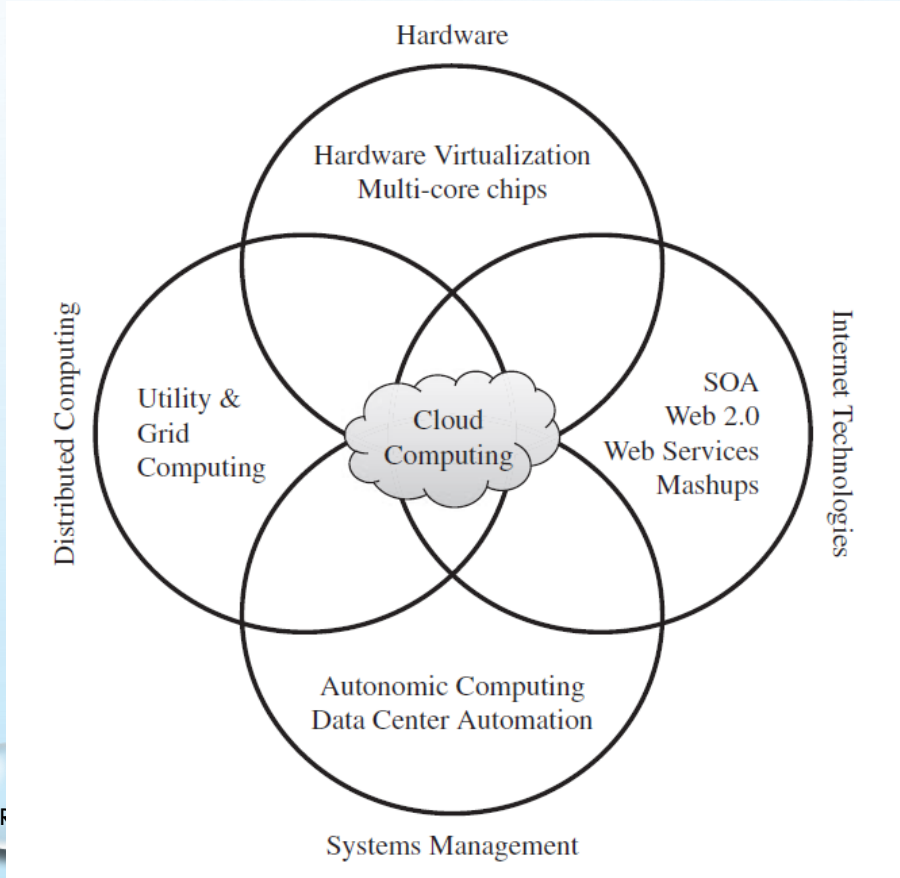
**Elasticity** is defined as "the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner

**Self-service** means that the consumer performs all the actions needed to acquire the service himself/herself, instead of going through an IT department

# ROOTS OF CLOUD COMPUTING

Roots of cloud computing can be tracked by observing the advancement in technologies….

1. Hardware (virtualization, multi-core chips)
2. Internet technologies (web services, service-oriented architectures, web 2.0)
3. Distributed computing (clusters, grids)
4. Systems management (autonomic computing, data center automation).



# ROOTS OF CC (CONTD..)

The technologies that form base of cloud computing ecosystem are:

1. From mainframes to clouds
2. SOA, web services, web 2.0 and mashups
3. Grid computing
4. Utility computing
5. Hardware virtualization
6. Virtual appliances and the open virtualization formats (OVM)
7. Autonomic computing

# MAINFRAME COMPUTERS

- Mainframes are a type of computer that generally are known for their **large size, amount of storage, processing power and high level of reliability**.
- A **single mainframe can replace** dozens or even hundreds of smaller servers.
- Widely used areas: **banking, finance, health care, insurance, public utilities, government**
- Mainframes are able to **handle large workloads**, and their **reliability, scalability, and performance** make them the system of choice for many organizations **for bulk processing, handling transactions, and for maintaining vital business records**.

# MAINFRAME COMPUTERS (CONTD..)

- Mainframes had to operate at **very high utilization rates and were very expensive**.
- The **mainframe era collapsed** with the advent of fast and **inexpensive microprocessors** .
- Next Cloud Computing have sprung up, **allowing applications to be run via third-party infrastructure** that requires little or no hardware investment and services to be paid for as they are needed.
- The **advent of fast fiber optic networks , helped enabling sharing of computing power over great distances.**

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS

Service-oriented architecture (SOA) is an evolution of distributed computing based on the request/reply design paradigm for synchronous and asynchronous applications

- What's key to these services is their loosely coupled nature; i.e., the service interface is independent of the implementation.
- For example, a service can be implemented either in .Net or J2EE, and the application consuming the service can be on a different platform or language.

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS (CONTD..)

- **WS standards have been created on top of HTTP and XML**
  - Providing a common mechanism for delivering services
  - Making them ideal for implementing a service-oriented architecture (SOA)
- **Web services can glue together applications running on different messaging product platforms**
  - Enabling information from one application to be made available to others

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS (CONTD..)

- SOA usually linked to **Web Services Description Language (WSDL)** and **Simple Object Access Protocol (SOAP)** specifications.
- The WS model of SOA uses the
  - → WSDL to **connect interfaces with services.**
  - → SOAP to **define procedure or component APIs**.
- RESTful web service -- is based on **representational state transfer** (REST) technology, an **architectural style and approach** for communications used in **web services** development.

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS (CONTD..)

- REST technology is generally preferred to the more robust technology because REST leverages less bandwidth, making it more suitable for internet usage.
  - Many service providers make their service APIs publicly accessible using standard protocols like SOAP and REST
  - Amazon, Facebook, and Google

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS (CONTD..)

- In the consumer Web, information and services may be programmatically aggregated
  - Acting as building blocks of complex compositions, called **service mashups**
  - Services like user authentication, e-mail, payroll management, and calendars are examples
    - Can be reused and combined in a business solution
  - Popular APIs are combined to produce a variety of interesting solutions
    - Google Maps, YouTube, Amazon ecommerce, and Twitter

# SOA, WEB SERVICES, WEB 2.0, AND MASHUPS (CONTD..)

A mashup is a Web page or application that uses and combines data, presentation or functionality from two or more sources to create new services.

Combination with Google Maps HousingMaps: combines rental listings (American classified advertisements website with sections devoted to jobs, housing, personals, for sale, items wanted, services, community and discussion forums.) with Google Maps for a visual representation of local apartments for rent.
Wikipedia vision: combines Google Map and a Wikipedia API

# GRID COMPUTING

- Grid Computing enables aggregation of distributed resources .
- It is a type of parallel and distributed system that enables the **sharing, exchange, selection, & aggregation of geographically distributed resources** depending on their availability, capability, cost, and user QoS requirements.
- **Open Grid Service Architecture(OGSA)** – helps in standardization by defining a set of core capabilities and behaviors.

# GRID COMPUTING (CONTD..)

**Issues in grid**

→ Guaranteeing execution time for critical applications

→Availability of resources with diverse configuration

Virtualization technology helped in finding the solution of some of the issues.

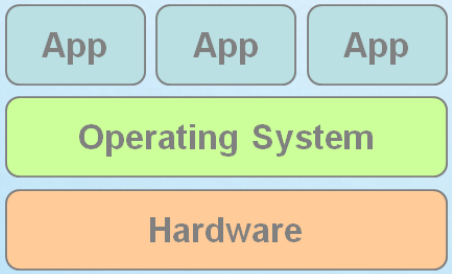
- **Some characteristics of Grid…**
  - Numerous
  - Owned and managed by different, multiple organisations and individuals.
  - Have different security requirements and policies
  - Heterogeneous
  - Connected by heterogeneous, multilevel networks
  - Have different resource management policies
  - Likely to be geographically separated

# UTILITY COMPUTING

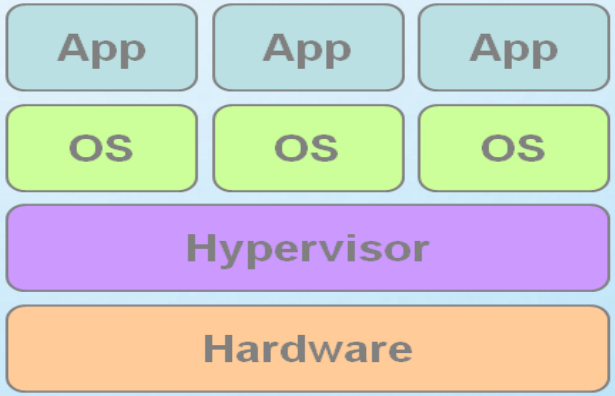
- In utility computing environments, users assign a "utility" value to their jobs, where utility is a fixed or time-varying valuation that captures various QoS constraints (deadline, importance, satisfaction).
- The valuation is the amount they are willing to pay a service provider to satisfy their demands.
- Utility computing is a computing business model in which the provider owns, operates and manages the computing infrastructure and resources, and the subscribers accesses it as and when required on a rental or metered basis.

# 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.
- Virtualization enhances :sharing and utilization of computer systems




Traditional Stack




Virtualized Stack

# HARDWARE VIRTUALIZATION (CONTD..)

- H/w virtualization allows running multiple OS and a s/w stacks on a single physical platform
- VMM (virtual machine monitor) also called hypervisor mediates access to physical hardware presenting each guest OS a VM
- A number of VMM platforms like VMWare, Xen…. that are the basis of many utility or cloud computing environments.

**Benefits**

→Sharing and utilization
→Better manageability
→Higher reliability

# HARDWARE VIRTUALIZATION (CONTD..)

Three basic capabilities regarding management of workload:

- **Workload isolation-**A virtual machine should not be able to read RAM that is in use by another virtual machine. It also should not be able to access another virtual machine's disk.
- **Workload migration** - encapsulating guest OS system state within VM and allowing it to suspend or migrate to different platforms and resumed immediately…( facilitates h/w maintenance, load balancing and disaster recovery).
- **Workload consolidation** – Consolidation of several individual and heterogenous workloads onto a single platform leads to better system utilization.

# 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.
- In other words, systems should manage themselves, with high-level guidance from humans
- 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

# LAYERS AND TYPES OF CLOUD SERVICES

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

  - **Infrastructure as a Service (IaaS):** hardware is provided by an external provider and managed for the customers
  - **Platform as a Service (PaaS):** in addition to hardware, your operating system layer is managed …..
  - **Software as a Service (SaaS):** further to the above, an application layer is provided and managed for you

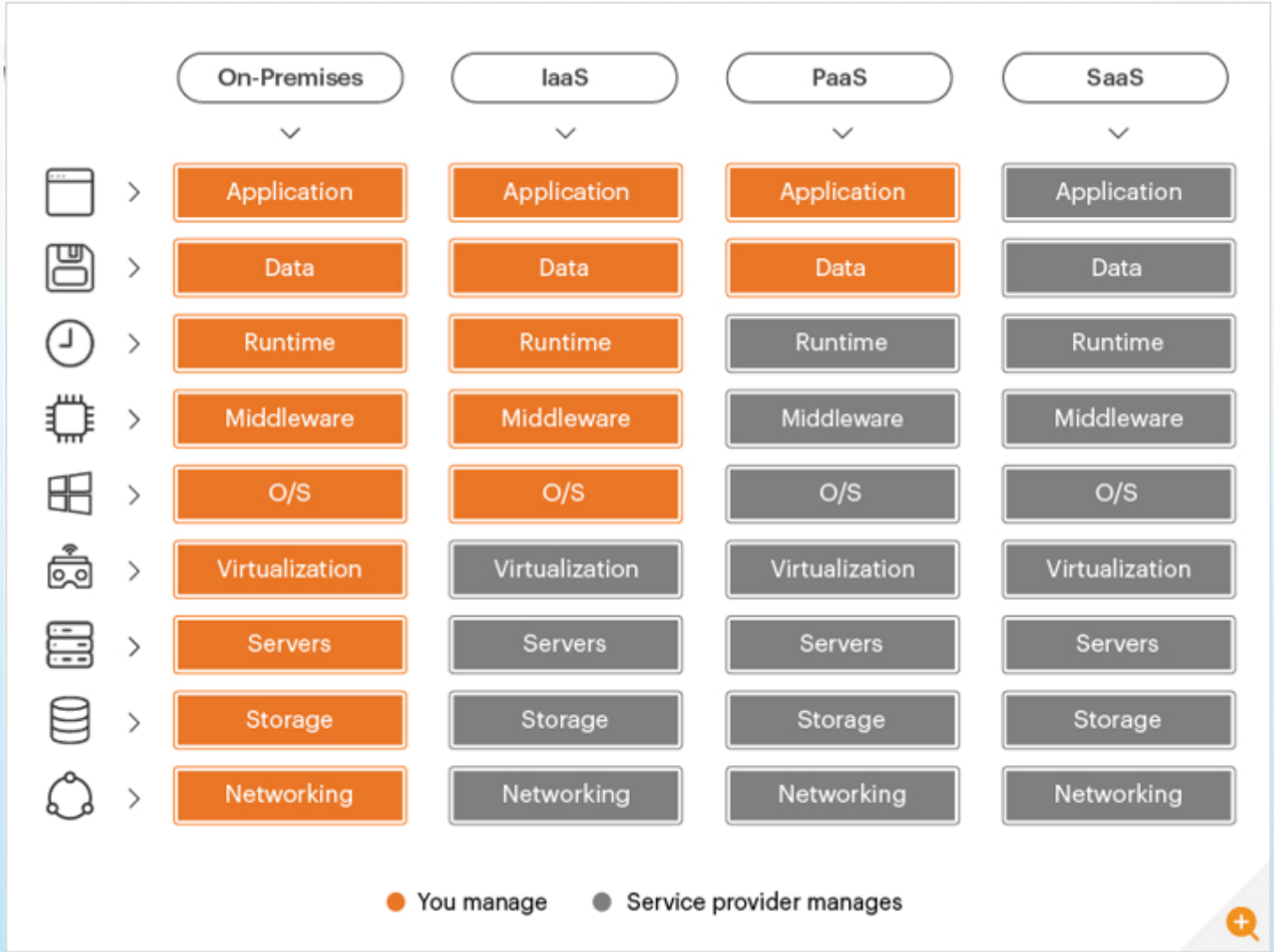
Also called as **The cloud computing stack**.

# 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 |

Fig: Cloud computing stack

# THE CLOUD COMPUTING STACK




https://www.eginnovations.com/blog/saas-vs-paas-vs-iaas-examples-differences-how-to-choose/

# INFRASTRUCTURE AS A SERVICE (IaaS)

- 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 and deployed applications**, and possibly limited control of **networking components** .
- The **cloud computing service provider** manages the infrastructure, - customers - install, configure and manage their own software like — operating systems, middleware and applications

# IaaS (contd..)

- **Examples :**

- **Amazon EC2** - Amazon Elastic Compute Cloud (EC2) - allowing users to rent virtual computers on which to run their own computer applications.
- **OpenNebula** - OpenNebula is a cloud computing platform for managing heterogeneous distributed data center infrastructures
- **Amazon Web Services (AWS)** and **Google Cloud Platform (GCP)** are examples of independent IaaS providers

# PLATFORM AS A SERVICE (PaaS)

- 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**.
- Apart from servers, storage and networking—it also includes **middleware, development tools, business intelligence (BI) services, database management systems** and more…
- Can avoid the expense and complexity of buying software licenses, and managing the underlying application infrastructure and middleware or the development tools and other resources

# PaaS (contd..)

- **Examples :**

- **Microsoft Windows Azure** - Microsoft Azure is a growing collection of cloud services for building, deploying and managing applications through global network of datacentres.
- **Google App Engine-** Google App Engine is a fully managed platform that completely abstracts away infrastructure so user can focus only on code.
- **Hadoop**- Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment.

# SOFTWARE AS A SERVICE (SaaS)

- 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).
- Often referred to as "**on-demand software**".
- Don't want to worry about the installation, setup and running of the application. Service provider will do that.

# SaaS (contd..)

- **Examples :**
  - Google Apps (e.g., Gmail, Google Docs, Google sites, …etc)
  - Google Apps, Microsoft Office 365.
  - Applications like email (Gmail, Yahoo mail etc), Social Networking sites (Facebook etc)

# CLOUD DEPLOYMENT MODELS

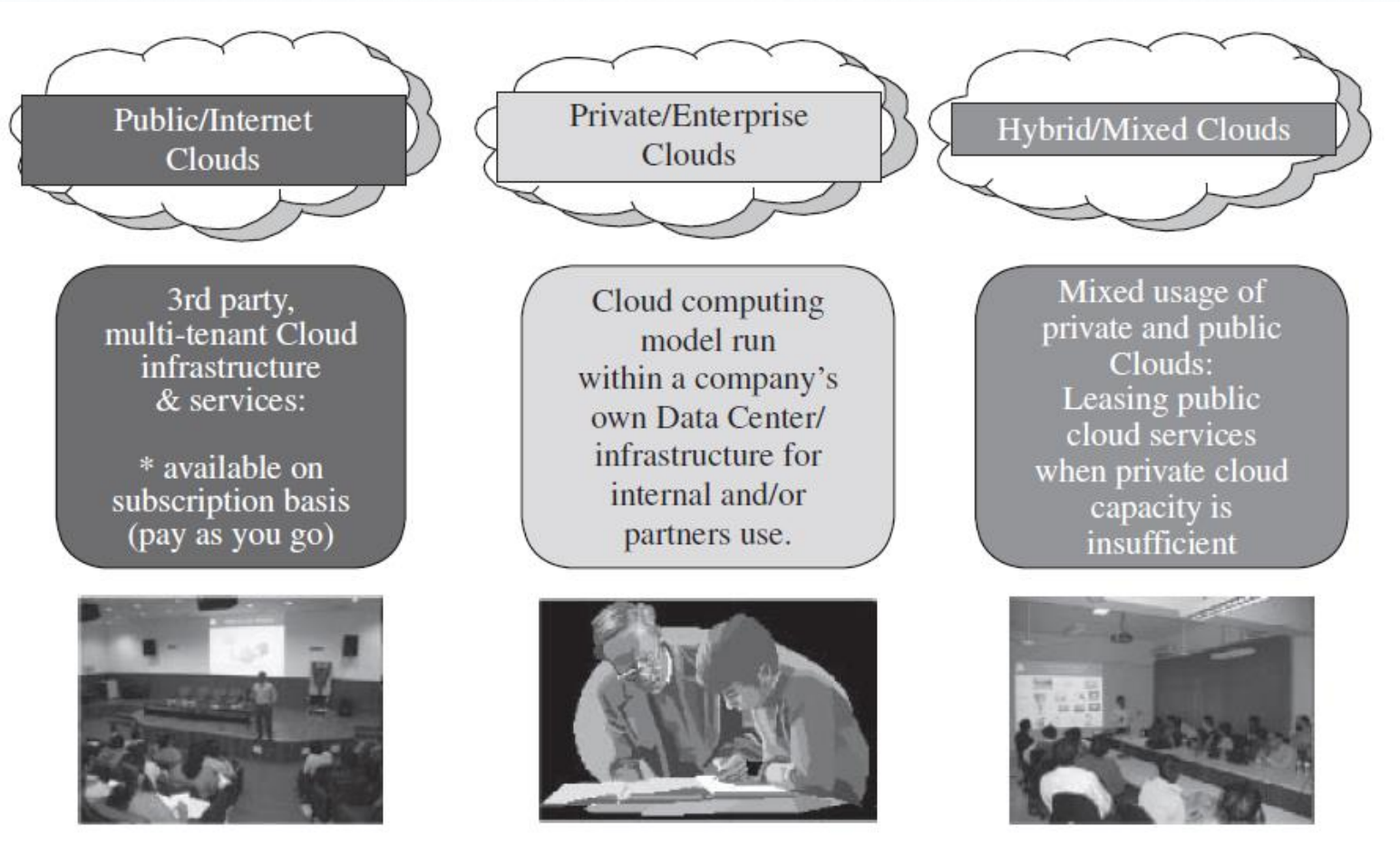
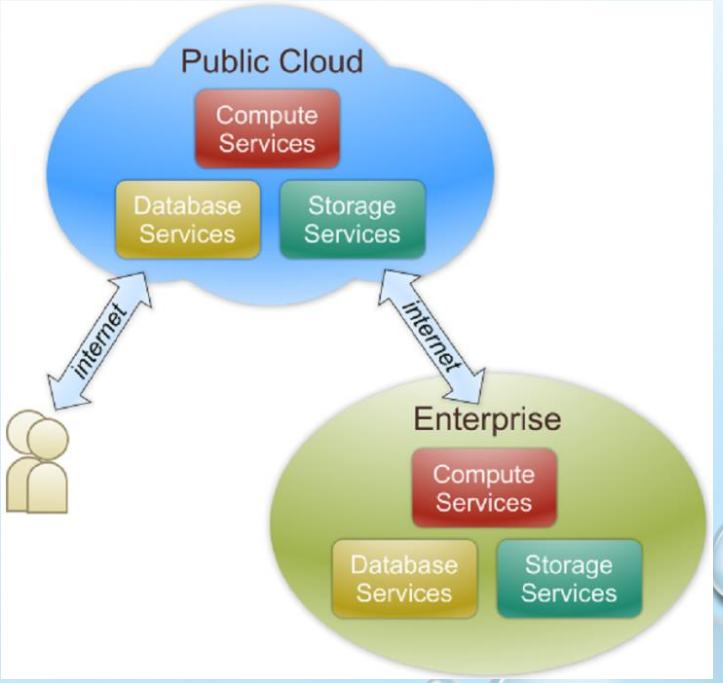



Fig. Types of clouds based on deployment models

# PUBLIC CLOUD

**Public cloud definition:**

- 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 known as external cloud or multi-tenant cloud, this model essentially represents a cloud environment that is openly accessible.
- **Basic characteristics :**
  - Homogeneous infrastructure
  - Common policies
  - Shared resources and multi-tenant
  - Leased or rented infrastructure
  - Economies of scale




**Examples:**

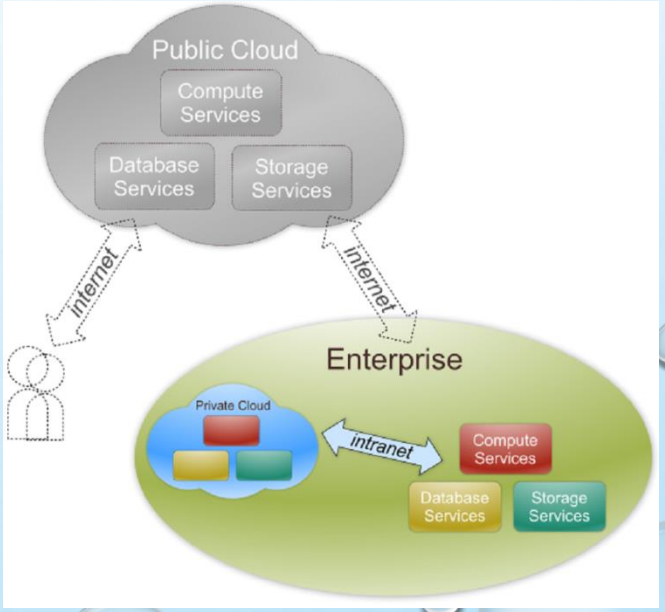
Amazon Elastic Compute Cloud (EC2), IBM's Blue Cloud, Sun Cloud, Google AppEngine and Windows Azure Services Platform.

Most of the google services: Gmail, Google Doc, Spreadsheet

# PRIVATE CLOUD

**Private cloud definition**

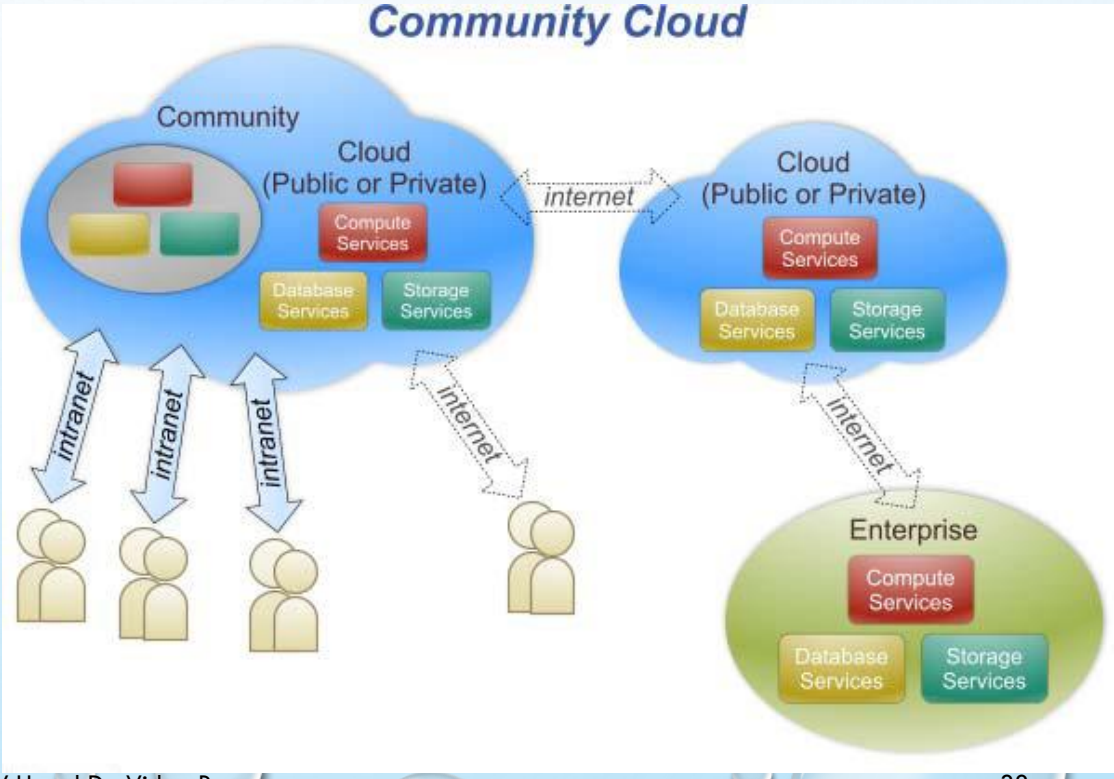
- The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- Also referred to as internal cloud or on-premise cloud, a private cloud intentionally limits access to its resources to service consumers that belong to the same organization that owns the cloud.
- Basic characteristics :
  - Heterogeneous infrastructure
  - Customized and tailored policies
  - Dedicated resources
  - In-house infrastructure
  - End-to-end control



# COMMUNITY CLOUD

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

Community clouds are often designed for businesses and organizations working on joint projects, applications, or research, which requires a central cloud computing facility for building, managing and executing such projects.

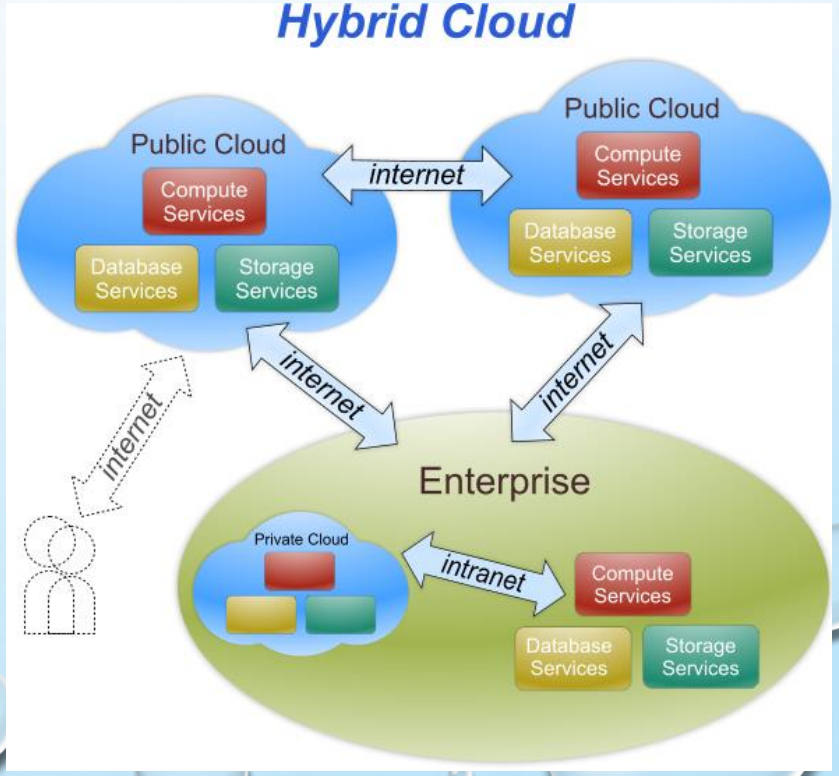


# HYBRID CLOUD

**Hybrid cloud definition:**

The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

- An example of a hybrid cloud solution is an organization that wants to keep confidential information secured on their private cloud, but make more general, customer-facing content on a public cloud.
- Cloud bursting is an application deployment model in which an application runs in a private cloud and burst into a public cloud when the demand for computing capacity spikes.



# FEATURES OF CLOUD

- Self Service
- Per usage metering and billing
- Elasticity
- Customization

# FEATURES OF CLOUD (CONTD..)

## ➢ Self Service

Consumers of cloud computing services expect on-demand

- Nearly instant access to resources

Clouds must allow self-service access

- Customers can request, customize, pay, and use services without intervention of human operators

## ➢ 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

Clouds must allow efficient trading of service

- Such as pricing, accounting, and billing

# FEATURES OF CLOUD (CONTD..)

## ➢ Elasticity

Cloud computing gives the illusion of infinite computing resources available on demand

- Users expect clouds to rapidly provide resources in any quantity at any time
- The additional resources can be
  - Provisioned, possibly automatically, when an application load increases
  - Released when load decreases
  - Scale up and down

## ➢ Customization

A multi-tenant cloud often reveals a great disparity between user needs

- Resources rented from the cloud must be highly customizable

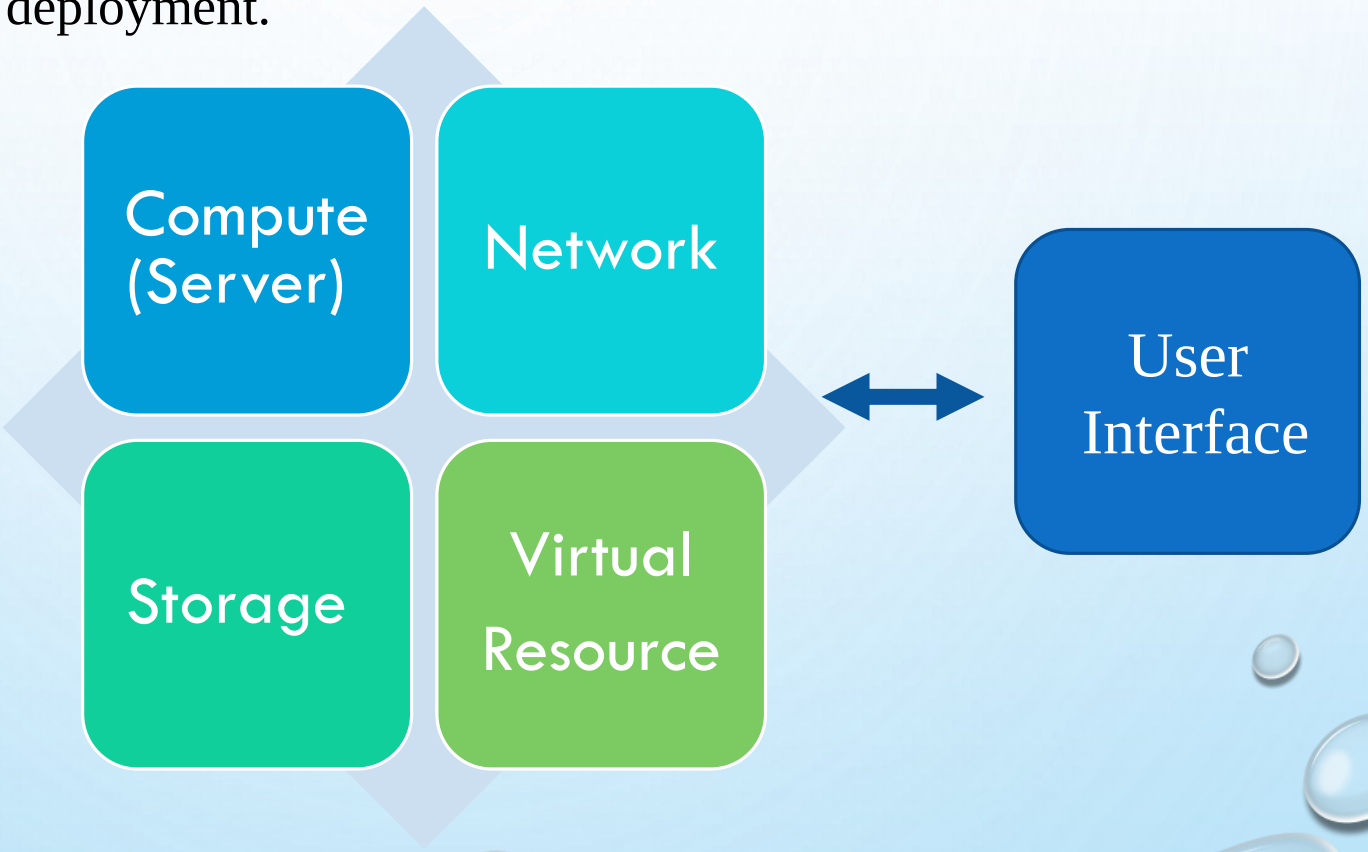
Infrastructure services allow users

- To deploy specialized virtual appliances
- To be given privileged (root) access to the virtual servers

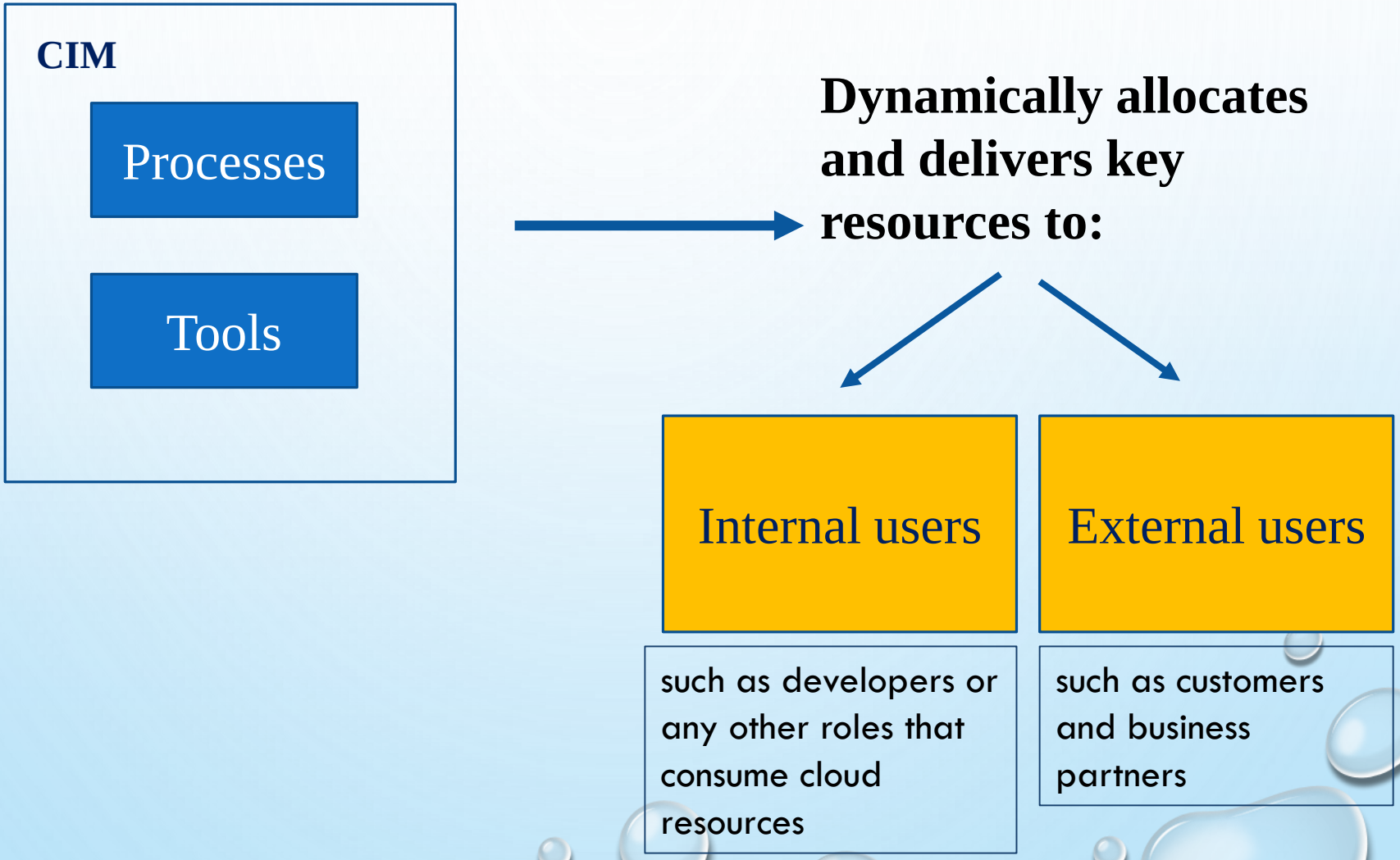
PaaS and SaaS offer less flexibility

# CLOUD INFRASTRUCTURE MANAGEMENT

**Cloud Infrastructure** includes all software and hardware resources needed for cloud deployment.



# CLOUD INFRASTRUCTURE MANAGEMENT(CIM) (CONTD..)



# CLOUD INFRASTRUCTURE MANAGEMENT (CONTD..)

**Why Cloud Infrastructure Management?**

- Without appropriate visibility, monitoring and governance, cloud computing costs can increase unnecessarily.
- A typical scenario would be an engineer who leaves a cloud development environment up and running 24/7, even if they only need it for several hours of work.
- In a pay-as-you-go model—which is common in Infrastructure-as-a-Service platforms—that kind of waste can lead to runaway cloud bills.

**What does Cloud Infrastructure Management do?**

- It's like a central nervous system.
- Tracks cloud usage
- Enables businesses to create, configure, scale and reuse.
- Maximises operational flexibility
- Maintains agility with cost-effectiveness
- Preserves privacy and security of transactions

# CLOUD INFRASTRUCTURE MANAGEMENT (CONTD..)

Cloud infrastructure management **tools** as per 2021:

- Apache CloudStack
- BMC Helix Cloud Security
- CloudHealth by VMware
- Microsoft Azure Management Tools
- Morpheus by Morpheus Data
- Terraform Enterprise by HashiCorp
- Turbonomic

https://www.bmc.com/blogs/hybrid-cloud-management-tools-how-to-choose/

# CLOUD INFRASTRUCTURE MANAGEMENT (CONTD..)

Factors to be considered while choosing the right cloud management tools include:

- Resource management
- Performance monitoring
- Scalability
- Automation and provisioning
- Cross-platform interoperability
- Compliance and governance
- Reporting

# CLOUD INFRASTRUCTURE MANAGEMENT

Features available in Virtual Infrastructure Manager(VIM):

- 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

# FEATURES

**Virtualization Support**

The multi-tenancy aspect of clouds requires **multiple customers with disparate requirements** to be served by a single hardware infrastructure

- Virtualized resources (CPUs, memory, etc.) Can be sized and resized with certain flexibility

Hardware virtualization

- The ideal technology to create a virtual infrastructure that partitions a data center among multiple tenants

# FEATURES (CONTD..)

**Self-Service, On-Demand Resource Provisioning**

- Enables users to directly obtain services from clouds
  - e.g., spawning the creation of a server
  - Tailoring its software, configurations, and security policies
  - Without interacting with a human system administrator
- Eliminates the need for more time-consuming, labor-intensive, human-driven procurement processes
- Users can easily interact with the system

# FEATURES (CONTD..)

**Multiple Backend Hypervisors**

- Some VI managers provide a uniform management layer regardless of the virtualization technology used
- Usually provide pluggable drivers to interact with multiple hypervisors

**Storage Virtualization**

Abstracting logical storage from physical storage

Consolidating all available storage devices in a data center to create virtual disks independent from device and location

Storage devices are commonly organized in a storage area network (SAN) attached to servers

Via protocols such as Fibre Channel, ISCSI, and NFS

A storage controller provides the layer of abstraction between virtual and physical storage

# FEATURES (CONTD..)

**Multiple Backend Hypervisors**

- Some VI managers provide a uniform management layer regardless of the virtualization technology used
- Usually provide pluggable drivers to interact with multiple hypervisors

**Storage Virtualization**

Abstracting logical storage from physical storage

Consolidating all available storage devices in a data center to create virtual disks independent from device and location

Storage devices are commonly organized in a storage area network (SAN) attached to servers

Via protocols such as Fibre Channel, ISCSI, and NFS

A storage controller provides the layer of abstraction between virtual and physical storage

# FEATURES (CONTD..)

## Interface to Public Clouds

- Extends the capacity by borrowing resources from public clouds
  - Make good use of the available resources
  - In case of spikes in demand, extra load can be offloaded to rented resources
- A VI manager can be used in a hybrid cloud setup
  - Offers a driver to manage the life cycle of virtualized resources obtained from external cloud providers
  - Must ideally be transparent

## Virtual Clusters

- Several VI managers can holistically manage groups of VMs
- Useful for provisioning computing virtual clusters on demand
- Interconnecting VMs for multi-tier Internet applications

# FEATURES (CONTD..)

## Virtual Networking

- Virtual networks allow creating an isolated network on top of a physical infrastructure
  - independently from physical topology and locations
- A virtual LAN (VLAN) allows isolating traffic that shares a switched network
  - Allowing VMs to be grouped into the same broadcast domain
  - Can be configured to block traffic originated from VMs from other networks
- The VPN (virtual private network) concept is a secure and private overlay network on top of a public network
  - Most commonly the public Internet
- VI managers support creating and configuring virtual networks to group VMs placed throughout a data center

# FEATURES (CONTD..)

## Dynamic Resource Allocation

- Increased awareness of energy consumption in data centers has encouraged dynamic consolidating VMs in a fewer number of servers
- Applications have variable and dynamic needs
- Also reallocates available resources among VMs according to application needs
- Energy consumption reduction and better management of SLAs can be achieved
  - By dynamically remapping VMs to physical machines at regular intervals
  - Machines that are not assigned any VM can be turned off or put on a low power state

# FEATURES (CONTD..)

## Reservation and Negotiation Mechanism

- Requests are termed advance reservations (AR)
  Users request computational resources to be available at a specific time
- In contrast to best-effort requests
  Users request resources whenever available
  The provider can offer a distinct slot that is still satisfactory to the user
- OpenPEX( Open Provisioning and execution system) incorporates a bilateral negotiation protocol

  - OpenPEX, a system that allows users to provision resources ahead of time through advance reservations
  - Allows users and providers to come to an alternative agreement by exchanging offers and counter offers

# FEATURES (CONTD..)

## High Availability (HA) and Data Recovery

- High availability feature minimizing application downtime and preventing business disruption
- Providing a failover mechanism
  - Detects failure of both physical and virtual servers and restarts VMs on healthy physical servers
  - Protects failures from host
- Frequent backup of a large number of VMs should be done with minimal interference in the systems performance
  - Each one is attached with multiple virtual disks

# IaaS FEATURES

- Geographic presence (build data centers)
- User Interface and access to users
- Advance reservation of capacity
- Automatic Scaling and Load Balancing
- Service level Agreement

# IaaS FEATURES (CONTD..)

- Geographic presence (build data centers)

  A provider of worldwide services would typically build several data centers distributed around the world

  To improve availability and responsiveness

# IaaS FEATURES (CONTD..)

## User Interfaces and Access to Servers

- A public IaaS provider must provide multiple access means to its cloud
  - For various users and their preferences
- Different types of user interfaces (UI) provide different levels of abstraction
  - The most common being graphical user interfaces (GUI), command-line tools (CLI), and Web service (WS) APIs
- GUIs are preferred by end users who need to launch, customize, and monitor a few virtual servers
  - Do not necessary need to repeat the process several times
- WS APIs offer programmatic access to a cloud using standard HTTP requests
  - Allowing complex services to be built on top of IaaS clouds

# IaaS FEATURES (CONTD..)

## Advance Reservation of Capacity

- Allow users to reserve resources for a specific time frame in the future
  - Cloud resources will be available at that time
- Most clouds only support best-effort requests
  - Users requests are served whenever resources are available
- Amazon Reserved Instances is a form of advance reservation of capacity
  - Allowing users to pay a fixed amount of money in advance to guarantee resource availability at anytime during an agreed period
  - Then paying a discounted hourly rate when resources are in use
  - Only long periods of 1 to 3 years are offered
  - Users cannot express their reservations in finer granularities, e.g., hours or days

# IaaS FEATURES (CONTD..)

**Automatic Scaling and Load Balancing**

- Elasticity is a key characteristic of the cloud
  - Applications often need to scale up and down to meet varying load conditions
- Automatic scaling is a highly desirable feature of IaaS clouds
  - Allow users to set conditions for when they want their applications to scale up and down
  - Based on application-specific metrics such as transactions per second, number of simultaneous users, request latency, etc.
- Incoming traffic must be automatically distributed among the available servers
  - Enables applications to promptly respond to traffic increase while also achieving greater fault tolerance

# IaaS FEATURES (CONTD..)

**Service-Level Agreement**

Offered by IaaS providers to express their commitment to delivery of a certain QoS

To customers it serves as a warranty

Include availability and performance guarantees

Metrics must be agreed upon by all parties as well as penalties for violating these expectations

# PAAS FEATURES

- Programming Models, Languages and Frameworks – how user can express their application and how efficiently they run in cloud platform.
- Persistence Options – to record their states and recover in case of crashes.

# PAAS FEATURES (CONTD..)

## Programming Models, Languages, and Frameworks

- Programming models made available by IaaS providers
  Define how users can express their applications using higher levels of abstraction and efficiently run them on the cloud platform
- The most common activities that require specialized models are
  MapReduce model: Processing of large dataset in clusters of computers
  Development of request-based Web services and applications
  High-performance distributed execution of various computational tasks
- PaaS providers usually support multiple programming languages
  Python and Java (e.g., Google appengine)
  .NET languages (e.g., Microsoft Azure)
  Ruby (e.g., Heroku)

# PaaS FEATURES (CONTD..)

## Persistence Options

- A persistence layer is essential to allow applications to record their state
  - Recover it in case of crashes
  - Also store user data
- Web and enterprise application developers have chosen relational databases as the preferred persistence method
  - Offer fast and reliable structured data storage and transaction processing
  - May lack scalability to handle several petabytes ($2^{50}$) of data stored in commodity computers
- Distributed storage technologies have emerged
  - Seek to be robust and highly scalable
  - At the expense of relational structure and convenient query languages

# CHALLENGES AND RISK

- Security , Privacy and Trust
- Data Lock in and standardization
- Availability , Fault Tolerance and Disaster Recovery
- Resource Management and Energy

# CHALLENGES AND RISKS (CONTD..)

A significant number of challenges and risks are inherent to cloud computing

- User privacy, data security
- Availability of service, disaster recovery
- Performance, scalability, energy-efficiency
- Programmability

# SECURITY, PRIVACY, AND TRUST

**Information security is a main issue**

- Current cloud offerings are essentially public
  - Exposing the system to more attacks
- Need to make cloud computing environments as secure as in-house IT systems
  - e.g., data encryption, VLANs, and firewalls
- The trust toward providers is fundamental
  - To ensure the desired level of privacy for applications hosted in the cloud

# SECURITY, PRIVACY, AND TRUST

**Legal and regulatory issues also need attention**

- Providers may choose to locate data anywhere
  - The physical location of data centers determines the set of laws applied to the management of data
- Specific cryptography techniques can not be used
  - They are not allowed in some countries
- Country laws can impose that sensitive data are to be stored within national borders
  - e.g., patient health records

# DATA LOCK-IN AND STANDARDIZATION

**A major concern of cloud computing users**

- Having their data locked-in by a certain provider
  - Users may want to move data and applications out from a provider that does not meet their requirements
  - Cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications.
  - User data are not portable

# DATA LOCK-IN AND STANDARDIZATION

## Open Virtual Format (OVF)

- OVF is a specification that describes an open-standard, secure, efficient, portable and extensible format for packaging and distributing software for virtual machines.
  - Hardware virtualization
  - Aims at facilitating packing and distribution of software to be run on VMs
    - Virtual appliances can be made portable
    - Seamlessly run on hypervisor of different vendors
    - The OVF standard is independent of any particular hypervisor or processor architecture.

# AVAILABILITY, FAULT-TOLERANCE, AND DISASTER RECOVERY

- Users will have certain expectations about the service level to be provided
  - Once their applications are moved to the cloud
  - Availability of the service, its overall performance
  - What measures are to be taken when something goes wrong in the system or its components
  - Users seek for a warranty before they can comfortably move their business to the cloud
- SLAs must be ideally set up between customers and cloud computing providers
  - To act as warranty

# AVAILABILITY, FAULT-TOLERANCE, AND DISASTER RECOVERY

- Including QoS requirements
- Specifies the details of the service to be provided
  - Including availability and performance guarantees
- Metrics must be agreed upon by all parties
- Penalties for violating the expectations must also be approved

# RESOURCE MANAGEMENT AND ENERGY-EFfiCIENCY

One important challenge is the efficient management of virtualized resource pools

- Physical resources like CPU cores, disk space, and network bandwidth must be sliced and shared among virtual machines
  - Running potentially heterogeneous workloads
- Dimensions to be considered include
  - Number of CPUs, amount of memory, size of virtual disks, and network bandwidth

# RESOURCE MANAGEMENT AND ENERGY-EFfiCIENCY

Data centers consume large amounts of electricity

- 100 server racks can consume 1.3 MW of power
- Another 1.3 MW are required by the cooling system
  - Costing USD 2.6 million per year
- Data centers significantly impact the environment
  - In terms of CO2 emissions from the cooling systems

# NEXT SESSION….

VIRTUALIZATION AND HYPERVISORS