CLOUD COMPUTING, UNIT-1: INTRODUCTION

Definition: "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. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

Cloud: A cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources.

Protocols: Protocols refer to standards and methods that allow computers to communicate with each other in a pre-defined and structured manner. A cloud can be based on the use of any protocols that allow for the remote access to its IT resources.

IT-Resource: An IT resource is a physical or virtual IT-related artifact that can be either software-based, such as a virtual server or a custom software program, or hardware-based, such as a physical server or a network device.

On-Premise: An IT resource that is hosted in a conventional IT enterprise within an organizational boundary (that does not specifically represent a cloud) is considered to be located on the premises of the IT enterprise, or on-premise for short.

Scaling: Scaling, from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands.

Horizontal Scaling (more common): The allocating or releasing of IT resources that are of the same type is referred to as horizontal scaling. The horizontal allocation of resources is referred to as scaling out and the horizontal releasing of resources is referred to as scaling in.

Vertical Scaling (less common): When an existing IT resource is replaced by another with higher or lower capacity, vertical scaling is considered to have occurred. Specifically, the replacing of an IT resource with another that has a higher capacity is referred to as scaling up and the replacing an IT resource with another that has a lower capacity is considered scaling down.

Horizontal Scaling	*Vertical Scaling*
less expensive	more expensive
IT resources instantly available	IT resources instantly available
resource replication, automated scaling	additional setup is normally required
additional IT resources needed	no additional IT resources needed
not limited by hardware capacity	limited by maximum hardware capacity

Cloud Service: A cloud service is any IT resource that is made remotely accessible via a cloud.

Cloud Service Customer: The cloud service consumer is a temporary runtime role assumed by a software program when it accesses a cloud service.

Goals and Benefits

1. **Reduced Investments and Proportional Costs**
   A cloud's Measured Usage characteristic represents a feature-set that allows measured operational expenditures (directly related to business performance) to replace anticipated capital expenditures. This is also referred to as proportional costs.
   Common measurable benefits to cloud consumers include:
   • On-demand access to pay-as-you-go computing resources on a short-term basis (such as processors by the hour), and the ability to release these computing resources when they are no longer needed.
   • The perception of having unlimited computing resources that are available on demand, thereby reducing the need to prepare for provisioning.
   • The ability to add or remove IT resources at a fine-grained level, such as modifying available storage disk space by single gigabyte increments.
   • Abstraction of the infrastructure so applications are not locked into devices or locations and can be easily moved if needed

2. **Increased Scalability**
   By providing pools of IT resources, along with tools and technologies designed to leverage them collectively, clouds can instantly and dynamically allocate IT resources to cloud consumers, on-demand or via the cloud consumer's direct configuration. This empowers cloud consumers to scale their cloud-based IT resources to accommodate processing fluctuations and peaks automatically or manually. Similarly, cloud-based IT resources can be released (automatically or manually) as processing demands decrease.

3. **Increased Availability and Reliability**
   - An IT resource with increased availability is accessible for longer periods of time (for example, 22 hours out of a 24 hour day). Cloud providers generally offer "resilient" IT resources for which they can guarantee high levels of availability.
   - An IT resource with increased reliability can better avoid and recover from exception conditions. The modular architecture of cloud environments provides extensive failover support that increases reliability.

Risks and Challenges

1. **Increased Security Vulnerabilities**
   The remote usage of IT resources requires an expansion of trust boundaries by the cloud consumer to include the external cloud. Another consequence

of overlapping trust boundaries relates to the cloud provider's privileged access to cloud consumer data. The overlapping of trust boundaries and the increased exposure of data can provide malicious cloud consumers (human and automated) with greater opportunities to attack IT resources and steal or damage business data.

2. **Reduced Operational Governance Control**
   Cloud consumers are usually allotted a level of governance control that is lower than that over on-premises IT resources. This can introduce risks associated with how the cloud provider operates its cloud, as well as the external connections that are required for communication between the cloud and the cloud consumer.
   Consider the following examples:
   - An unreliable cloud provider may not maintain the guarantees it makes in the SLAs that were published for its cloud services. This can jeopardize the quality of the cloud consumer solutions that rely on these cloud services.
   - Longer geographic distances between the cloud consumer and cloud provider can require additional network hops that introduce fluctuating latency and potential bandwidth constraints.

   Legal contracts, when combined with SLAs, technology inspections, and monitoring, can mitigate governance risks and issues.

3. **Limited Portability between Cloud Providers**
   Portability is a measure used to determine the impact of moving cloud consumer IT resources and data between clouds. For cloud consumers that have custom-built solutions with dependencies on these proprietary environments, it can be challenging to move from one cloud provider to another.

4. **Multi-Regional Compliance and Legal Issues**
   Third-party cloud providers will frequently establish data centres in affordable or convenient geographical locations. A potential legal issue pertains to the accessibility and disclosure of data. Countries have laws that require some types of data to be disclosed to certain government agencies or to the subject of the data.

Roles and Boundaries

1. **Cloud Provider**: The organization that provides cloud-based IT resources is the cloud provider. When assuming the role of cloud provider, an organization is responsible for making cloud services available to cloud consumers, as per agreed upon SLA guarantees.

2. **Cloud Consumer**: A cloud consumer is an organization (or a human) that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider. Specifically, the cloud consumer uses a cloud service consumer to access a cloud service.

3. **Cloud Service Owner**: The person or organization that legally owns a cloud service is called a cloud service owner. The cloud service owner can be the cloud consumer, or the cloud provider that owns the cloud within which the cloud service resides.

4. **Cloud Resource Administrator**: A cloud resource administrator is the person or organization responsible for administering a cloud-based IT resource (including cloud services). The cloud resource administrator can be (or belong to) the cloud consumer or cloud provider of the cloud within which the cloud service resides. Alternatively, it can be (or belong to) a third-party organization contracted to administer the cloud-based IT resource.

5. **Cloud Auditor**: A third-party (often accredited) that conducts independent assessments of cloud environments assumes the role of the cloud auditor. The typical responsibilities associated with this role include the evaluation of security controls, privacy impacts, and performance.

6. **Cloud Broker**: This role is assumed by a party that assumes the responsibility of managing and negotiating the usage of cloud services between cloud consumers and cloud providers. Mediation services provided by cloud brokers include service intermediation, aggregation, and arbitrage.

7. **Cloud Carrier**: The party responsible for providing the wire-level connectivity between cloud consumers and cloud providers assumes the role of the cloud carrier. This role is often assumed by network and telecommunication providers.

Organizational Boundary: An organizational boundary represents the physical perimeter that surrounds a set of IT resources that are owned and governed by an organization. The organizational boundary does not represent the boundary of an actual organization, only an organizational set of IT assets and IT resources. Similarly, clouds have an organizational boundary.

Trust Boundary: A trust boundary is a logical perimeter that typically spans beyond physical boundaries to represent the extent to which IT resources are trusted.

Cloud Characteristics

1. **On-Demand Usage**
   A cloud consumer can unilaterally access cloud-based IT resources giving the cloud consumer the freedom to self-provision these IT resources. Once configured, usage of the self-provisioned IT resources can be automated, requiring no further human involvement by the cloud consumer or cloud provider.

2. **Ubiquitous Access**
   Ubiquitous access represents the ability for a cloud service to be widely accessible.

3. **Multitenancy and Resource Pooling**
   The characteristic of a software program that enables an instance of the program to serve different consumers (tenants) whereby each is isolated from the other, is referred to as multitenancy. A cloud provider pools its IT resources to serve multiple cloud service consumers by using multitenancy models that frequently rely on the use of virtualization technologies. Using multitenancy technology, IT resources can be dynamically assigned and reassigned, according to cloud service consumer demands.
   Resource pooling allows cloud providers to pool large-scale IT resources to serve multiple cloud consumers.

4. **Elasticity**
   Elasticity is the automated ability of a cloud to transparently scale IT resources, as required in response to runtime conditions or as pre-determined by the cloud consumer or cloud provider.

5. **Measured Usage**
   The measured usage characteristic represents the ability of a cloud platform to keep track of the usage of its IT resources, primarily by cloud consumers. Based on what is measured, the cloud provider can charge a cloud consumer only for the IT resources actually used and/or for the timeframe during which access to the IT resources was granted.

6. **Resiliency**
   Resilient computing is a form of failover that distributes redundant implementations of IT resources across physical locations. IT resources can be pre-configured so that if one becomes deficient, processing is automatically handed over to another redundant implementation. Within cloud computing, the characteristic of resiliency can refer to redundant IT resources within the same cloud (but in different physical locations) or across multiple clouds.

Cloud Delivery Models: A cloud delivery model represents a specific, pre-packaged combination of IT resources offered by a cloud provider. Three common cloud delivery models have become widely established and formalized:

1. **Infrastructure-as-a-Service (IaaS)**
   The IaaS delivery model represents a self-contained IT environment comprised of infrastructure-centric IT resources that can be accessed and managed via cloud service-based interfaces and tools.
   The general purpose of an IaaS environment is to provide cloud consumers with a high level of control and responsibility over its configuration and utilization. The IT resources provided by IaaS are generally not pre-configured, placing the administrative responsibility directly upon the cloud consumer. This model is therefore used by cloud consumers that require a high level of control over the cloud-based environment they intend to create.

2. **Platform-as-a-Service (PaaS)**

The PaaS delivery model represents a pre-defined "ready-to-use" environment typically comprised of already deployed and configured IT resources. Specifically, PaaS relies on (and is primarily defined by) the usage of a ready-made environment that establishes a set of pre-packaged products and tools used to support the entire delivery lifecycle of custom applications.
By working within a ready-made platform, the cloud consumer is spared the administrative burden of setting up and maintaining the bare infrastructure IT resources provided via the IaaS model. Conversely, the cloud consumer is granted a lower level of control over the underlying IT resources that host and provision the platform.

3. **Software-as-a-Service (SaaS)**
   A software program positioned as a shared cloud service and made available as a "product" or generic utility represents the typical profile of a SaaS offering. The SaaS delivery model is typically used to make a reusable cloud service widely available (often commercially) to a range of cloud consumers. A cloud consumer is generally granted very limited administrative control over a SaaS implementation.

Model	*Level of control granted to Cloud Customer*	*Functionality made available to Cloud Customer*
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 customer's usage of platform
IaaS	full administrative	full access to virtualized infrastructure related IT resources and possibly to underlying physical resources
Model	*Common cloud consumer activities*	*Common cloud provider activities*
SaaS	uses and configures cloud service	implements, manages, and maintains cloud service
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
IaaS	sets up and configures bare infrastructure, and installs, manages, and monitors any needed software	provisions and manages physical processing, storage, networking, and hosting required

(All cloud providers also monitor usage by the cloud customers)

Combining Cloud Delivery Models

1. **IaaS + PaaS**: A PaaS environment will be built upon an underlying infrastructure comparable to the physical and virtual servers and other IT

resources provided in an IaaS environment. The motivation for such an arrangement may be influenced by economics or maybe because the first cloud provider is close to exceeding its existing capacity by serving other cloud consumers. Or perhaps a particular cloud consumer imposes a legal requirement for data to be physically stored in a specific region (different from where the first cloud provider's cloud resides).

2. **IaaS + PaaS + SaaS**: All three cloud delivery models can be combined to establish layers of IT resources that build upon each other. For example, by adding on to the preceding layered architecture shown in Figure 4.15, the ready-made environment provided by the PaaS environment can be used by the cloud consumer organization to develop and deploy its own SaaS cloud services that it can then make available as commercial products

Cloud Deployment Models: A cloud deployment model represents a specific type of cloud environment, primarily distinguished by ownership, size, and access.

1. **Public Clouds**: A public cloud is a publicly accessible cloud environment owned by a third-party cloud provider.

2. **Community Clouds**: A community cloud is similar to a public cloud except that its access is limited to a specific community of cloud consumers. The community cloud may be jointly owned by the community members or by a third-party cloud provider that provisions a public cloud with limited access.

3. **Private Clouds**: A private cloud is owned by a single organization. Private clouds enable an organization to use cloud computing technology as a means of centralizing access to IT resources by different parts, locations, or departments of the organization.
   With a private cloud, the same organization is technically both the cloud consumer and cloud provider. To differentiate these roles:
   - A separate organizational department typically assumes the responsibility for provisioning the cloud (and therefore assumes the cloud provider role).
   - Departments requiring access to the private cloud assume the cloud consumer role.

4. **Hybrid Cloud**: A hybrid cloud is a cloud environment comprised of two or more different cloud deployment models. For example, a cloud consumer may choose to deploy cloud services processing sensitive data to a private cloud and other, less sensitive cloud services to a public cloud.

5. **Virtual Cloud**: Also known as a "dedicated cloud" or "hosted cloud," this model results in a self-contained cloud environment hosted and managed by a public cloud provider and made available to a cloud consumer.

6. **Inter Cloud**: This model is based on an architecture comprised of two or more inter-connected clouds.