# **Definition of Cloud Computing**

Cloud Computing is a **model**. SOA is an architecture. Cloud Computing is not associated with any specific technology, protocol or vendor. It allows cloud applications to service the consumers (via websites, client applications) assuring that the service will have a single point of access and all the scaling, parallel computation, virtualization and whatever technology is used on the back end will be transparent to the client.

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: Five Characteristics**

1. On-demand self-service. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

2. Broad network access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

3. Resource pooling. Computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence - the client generally has no control or knowledge on the exact location of the resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Resources include storage, processing, memory, and network bandwidth.

4. Rapid elasticity. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the client, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

5. Measured service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

# **Three Service Models**

1.Software as a Service (SaaS) - Salesforce

• *SaaS* is the capability for the consumer to use a provider’s applications running on a cloud infrastructure.

• Applications are accessible from various client devices.

• 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.

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

• *PaaS* is capability for the consumer to deploy consumer created or consumer-acquired applications to the cloud infrastructure.

• These applications’ programming languages, libraries, services, and tools are supported by the cloud 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 configuration settings for the application-hosting environment.

3. Infrastructure as a Service (*IaaS*) MS Azure

• *IaaS* is the capability for the consumer to provision processing, storage, networks, and other fundamental computing resources which allows them 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, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

# **Four Deployment Models**

1. Private cloud. The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or a combination, and may exist on or off premises.

Example Bank of America runs an internal, restricted access cloud for development & application deployment within the banks intranet.

2. Community cloud. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). May be owned, managed, and operated by one or more organizations in the community, a third party, or a combination, and it may exist on or off premises.

3. Public cloud. The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

4. Hybrid cloud. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (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).

**Distributed Systems** A distributed system consists of a collection of autonomous computers, connected through a network and distribution middleware, which enables computers to coordinate their activities and to share the resources of the system, so that users perceive the system as a single, integrated computing facility.

**Cloud Mode** The ideal capacity-utilization curve is one where capacity acquired meets or just barely exceeds that which is required for an appropriate SLA. Amazon over built capacity because they had to… but created an all new business as a result.

