Chapter One

Introduction to Cloud Computing

- College of Science
- Department of Computer Science
- 4th Class/ CS

Dr. Ethar Abdul Wahhab

Introduction to Cloud Computing

1.1 Introduction

Today, cloud computing considered as a modern generation in the information technology revolution. It provides a wide range for sharing services, resources, and storing data. In a cloud environment, clients can access the cloud and use available services and their data from anywhere without knowing where it is hosted.

There is an increasingly trend in many organizations to move a large portion or even all operations to an Internet-connected infrastructure known as enterprise cloud computing. Cloud computing has been coined as an umbrella term to describe a category of on-demand computing services initially offered by commercial providers, such as Amazon, Google, and Microsoft. It denotes a model where 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". 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 collecting resources and offering a single system view. Cloud computing can be thought of as a new approach to services-based utility computing in which a company pays only for the services it requires, rather than putting effort into the technical details of setting up an

entire infrastructure or requiring licensing for all of the company's software.. Cloud computing allows for the real-time allocation of computer power, storage resources and network resources to various clients. Clients of cloud services are always concerned about who has access to their data stored on cloud servers. For cloud computing to be widely accepted, trustworthiness between Cloud Service Providers (CSPs) and cloud clients is critical.

1.2 Cloud Computing Definition

Many specialists in the commercial and academic scopes have attempted to define exactly what "cloud computing" is and what unique characteristics it presents. Cloud computing, can be defined as "a type of parallel and distributed system made up of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resource(s) based on Service Level Agreements (SLA) established through negotiation between the service provider and clients".

Cloud computing is also defined as "both the applications offered as services via the internet and the hardware and systems software in the data centers that support those services". It can be described as "the new computing model that transforms the process of resources maintaining, purchasing, and disposal it through enabling the demand mechanism of the dynamic basket of resources, which are existing in special data centers through the Internet".

The National Institute of Standards and Technology's (NIST) definition of cloud computing is the most widely accepted in the literature. NIST defines cloud computing as "**a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources such as networks,**

servers, storage, applications, and services that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Cloud computing has five fundamental qualities such as broad access to the network, measurable service, fast elasticity, resource pooling and on demand self service. It has three service delivery methods such as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS)and four deployment strategies such as private cloud, public cloud, hybrid cloud, and community cloud are the service models. Figure (1.1) depicts the five ***essential characteristics***, three ***cloud service*** models, and four ***cloud deployment*** models.

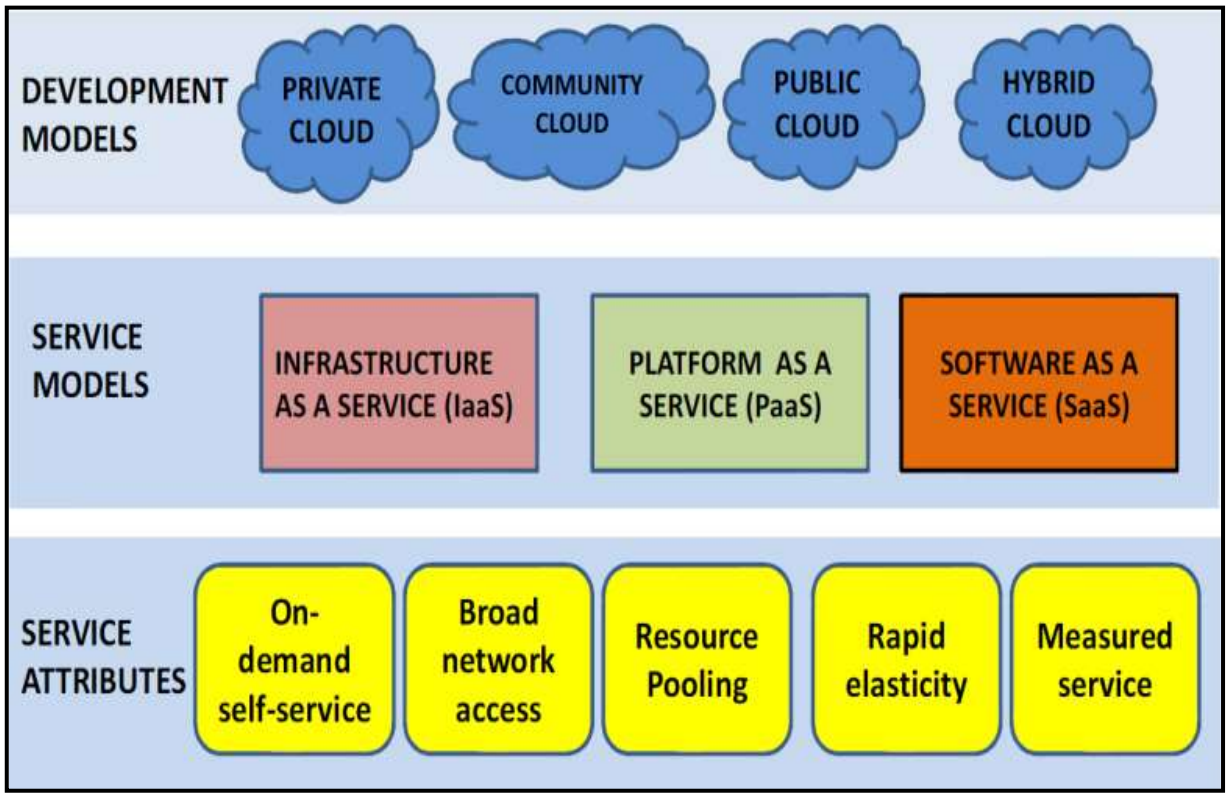

Figure (1.1). NIST visual model of cloud computing definition

1.3 Cloud Computing Components

Cloud computing consist of three components: Clients, Distributed servers and Data centers as show in figure (1.2). Each of these components has an essential role in providing cloud services.

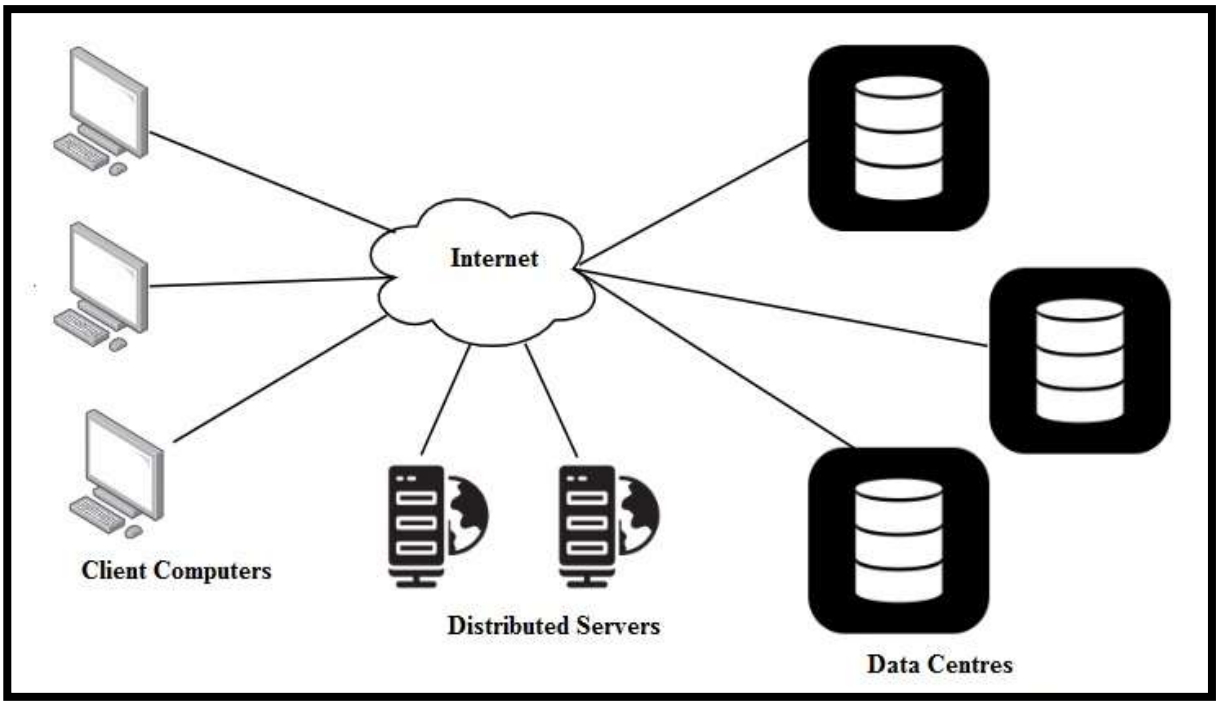

Figure (1.2). Cloud computing component

Clients: Clients in cloud computing a person or organization that requests access to a service provided by a cloud. These might be in the form of laptops, tablets, home automation devices, mobile phones or web browser. Clients hold the responsibility of interaction which pushes for the management of data on cloud servers.

Distributed Servers: In order to provide the availability, the cloud has distributed the servers in the different geographic area. It includes multiple

servers working together to provide a service or application. In the case of failure in the specific server, then the other server will take the action. Cloud servers are virtual servers running in a cloud computing environment that can be accessed on demand by unlimited users.

Data center: A data center is a physical location that stores computing machines and their related hardware equipment. It contains the computing infrastructure that IT systems require, such as servers, data storage drives, and network equipment. It is the physical capability that stores any company's digital data. Cloud providers maintain large data centers with full security and compliance.

1.4 Cloud Computing Architecture:

The cloud computing architecture is generally classified into three categories: a front-end platform, a backend platform, and cloud-based delivery. The architecture of the system needs the Internet for communication between the front end and the back end. The delivery system, as the name suggests, is what allows information to be delivered between the front end and the backend.

Frontend is a term used to denote any user-facing part of cloud computing architecture. This is the part that the end user interacts with generally takes the form of a user interface and is an integral part of how the user interacts with cloud computing software. Frontend is made up of three parts:

a) *Software* that allows cloud computing to be run from the user's side such as web browser Google Chrome, Firefox, Internet Explorer and others.

b) *User interface* is what the end user directly interacts with in order to perform tasks on the cloud. This includes the text editor in services such as Google Docs, or the interface to send and receive emails on Gmail.

c) *Client device* includes client-side hardware such as the user's PC and input devices. Generally, in cloud computing, the client device does not require a lot of computing power. This is because most of the "heavy" tasks are processed in the cloud.

Backend is the part of the cloud computing architecture that runs the front-end. This includes the core components of the system such as hardware and storage, and is generally located in a servers farm in a geographically distant location. Backend architecture is made up of many different parts such as:

a) *Application* can be any software or platform which a client wants to access, this layer coordinates the various needs of the client with resources in the backend.

b) *Service* refers to the major three types of cloud based services like SaaS, PaaS and IaaS. It manages which type of service the client accesses according to its requests.

c) *Storage* Cloud storage is where the data resides within a cloud application. Storage varies with cloud service providers today, with all of them having a product dedicated solely to cloud storage.

d) *Management* software is responsible for allocating specific resources for certain tasks. Management usually takes the form of what is known as 'middleware', as it boundaries between the backend and frontend.

Middleware is used to divide system resources and infrastructure in effective and dynamic manner.

e) *Security* refers to implementing different security mechanisms for secure cloud environment, resources, files, and infrastructure to prevent data loss and keep it safe from attackers.

Figure (1.3) explain the architecture of cloud computing

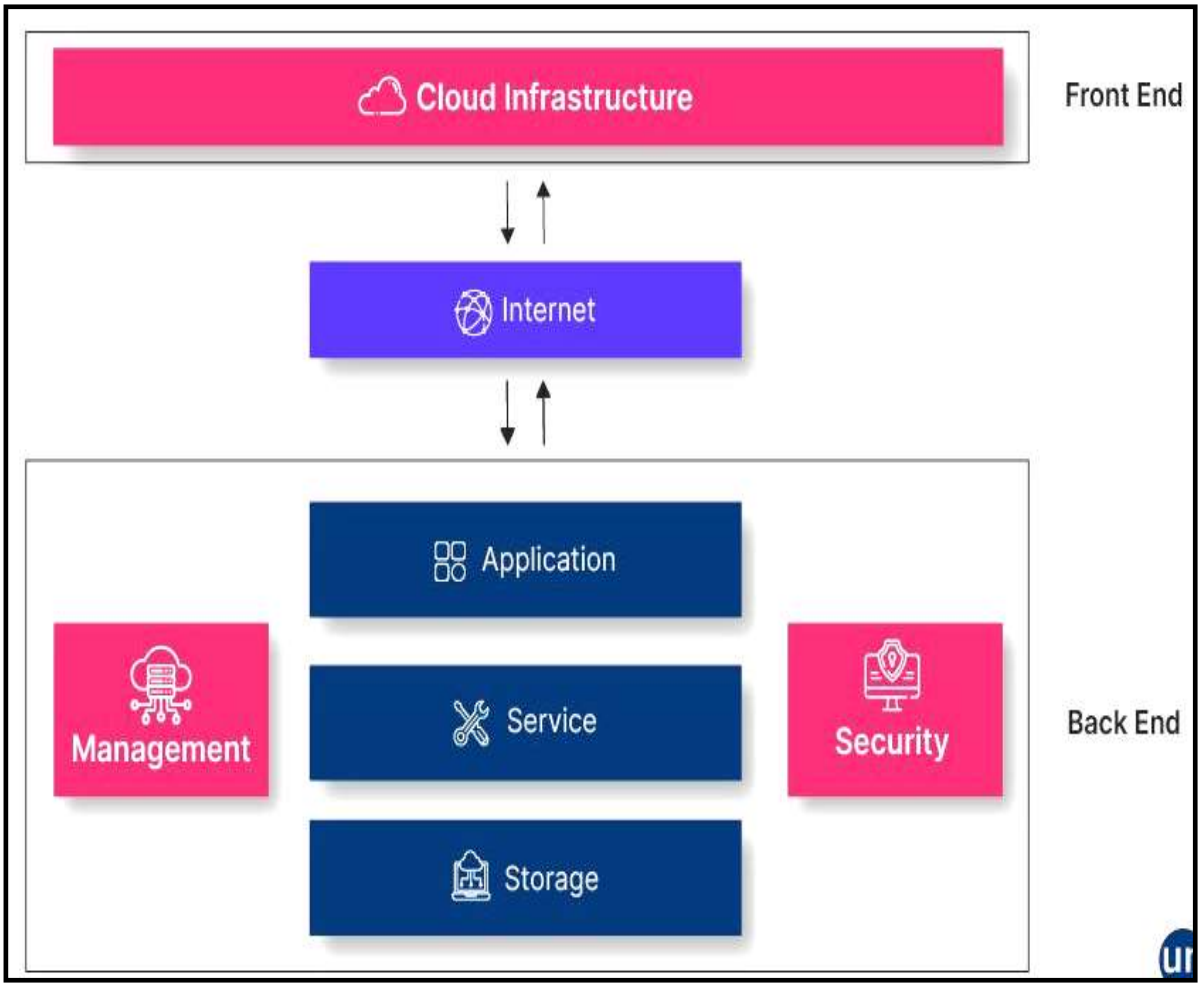

Figure (1.3). Cloud computing architecture

1.5 Evolution of Cloud Computing

Cloud Computing has emerged from the 1950s to the current year and companies are relying completely on cloud computing based on their specific needs. The cloud symbol was used to represent the Internet as early as 1994, in which servers were then shown connected to, but external to, the cloud. In making cloud computing what it is today, five technologies played a vital role.

1) Mainframe Computing:

In the past, companies powered their information infrastructure from a mainframe. In one physical location (i.e. a building or an office), this large, powerful computer stored data and ran all the software applications. While it was easy to support multiple applications through one mainframe, maintaining such a large piece of hardware was expensive and inefficient. Mainframes can be defined as high-performance computers with large amounts of memory and data processors that process billions of simple calculations and transactions in real time. The presence of a mainframe often involves a centralized form of computing, as opposed to a distributed form of computing.

2) Distributed computing:

As lower cost computing became more available (enter the IBM PC in 1981) and as more people wanted access to more powerful applications, mainframe computing became less effective. The next solution for businesses was to replace the mainframe with multiple cheaper computers, each with enough computing power to store data and run

applications. Distributed computing is a composition of multiple independent computer but all of them are depicted as a single entity to the users to utilize computational resources across multiple, separate computation nodes to achieve a common, shared goal. The purpose of distributed computing is to share resources and also use them effectively and efficiently.

Distributed computing systems have the following characteristics:

- Resource sharing : A distributed system can share hardware, software, or data and every node can access and communicate with other nodes in the system.
- Simultaneous processing: Multiple machines can process the same function simultaneously and a complex requests can be broken down into smaller pieces and distributed among different systems.
- Scalability: The computing and processing capacity can scale up as needed when extended to additional machines.
- Error detection: Failures can be more easily detected and if one machine isn't available, others can fill in for the service.
- Cost-effectiveness:. Distributed computing can use low-cost, off-the-shelf hardware reducing the time needed to compute requests.

Despite these advantages of the distributed computing but the main problem with this system was that all the systems were required to be present at the same geographical location. Thus to solve this problem,

distributed computing led to other types of computing and they are cluster computing, and grid computing.

3) Parallel computing:

Parallel computing refers to the process of breaking down larger problems into smaller, independent that can be executed simultaneously by multiple processors communicating via shared memory, the results of which are combined upon completion as part of an overall algorithm. The primary goal of parallel computing is to increase available computation power for faster application processing and problem solving. Parallel computing infrastructure is typically housed within a single datacenter and computation requests are distributed in small parts and each part is further broken down to a series of instructions. Instructions from each part execute simultaneously on different processors.

4) Cluster Computing:

Cluster computing is a type of parallel or distributed computing represent a system that consists of two or more homogeneous computers often known as nodes. These nodes work together for executing applications and performing other tasks in a minimum amount of time. A computer node can be a single or multiprocessor system (PC_s or workstations) with memory, I/O devices and an operating system. These nodes connected via Local Area Network (LAN) such as Ethernet, and can appear as a single system. Cluster computing can provides the following benefit:

- Increased resource availability: If one server in a cluster fails, the other servers in the cluster can pick up the workload. This prevents the loss of valuable time and information if a server fails.
- Strategic resource usage: the user can distribute projects across nodes in whatever configuration prefer. This reduces overhead because not all machines need to be running all projects, and allows user to use the resources flexibly.
- Increased performance: Multiple machines provide greater processing power.
- Greater scalability: As the user base grows and report complexity increases, the resources can grow.

However, the cluster computing suffers from some disadvantages such as the high cost since the cluster needs good hardware and a design, it will be costly comparing to a non-clustered server management design. Moreover, cluster computing needs more servers and hardware to establish one, monitoring and maintenance is hard.

5) Grid Computing:

Grid computing is a form of distributed computing can be defined as a network of homogeneous or heterogeneous computers working together over a long distance (Wide Area Network (WAN) or Internet) to perform a task that would rather be difficult for a single machine. Each computer in the grid contributes its processing power to the overall effort.

There are many benefits to using grid computing, including increased processing power, improved resource utilization, and reduced costs. It also allows organizations to share computing resources, reducing the need for each organization to spend in expensive hardware and software.

6) Cloud Computing:

Cloud computing is more flexible than grid computing and follows client-server computing architecture. Nowadays the companies are in need of some virtual space to store the data which generates. So cloud computing provides this facility which is flexible and compatibility issues. It helps to run the infrastructure in an efficient manner and integrate the business very quickly. The difference between the grid computing and cloud computing can be summarized in the following table:

Table (1.1). Deference between grid computing & cloud computing

	Grid Computing	Cloud Computing
1	Grid computing follows a distributed computing architecture.	Cloud Computing follows client-server computing architecture.
2	Grid Computing is less flexible than cloud computing.	Cloud Computing is more flexible than grid computing.
3	Grid operates as a decentralized management system.	Cloud operates as a centralized management system.
4	Grid computing is Application-oriented.	Cloud Computing is Service-oriented.

5	Low level of abstraction (more details)	High level of abstraction (eliminate details)
6	Normal scalability	High scalability
7	The users have full control and management over the grid system	The users have less control over the system in the cloud
8	Resources are managed on a collaboration pattern.	Resources are centrally managed.
9	Allocation of multiple servers onto single task or job	Virtualization of servers (one server to compute several tasks concurrently)

1.6 History of Cloud Service Providers

Cloud Service Provider (CSP) is IT company that provides scalable computing resources that businesses can access on demand over a network, including cloud-based compute, storage, platform, and application services. Big organization providing the service like Amazon AWS, Microsoft Azure, Google Cloud, Alibaba cloud etc. is some Cloud Computing service Provider.

Cloud service providers have a rich history that can be summarized as follows:

- In 1960s, the first general-purpose computer network that connects different types of computers together on a large-scale has been created.

This is important because it paved the way for the modern Internet, the platform on which cloud computing runs.

- In 1974, IBM advertises its Systems Network Architecture (SNA). It is a set of protocols they designed to support less centralized networks. The personal computers are realized at the end of 1970s .
- By 1985, data storage tapes capable of holding 200 megabytes are available. There are approximately 100,000 Internet-connected computers by this time.
- The use of the "cloud" metaphor to denote virtualized services traces back to 1994.
- In 1999, Salesforce.com became the first company started delivering an enterprise application over the internet and this way the boom of Cloud Computing was started.
- In 2002, Amazon started Amazon Web Services (AWS), Amazon will provide storage, computation over the internet.
- In 2006 Amazon will launch Elastic Computer Cloud (EC2) Commercial Service which is open for everybody to use.
- After that in 2009, Google Play also started providing Cloud Computing Enterprise Application and other applications such as Gmail, Google Calendar, Google Docs, Google Search engine and YouTube .
- Thus, in 2010, Microsoft launches Microsoft Azure which supports streamlined development of web and mobile apps, to support a rapidly growing mobile industry.

- After that other companies like Alibaba, IBM, Oracle, HP also introduces their Cloud Services.
- In today the Cloud Computing become very popular and important skill.

1.7 Cloud Computing Characteristics

There are five essential characteristics of cloud computing as shown in figure (1.4):

1. **On-demand self-service:** Clients can get cloud computing services anytime they need them.
2. **Resource pooling:** Various resources are pooled at the same time and dynamically distributed to clients.
3. **Rapid elasticity:** Rapid elasticity provides scalable services. This ability can be used by users who require additional cloud storage.
4. **Broad network access:** Clients can use a variety of methods to access all of the cloud services that are offered.
5. **Measured services:** Resource use is closely monitored and managed. As a result, both the client and the provider receive transparent service.

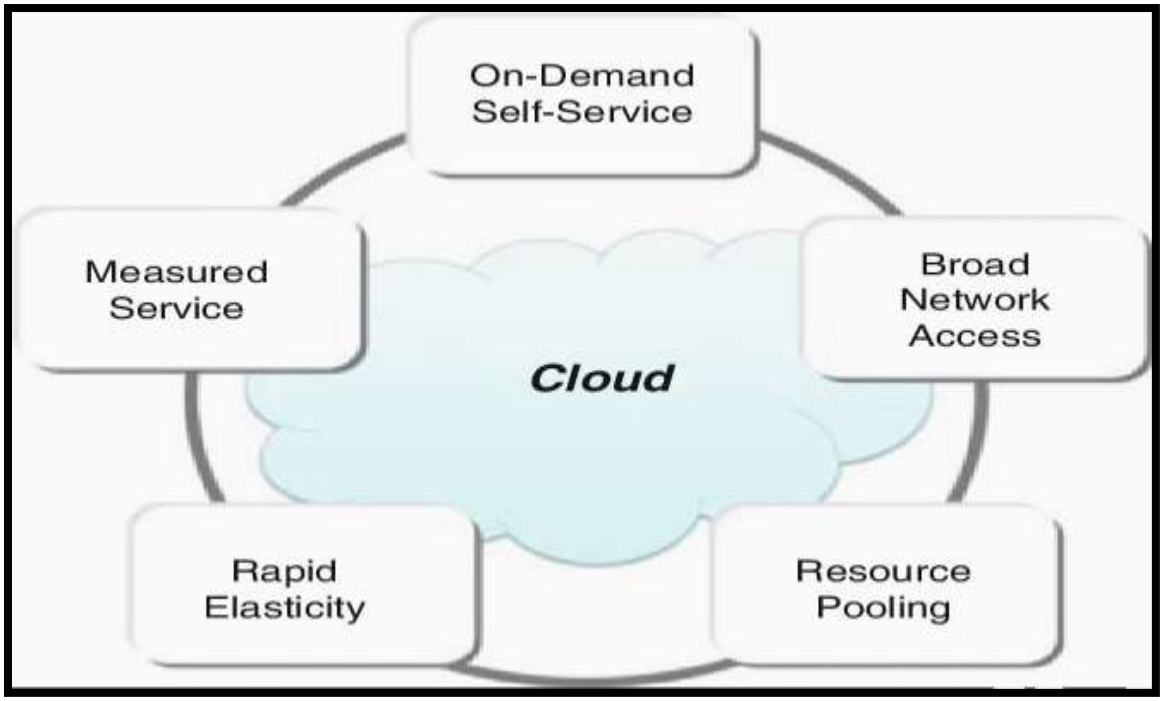

Figure (1.4). Characteristics of cloud computing

1.8 Benefits of Cloud Computing

Cloud computing benefits can be categorized into:

1. **Reduced IT costs:** Moving to cloud computing may reduce the cost of managing and maintaining IT systems. Rather than purchasing expensive systems and equipment for business, the costs can reduced by using the resources of the cloud computing service provider.

2. **Mobility and Flexibility:** Cloud computing allows employees to be more flexible in their work practices. For example, customer has the ability to access data from home anytime.

3. **Backup and Recovery:** Since all the data is stored in the cloud, backing it up and restoring the same is relatively much easier than storing the same on a physical device. Also it has many techniques to recover it from any type of disaster.

4. **Collaboration:** Cloud computing provide a convenient way to work group of people together on a common project or applications in an effective manner.

5. **Automatic Software Updates:** Cloud-based applications automatically refresh and update themselves, instead of forcing an IT department to perform a manual organization-wide update.

1.9 Challenges of Adapting Cloud Computing

1. **Security:** Cloud computing providers must implement robust security measures to safeguard data from cyber threats such as hacking, malware, and data breaches and ensure that data is secure and protected from unauthorized access.

2. **Data privacy:** Data privacy is another significant challenge in cloud computing. Cloud computing providers must implement measures to ensure data privacy and confidentiality while allowing clients to access and collaborate on data.

3. **Availability:** Clients need access to cloud computing services around the clock to ensure that they can work on their projects at any time and ensure that their services are always available and that there are no downtimes.

4. **Data integration:** Cloud computing providers must offer tools and services that make it easy for clients to integrate data from various sources, ensuring that the data is accurate, reliable, and up-to-date.

5. **Network Latency Issues:** another limitation of cloud computing for businesses is network latency issues, which occur when there is a delay in transferring data between two points due to network congestion or other factors such as distance between servers or slow internet speeds at either end point.