Introduction to Cloud Computing

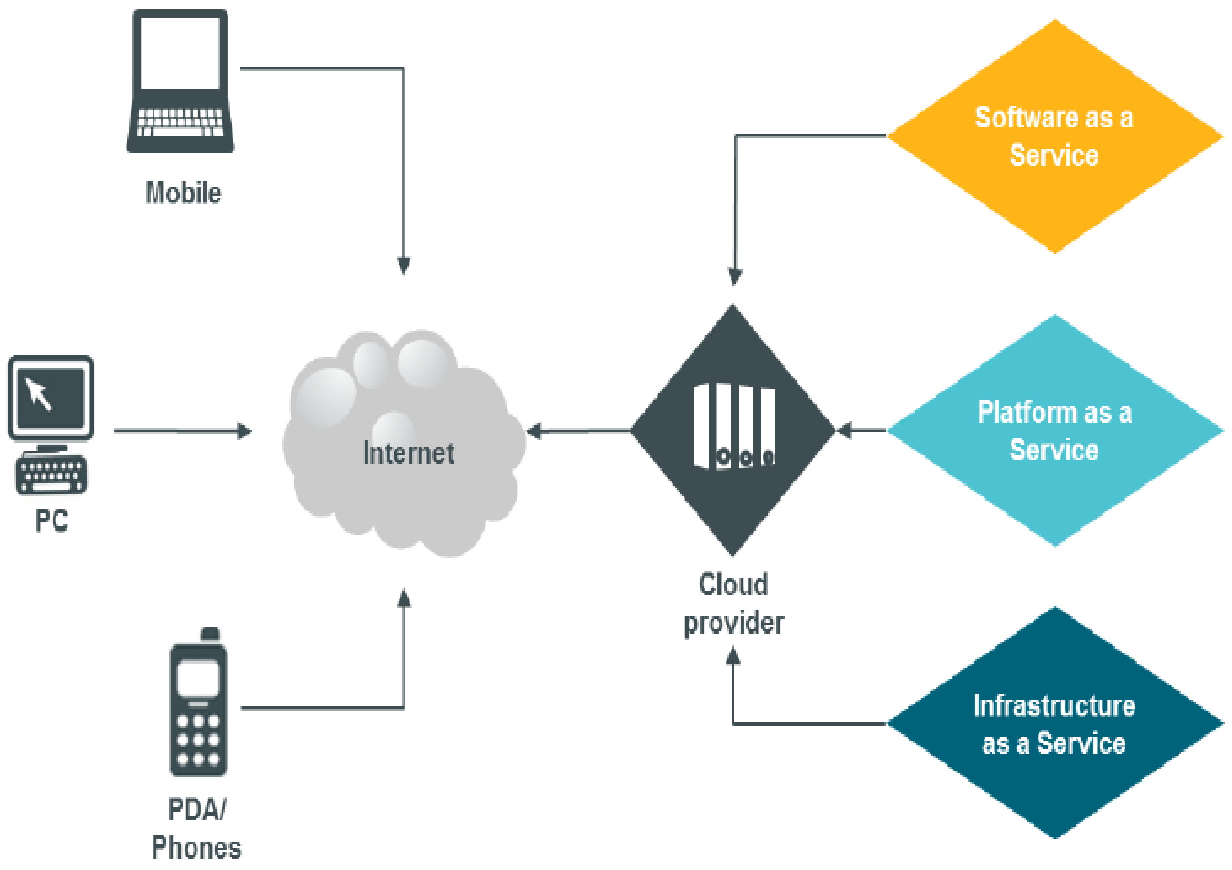
BY KHATIYA KATO PHILIP

Topic Outline

- What is Cloud Computing
- Cloud Computing Evolution
- Basic Concepts and Terminologies
- Features and characteristics
- Cloud Drivers
- Benefits of cloud computing
- Challenges of cloud computing

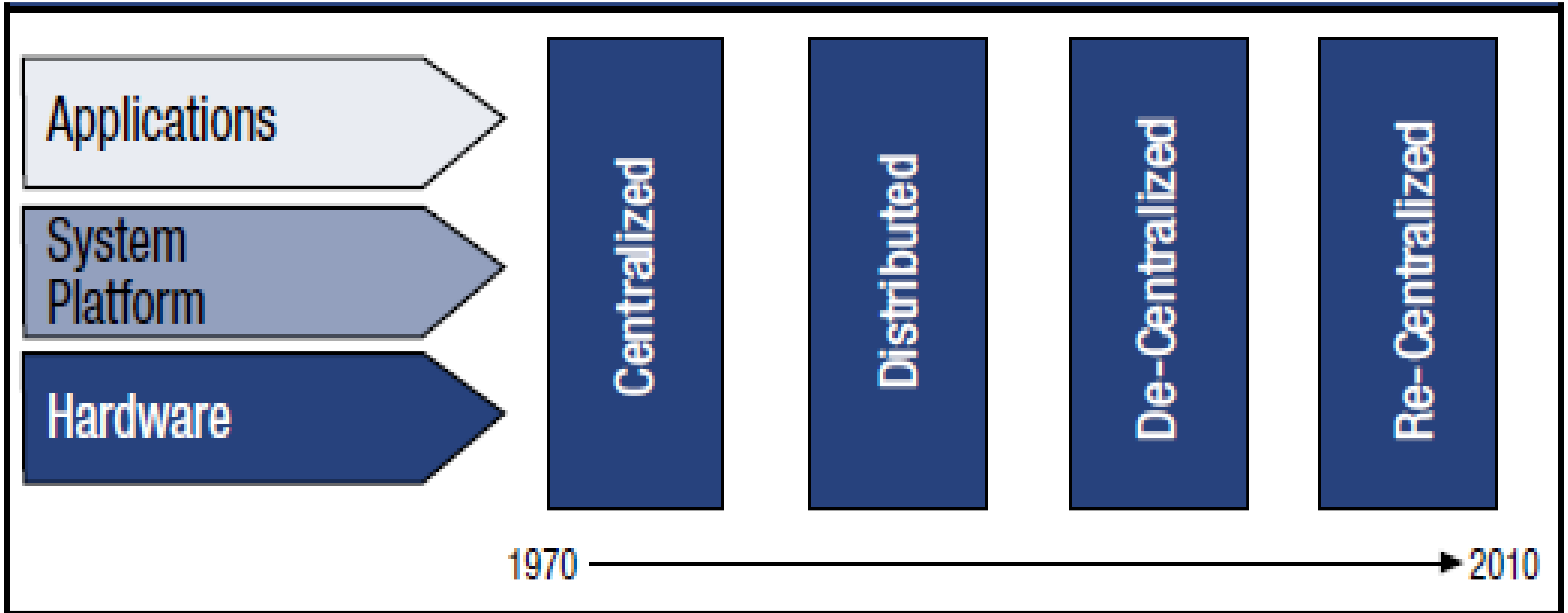
What is cloud computing?

'Cloud computing' is a term used to describe the delivery of computing services such as software, data access and data storage facilities over the internet rather than through a personal computer or location server. This enables users to access applications and data on-demand through a web browser regardless of their physical location or device (PC, notebook, iPad or mobile phone).



Source	Definition
Gartner	"a style of computing in which massively scalable IT-related capabilities are provided "as a service" using Internet technologies to multiple external customers" (Gartner 2008b)
IDC	"an emerging IT development, deployment and delivery model, enabling real-time delivery of products, services and solutions over the Internet (i.e., enabling cloud services)" (Gens 2008)
The 451 Group	"a service model that combines a general organizing principle for IT delivery, infrastructure components, an architectural approach and an economic model – basically, a confluence of grid computing, virtualization, utility computing, hosting and software as a service (SaaS)" (Fellows 2008)
Merrill Lynch	"the idea of delivering personal (e.g., email, word processing, presentations.) and business productivity applications (e.g., sales force automation, customer service, accounting) from centralized servers" (Merrill Lynch 2008)

Cloud Computing Evolution



Source: © Novell, Cloud Presentation 9.09. Used with permission.

The aggregation of technologies into today's cloud computing services was first successfully accomplished by several of today's largest CSPs for their own internal use. Enterprises such as Amazon and Google demonstrated internally the business benefits obtained by successfully implementing the cloud's "technical building blocks," described later in this chapter. These enterprises then leveraged their own in-house expertise in virtual computing and created the cloud computing service offerings that are now available to the public. Since then, cloud computing has evolved and is now commonly viewed as a major technology enhancement similar to the Internet.

However, cloud computing is not really new; it has been built on existing infrastructure and processes. Cloud computing has many similarities to the computer processing methods of the 1960s and 1970s. For example, 40 years ago, computing was centralized within enterprises, with large-scale operations using interfaces with mainframe computers.

User interfaces were limited primarily to dumb terminals or punch cards. The 1980s delivered midsized computers and minicomputers, which enabled computer processing to be distributed and accessed more readily throughout an enterprise. With the adoption of the Windows OS in the 1990s, computer processing was further distributed via client-server or simply client applications to nearly every office desktop, factory or warehouse station in an enterprise. In 2011, cloud computing is now returning users to centralized processing. Services are provided from hosts within the Internet. Through the World Wide Web, cloud computing is seen as the new mainframe. While many similarities exist, there are major differences between today's centralized cloud and the original mainframes. Among the notable differences are:

- Cloud processing power is much greater than that of the original mainframes.
- Storage capabilities have increased exponentially.
- The cloud allows a much larger number of user clients to connect.
- Connectivity is now over the World Wide Web; the transport protocols have changed.

Examples of Cloud Services

Company	Cloud computing platform	Year of launch	Key offerings
Amazon. com	AWS (Amazon Web Services)	2006	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Microsoft	Azure	2009	Application platform as a service (.Net, SQL data services)
Google	Google App. Engine	2008	Web Application Platform as a service (Python run time environment)
IBM	Blue Cloud	2008	Virtualized Blue cloud data center
Salesforce.com	Force.com	2008	Proprietary 4GL Web application framework as an on Demand platform

Basic Concepts and Terminologies

There are many components and terms used in cloud computing that are helpful in understanding the internal working of cloud technologies. Some of these terms include:

Hypervisor

A computer tool allowing various software applications running on different OSs to coexist on the same server at the same time. This means that Windows, Java, Linux, C++, Simple Object Access Protocol (SOAP) and Pearl-based applications can operate concurrently on the same machine. The hypervisor is the enabling technology for server virtualization.

Virtualization

The process of adding a “guest application” and data onto a “virtual server,” recognizing that the guest application will ultimately part company from this physical server

Dynamic partitioning

The variable allocation of CPU processing and memory to multiple applications and data on a server. Also known as logical partitioning (LPAR), dynamic partitioning provides variable CPU and server memory capacity to the various concurrently operating applications as needed. This is important because of the variable processing requirements experienced with batch jobs and real-time processing. Multiple concurrent applications may require near equal portions of CPU cycles and memory, but in some instances, one of the applications may need a much larger appropriation of processing power and memory space to avoid throughput delays. Dynamic partitioning reallocates the CPU and memory capacity as needed.

Web 2.0 and Mashup

Web 2.0 is a new concept that refers to the use of Web technology and Web design to enhance creativity, information sharing, and collaboration among users.

Cloud Computing Standards

Cloud computing standards have not been yet fully developed; however, a number of existing typically lightweight, open standards have facilitated the growth.

OS, application and data migration

The process of migrating data, the application and the underlying OS onto another server. Dynamic partitioning reallocates server processing and memory capacity as needed, automatically, on the fly. However, when the hypervisor senses that there is too much demand from the various applications for the host server's horsepower, tools exist to migrate data, the application and the underlying OS onto another server identified as available.

Cloud client usage measurement

The ability to measure usage of CPU processing, input/output and memory utilization per customer, per application. This measured services tool allows the CSPs that operate the servers for the cloud to charge clients usage fees based on the actual processing consumed.

Service Flow and Workflows

The concept of service flow and workflow refers to an integrated view of service-based activities provided in clouds. Workflows have become one of the important areas of research in the field of database and information systems.

Web Service and Service Oriented Architecture

Web Services and Service Oriented Architecture (SOA) are not new concepts; however, they represent the base technologies for cloud computing. Cloud services are typically designed as Web services, which follow industry standards including WSDL, SOAP, and UDDI. A Service Oriented Architecture organizes and manages Web services inside clouds. A SOA also includes a set of cloud services, which are available on various distributed platforms.

Features and characteristics

Characteristic	Description
On-demand self-service	The CSP can automatically provision computing capabilities such as server and network storage as needed, without requiring human interaction with each service's provider.
Broad network access	The cloud network should be accessible anywhere, by almost any device (e.g., smart phone, laptop, mobile device, PDA).
Resource pooling	The CSP's computing resources are pooled to serve multiple customers using a multitenant model, with different physical and virtual resources dynamically assigned and reassigned according to demand. There is a sense of location independence: The client generally has no control over or knowledge of the exact location of the provided resources, but may be able to specify location at a higher level of abstraction (e.g., country, region or data center). Examples of resources include storage, processing, memory, network bandwidth and virtual machines.
Rapid elasticity	Capabilities can be rapidly and elastically provisioned—in many cases, automatically—to accommodate customer needs. To the customer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
Measured service	Cloud computing systems automatically control and optimize resource usage by leveraging a metering capability (e.g., storage, processing, bandwidth and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for both the provider and customer of the utilized service.
Multitenancy of data	Multitenancy is the sharing of an application by multiple customers.

Cloud Drivers

Cloud computing is viewed as a significant change in the platform in which business services will be translated, used and managed. Many consider it to be as large a shift in IT as was the advent of the personal computer (PC) or of Internet access. However, a major difference between the cloud and those technologies is that the introductions of those earlier technologies encompassed a slower development phase. With the cloud, the required pieces for use have come together more rapidly for implementation. Some of the drivers bringing the cloud to the attention of enterprise decision makers are:

Optimized server utilization

Enterprises typically utilize just 15 to 20 percent of server computing resources.2 This means that they have five times the computing capacity than is typically used. By using many of the cloud-enabling tools described in this chapter, server utilization rates can increase four- to fivefold.

Cost savings

Increased server utilization plus the transition of computational capability from acquired and maintained computers to rented cloud services change the computing cost paradigm from a CAPEX to an OPEX, with potentially significant up-front and total cost savings.

Dynamic scalability

Many enterprises install five times their average computing requirements just to ensure that capacity exists to meet the large batch or peak demand. The cloud provides an extra processing buffer as needed, at low cost and without capital investment or a contingency fee to users.

Shortened development life cycle

Using cloud computing's SOA development approach, new business applications can be developed online, connecting proven functional application building blocks together. SOA-developed applications have measured completion times of one-fifth the time required for traditionally developed applications.

Reduced time for implementation

Cloud computing provides processing power and data storage as needed and at the capacity needed. This can be obtained in near-real time, not requiring the weeks or months (or CAPEX) that accrue when a new business initiative is brought online in a traditional IT enterprise.

Cloud Computing Benefits

Reduced Cost

There are a number of reasons to attribute Cloud technology with lower costs. The billing model is pay as per usage; the infrastructure is not purchased thus lowering maintenance. Initial expense and recurring expenses are much lower than traditional computing.

Increased Storage

With the massive Infrastructure that is offered by Cloud providers today, storage & maintenance of large volumes of data is a reality. Sudden workload spikes are also managed effectively & efficiently, since the cloud can scale dynamically.

Flexibility

This is an extremely important characteristic. With enterprises having to adapt, even more rapidly, to changing business conditions, speed to deliver is critical. Cloud computing stresses on getting applications to market very quickly, by using the most appropriate building blocks necessary for deployment.

Challenges of Cloud Computing

Data location

Regardless of the deployment model selected, customers may not know the physical location of the server used to store and process their data and applications. Cloud computing technology allows cloud servers to reside anywhere. From a technology standpoint, location becomes mostly irrelevant. However, for many compliance and data governance requirements, the physical location of the cloud computing server hosting user data is a critical issue.

Cloud security policy/procedure transparency

Some CSPs may have less transparency than others when it comes to their current information security policies. The rationalization for this is that the policies may be proprietary. This practice may cause conflict with clients’ information compliance requirements.

Cloud data ownership

Contract agreements may state that the CSP owns the data placed in the cloud computing environment that it maintains. The CSP may also require significant service fees for data to be returned to clients if and when a cloud computing services agreement terminates.

CSP business viability

As cloud computing continues to mature, there will be CSPs going out of business. Clients need to consider the risk and how data and applications can be easily transferred back to the traditional enterprise or to another CSP.

Record protection for forensic audits

Clients must also consider the availability of data and records if required for forensic audits. Since data may have been commingled and migrated among multiple servers located widely apart, it may be possible that the data for a specific point in time cannot be identified. Furthermore, local authorities may impound a cloud computing server to assess court-warranted data records of a suspect client taking with it the data of all the cloud computing clients sharing this impounded server.

Identity and access management (IAM)

Current CSPs may not develop and implement adequate user access privilege controls. With ever more sophisticated applications going online available for access by enterprise users, partners and clients highly granular, least privilege-based user access tools are required.

Penetration detection

Consideration should be given to whether the CSP has a penetration detection system in use. If such a system is in use, it is important to ensure that it has the required sophistication to monitor all cloud computing activities adequately. It is also important to consider whether a real-time digital dashboard is provided to user managers, along with audit logs and records of security incidents.

Screening of other cloud computing clients

By definition, CSPs leverage their cloud computing technology for many clients concurrently to maximize revenues. Clients should consider whether the other clients who share the same servers and, in the case of SaaS, the same application and data files are of the same repute as their own enterprise.

Compliance requirements

For the many compliance requirements including privacy and PII laws, Payment Card Industry (PCI) requirements, or various financial reporting laws today’s cloud computing services can challenge various compliance audit requirements currently in place. Data location, cloud computing security policy transparency and IAM are all challenging issues in compliance auditing efforts.

Public cloud server owners’ due diligence

Trust is a major component in the cloud computing business model. When contemplating transferring critical organizational data to the cloud computing platform, it is important to understand who and where all of the companies are that may touch the enterprise data. This includes not only the CSP, but all vendors that are in the critical path of the CSP.

Data erasure for current SaaS or PaaS applications

When an application and data are transferred from one server to another, as would be expected with dynamic scalability, the earlier application and data files may remain and may not be erased. Their space on the original hard drives is now available for overwrites. Customers need to ensure that this confidentiality is implemented by including language in the contract that provides for immediate data erasure upon contract termination.

Disaster recovery

Disaster recovery is a concern for potential cloud customers. In traditional hosting or colocation sites, customers know exactly where their data are in the event that they need to quickly retrieve them. The cloud model can change in the sense that public CSPs may outsource capabilities to third parties who may also outsource the original CSP may not be the CSP ultimately holding the data

Regulatory and Compliance Restrictions

In some of the some countries, Government regulations do not allow customer's personal information and other sensitive information to be physically located outside the state or country. In order to meet such requirements, cloud providers need to setup a data center or a storage site exclusively within the country to comply with regulations. Having such an infrastructure may not always be feasible and is a big challenge for cloud providers.

Cloud Models

BY KHATIYA KATO PHILIP

Content Outline

- Cloud Delivery Models
- Cloud Deployment Models

SERVICE MODELS

There are three models which differ in the capabilities that are offered to the consumer. It can be software, a platform, or infrastructure.

Software as a Service (SaaS)

In this service, the cloud service provider provides software and the cloud infrastructure to the clients so they can use this software on the cloud infrastructure for their applications. Since the clients can only run the software and use it, the client does not have control over the underlying infrastructure and physical setting of the cloud such as network, operating system, and storage. The cloud service provider is responsible and is the only one who is in charge of controlling underlying physical setting without client intervention.

The client can access this software as a thin client through a web browser. SaaS provides the most used cloud applications to nearly everyone online. Facebook, G-mail, LinkedIn, Yahoo user applications, Google Docs and Microsoft Online Services are all popular consumer-directed SaaS applications. SaaS allows customers 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.

Platform as a Service (PaaS)

This service is similar to SaaS in that the infrastructure is controlled by the cloud service provider but is different in that the users can deploy their software. In this model, the clients can install and deploy their customized applications by using the tool offered by the cloud service provider. Physical settings are controlled and restricted by the cloud service provider and application settings are given to each user to control them.

Infrastructure as a Service (IaaS)

In this service, computing resources such as processing, storage and networks can be provisioned. The client of IaaS can install and use any arbitrary operating system. Also, the clients can install and deploy their applications on this operating system. Cloud services such as Amazon EC2 are adopting this model and charging their clients according to the resources are being utilized.

SaaS	Highly scalable internet based applications are hosted on the cloud & offered as services to the end user.	Google Docs, acrobat.com, salesforce.com
PaaS	Here, the platforms used to design, develop, build & test applications are provided by the cloud infrastructure.	Azure Service Platform, force.com, Google App Engine.
IaaS	In this pay per use model, services like storage, database management & compute capabilities are offered on demand.	Amazon Web Services, GoGrid, 3 Tera

Service Model	Description	Considerations
IaaS	Capability to provision processing, storage, networks and other fundamental computing resources, offering the customer the ability to deploy and run arbitrary software, which can include OSs and applications. IaaS puts these IT operations into the hands of a third party.	IaaS can provide infrastructure services such as servers, disk space, network devices and memory. Example CSPs: • Amazon Web Services™ • Mosso from Rackspace®
PaaS	Capability to deploy onto the cloud infrastructure customer-created or customer-acquired applications developed using programming languages and tools supported by the provider	PaaS is designed for developers. Example vendors and services: • Microsoft's Azure™ Services Platform • Google's Google App Engine • Salesforce.com's Force.com®

Service Model	Description	Considerations
SaaS	Capability to use the provider's applications running on cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based e-mail).	Applications are complete and available on demand to the customer. Traditional licensing and asset management are changed. Example CSPs: • Microsoft Online Services • Salesforce customer relationship management (CRM) • LotusLive™ from IBM®
Source: Pijanowski, Keith; "Understanding Public Clouds: IaaS, PaaS and SaaS," Keith Pijanowski's Blog, 31 May 2009, *www.keithpij.com/Home/tabid/36/EntryID/27/Default.aspx*		

DEPLOYMENT MODELS:

There are four deployment models as following:

Private cloud: In this model, the cloud provider provides cloud infrastructure to a single organization that has many consumers. This infrastructure is to be used exclusively for their use and need. The owner, manager, and operator of this cloud could be the organization itself, a third party, or the organization and third party together. This private cloud could be on premises or off premises.

Community Cloud: In this model, the cloud provider provides cloud infrastructure to many organizations that forms community that shares mission, security requirements, compliance consideration, or policy.

This infrastructure is to be used exclusively for their uses and needs. The owner, manager, and operator of this cloud could be one of organizations, a third party, or the organization and third party together. This Community cloud could be on premises or off premises.

Public Cloud: This model differs from the previous model in that it is open for the public; it is not private and not exclusively for community. In this model, a public cloud can be provisioned for public to use it to satisfy their needs. The owner, manager, and operator of this cloud could be a government, private organization, a business or academic organization, and sometimes many of them can be in one cloud and get the service from the same provider.

Hybrid Cloud: This model comprises two or more deployment models (private, community, or public). The cloud infrastructure can be combination of those models. Data center within an organization, private cloud, and public cloud can be combined in order to get services and data from both in order to create a well-managed and unified computing environment. A cloud can be considered hybrid if the data moves from a data center to a private cloud or public cloud or vice versa.

Deployment Model	Description
Private cloud	• Operated solely for an enterprise • May be managed by the enterprise or a third party • May exist on- or off-premise
Public cloud	• Made available to the general public or a large industry group • Owned by an organization selling cloud services
Community cloud	• Shared by several enterprises • Supports a specific community that has a shared mission or interest • May be managed by the enterprises or a third party • May reside on- or off-premise
Hybrid cloud	• 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)