

Cloud Computing

Introduction to Cloud Computing: Computing Paradigms, Cloud Computing Fundamentals, Motivation for Cloud Computing, Defining Cloud Computing, Principles of Cloud Computing,

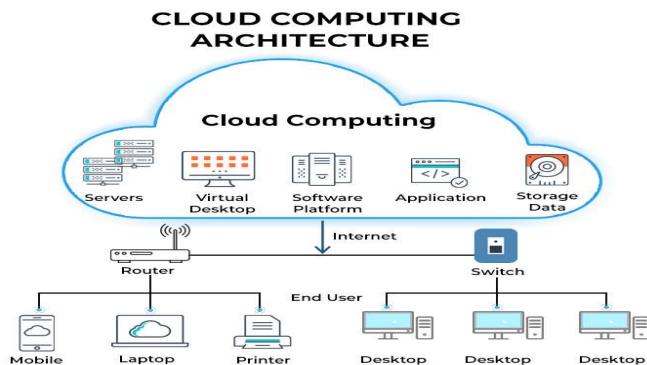
Cloud Computing Architecture and Management, Cloud Architecture, Network Connectivity in Cloud Computing, Applications on Cloud, Managing the Cloud, Migrating Application to Cloud.

Introduction to Cloud Computing:

Cloud Computing provides us means by which we can access the applications as utilities over the internet. It allows us to create, configure, and customize the business applications online.

What is Cloud Computing?

Cloud computing is a way for providing computing resources over internet. The computing resources include physical or virtual servers, data storage, networking capabilities, application development tools, software, analytic platforms and more. Cloud can provide services over network, that is, on public networks or on private networks, that is, Wide Area Networks (WANs), Local Area Networks (LANs), or Virtual Private Networks (VPNs). Applications such as e-mail, web conferencing, customer relationship management (CRM), all run in cloud.



Instead of storing files on a storage device or hard drive, a user can save them on cloud, making it possible to access the files from anywhere, as long as they have access to the web. The services hosted on cloud can be broadly divided into infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). Based on the deployment model, cloud can also be classified as public, private, and hybrid cloud.

Further, cloud can be divided into two different layers, namely, front-end and back-end. The layer with which users interact is called the front-end layer. This layer enables a user to access the data that has been stored in cloud through cloud computing software.

The layer made up of software and hardware, i.e., the computers, servers, central servers, and databases, is the back-end layer. This layer is the primary component of cloud and is entirely responsible for storing information securely. To ensure seamless connectivity between devices linked via cloud computing, the central servers use a software called middleware that acts as a bridge between the database and applications.

1. Computing Paradigms

A computing paradigm refers to a fundamental approach or model for performing computation, organizing data, designing systems of higher levels, and solving problems using computers.

Characteristics of Computing Paradigms

- It encompasses the principles, techniques, methodologies, and architectures that guide the design, development, and deployment of computational systems.
- Computing paradigms can vary widely based on factors such as the underlying hardware, programming models, and problem-solving strategies.
- The choice of paradigm depends on factors such as the nature of the problem, performance requirements, scalability, and ease of development.
- Many modern computing systems and applications combine multiple paradigms to use their respective strengths and address complex challenges.

Types of High-Performance Computing Paradigm

- High-performance computing (HPC) refers to the use of advanced computing techniques and technologies to solve complex problems and perform data-intensive tasks at speeds beyond what a conventional computer could achieve.
- Different high-computing performance paradigms arise from various methodologies, principles, and technologies used to solve complex computational problems.
- Each computing paradigm has its strengths, weaknesses, and specific use cases. The choice of use of a paradigm depends on the nature of the problem to be solved, performance requirements, and the available technology.
- There are some key computing paradigms are as follows:-

1. Sequential Computing

- This is a traditional paradigm that involves the execution of instructions in a linear sequence.
- It is the foundation of classical computing architectures, where a processor executes one instruction at a time.

2. Parallel Computing

- Parallel computing is a computing paradigm where multiple computations or processes are executed simultaneously to solve a single problem, typically to improve performance, efficiency, and scalability.
- In parallel computing, multiple processors or multi-cores work simultaneously on different parts of a problem.
- In parallel computing, tasks are divided into smaller subtasks that can be executed concurrently on multiple processing units or cores, allowing for faster execution and higher throughput compared to sequential processing.
- Parallel computing is used in various domains, including scientific simulations, data analytics, image and signal processing, artificial intelligence, and computer graphics.

- Examples of parallel computing applications include weather forecasting, molecular dynamics simulations, genome sequencing, deep learning training, and rendering complex 3D graphics.

3. Distributed/Network Computing

- Distributed computing involves the use of multiple computers connected to a network.
- Tasks are distributed across these distributed computers, and they work collaboratively to achieve a common goal.
- Network computing, also known as distributed computing, refers to the use of interconnected computers and resources to perform tasks collaboratively over a network.
- This infrastructure can include local area networks (LANs), wide area networks (WANs), and the Internet.
- The client-server model is a common architecture in network computing.
- Network computing allows users to access resources and applications remotely.
- Resources such as processing power, storage, and applications are distributed across multiple computers within the network. This enables users to access and utilize resources located on different machines.
- Network computing facilitates collaboration among users by enabling them to share files, work on documents simultaneously, and communicate in real time. Collaboration tools, such as email, video conferencing, and collaborative document editing, are common in networked environments.
- Network computing systems can be easily scaled by adding more computers or resources to the network. This scalability allows organizations to adapt to changing demands and accommodate growing workloads.
- In network computing, computers are connected to share resources, exchange information, and work together to achieve common goals.
- This paradigm allows for the efficient use of resources, improved collaboration, and the distribution of computational tasks across multiple devices.
- The evolution of network computing has contributed to the development of various technologies, including cloud computing, edge computing, and distributed computing systems.
- Cloud computing is an example of distributed computing.

4. Grid Computing

- Grid computing is a distributed computing paradigm that manages the computational resources of multiple networked computers or clusters to solve large-scale computational problems.
- Grid computing involves the coordination of geographically dispersed resources to work on a common task.
- Grid computing systems consist of multiple nodes or sites interconnected by high-speed networks, such as the Internet or dedicated communication links. Each node in the grid can contribute its computational power and resources to the collective pool, creating a distributed computing infrastructure.

- Grid computing relies on middleware software to manage resource discovery, allocation, scheduling, security, and communication within the grid. Grid middleware provides a set of services and APIs (Application Programming Interfaces) that abstract the underlying infrastructure and facilitate the development and execution of grid applications.
- It typically involves pooling together computing resources from multiple locations to solve large-scale problems.
- In grid computing, resources such as processing power, storage, and software applications are shared across geographically distributed sites, allowing organizations to leverage idle resources and collaborate on complex tasks.
- Grid computing facilitates collaborative research and scientific discovery by enabling researchers and organizations to share data, computational resources, and expertise across institutional boundaries.
- Grid computing offers several benefits, including increased computational power, scalability, fault tolerance, and cost-effectiveness.

5. Cluster Computing

- Cluster computing is a type of parallel computing in which multiple interconnected computers or nodes work together as a single integrated system to solve large-scale computational problems or perform complex tasks.
- Cluster computing uses the collective computational power and resources of the individual nodes to achieve higher performance, scalability, and reliability compared to single-node systems.
- In a cluster computing environment, multiple independent computers, servers, or nodes are interconnected through a high-speed network, such as Ethernet or InfiniBand. Each node typically has its own CPU (Central Processing Unit), memory, storage, and operating system and each node in the cluster performs a specific function, contributing to the overall computing power of the cluster.
- Cluster computing uses parallel processing techniques to divide computational tasks into smaller sub-tasks that can be executed concurrently across multiple nodes. This allows for faster execution of tasks and improved overall performance, as the workload is distributed among the nodes.
- Cluster computing systems often employ load-balancing techniques to evenly distribute computational tasks among the nodes and maximize resource utilization. Load-balancing algorithms dynamically allocate tasks based on factors such as node performance, workload, and availability.
- Cluster computing is widely used in scientific research, engineering simulations, data analysis, machine learning, financial modeling, and other computationally intensive applications. Examples include weather forecasting, drug discovery, genomic sequencing, financial risk analysis, and large-scale data processing.

- **Types of Clusters:** There are different types of clusters, including:-
 - **Homogeneous Clusters:** All nodes in the cluster have similar hardware configurations and run the same operating system or software applications.
 - **Heterogeneous Clusters:** Nodes in the cluster have varying hardware configurations and may run different operating systems or software environments.
 - **High-Performance Computing (HPC) Clusters:** Designed for scientific and engineering simulations, these clusters use specialized hardware, such as GPUs (Graphics Processing Units), for accelerated computation.
 - **Cloud-based Clusters:** Clusters are hosted in cloud computing environments, where resources are provisioned and managed dynamically using cloud services.

6. Cloud Computing

- **Cloud computing** refers to the delivery of computing services, including servers, storage, databases, networking, processing power, software applications, and analytics, over the internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale.
- Cloud computing is a modern form of network computing where computing resources, applications, and services are delivered over the Internet.
- Cloud computing allows users to access and manage data and applications without having to buy or maintain physical hardware and software.
- Cloud services have a global presence, allowing organizations to expand their reach and serve customers in different geographic locations without the need for establishing physical infrastructure in each region.
- Cloud computing eliminates the need for organizations to own and maintain physical hardware and infrastructure, enabling them to access computing resources on-demand and pay only for what they use.
- Cloud computing has many benefits, such as cost reduction, scalability, performance, reliability, and security.
- Security in cloud computing is a critical aspect of cloud structure that involves protecting data, applications, and infrastructure in cloud environments.
- Cloud security is a shared responsibility between cloud service providers and their customers.

7. Quantum Computing

- Quantum computing is an emerging field of computing that manages the principles of quantum mechanics to perform computations using quantum bits or qubits.
- Quantum computing has the potential to revolutionize fields such as cryptography, optimization, machine learning, material science, and drug discovery by solving complex problems that are intractable for classical computers.
- Quantum computers have the potential to solve certain problems exponentially faster than classical computers.

- Quantum computing is a subset of nano computing that uses the principles of quantum mechanics to perform computation.
- Unlike classical computing, which uses bits that can be in a state of either 0 or 1, quantum computing utilizes quantum bits, which can exist in multiple states simultaneously due to superposition and entanglement, enabling quantum computers to solve certain problems exponentially faster than classical computers.
- Qubits are the basic units of quantum information in quantum computing. Unlike classical bits, which can be in either a 0 or 1 state, qubits can exist in a superposition of both states simultaneously. This enables quantum computers to perform parallel computations and process vast amounts of information more efficiently than classical computers.
- Superposition is a fundamental principle of quantum mechanics that allows qubits to exist in multiple states at the same time. This property enables quantum computers to consider and process many possible solutions simultaneously, exponentially increasing their computational power for certain types of problems.
- Quantum computers require specialized hardware to create and manipulate qubits. Various physical systems are being explored for implementing qubits, including superconducting circuits, trapped ions, quantum dots, and topological qubits. Building scalable and error-corrected quantum hardware is a significant challenge in the development of practical quantum computers.

8. Biological/Bio/DNA Computing

- Biological computing, also known as **biocomputing or bioinformatics**, refers to the use of biological systems, molecules, and processes to perform computation and solve computational problems.
- Biological computing draws inspiration from the complex and highly parallel nature of biological systems, such as DNA, RNA, proteins, cells, and organisms, to develop new computational paradigms, algorithms, and technologies.
- Biological computing explores the use of biological processes or materials (such as DNA computing) to perform computational tasks.
- **Bioinformatics** is a multidisciplinary field that combines biology, computer science, mathematics, and statistics to analyze and interpret biological data, such as DNA sequences, protein structures, and gene expression profiles. Bioinformatics techniques and tools enable researchers to discover new insights into the structure, function, and evolution of biological systems.
- Biological computing offers exciting opportunities for innovation and discovery by integrating principles from biology and computer science to develop new computational technologies and approaches.
- Biological computing faces several challenges, including scalability, reliability, and the need for new computational models and algorithms that can effectively harness the complexity and diversity of biological systems.

- Biological computing has applications in various domains, including healthcare, biotechnology, agriculture, environmental monitoring, and bioenergy.
- Examples include drug discovery, personalized medicine, DNA sequencing, gene editing, metabolic engineering, and biomolecular sensing.
- Biological computing encompasses the development of algorithms and computational models inspired by biological processes and systems. Examples include genetic algorithms, evolutionary algorithms, neural networks, and swarm intelligence algorithms, which mimic the behavior of biological systems to solve optimization, classification, and pattern recognition problems.

9. Mobile Computing

- Mobile computing has transformed the way people work, communicate, shop, entertain themselves, and navigate the world. It has enabled new modes of interaction, collaboration, and commerce, and has become an indispensable part of modern life for billions of users worldwide.
- Mobile computing refers to the use of portable computing devices, such as smartphones, tablets, laptops, wearables, and other wireless-enabled devices, to access and process information, communicate, and run applications from portable devices from any location wirelessly/while on the move.
- Mobile computing enables users to stay connected, productive, and entertained regardless of their location, as long as they have access to a wireless network, such as Wi-Fi, cellular data, or Bluetooth.
- Mobile computing has become an integral part of our daily lives, enabling users to stay connected, productive, and entertained while on the go.
- The evolution of mobile computing has significantly impacted communication, business, education, healthcare, and various other sectors, providing users with unprecedented flexibility and connectivity.
- Mobile computing devices are designed to be lightweight, compact, and easy to carry, allowing users to take them anywhere and use them on the go.
- Mobile operating systems and applications support multitasking, allowing users to perform multiple tasks simultaneously, such as browsing the web, checking email, streaming media, and using productivity apps.
- Mobile computing platforms, such as iOS (Apple), Android (Google), and Windows Mobile (Microsoft), offer extensive lightweight app ecosystems with millions of applications available for download from app stores, providing a wide range of functionality and entertainment options.
- Mobile devices are equipped with wireless networking capabilities, such as Wi-Fi, cellular, Bluetooth, NFC (Near Field Communication), and GPS (Global Positioning System), enabling communication and data exchange without the need for physical connections.

- Mobile computing often involves integration with cloud services, allowing users to store, access, and synchronize data across multiple devices and platforms, and enabling seamless access to resources and services from anywhere with an internet connection.

10. Optical Computing

- Optical computing is an approach to computation that utilizes light instead of traditional electronic signals to perform various computing tasks.
- The fundamental idea behind optical computing is to use photons, the particles of light, to carry and process information.
- Optical switches and modulators play a crucial role in optical computing. These components are used to control the flow of optical signals, enabling the creation of optical circuits and devices.
- Optical computing takes advantage of principles like interference and superposition, which are properties of light waves. These properties allow for complex computations to be performed in parallel.
- Optical computing has the potential to overcome some of the limitations associated with traditional electronic computing, such as speed, power consumption, and heat generation.
- One of the strengths of optical computing lies in its ability to perform parallel processing efficiently. Optical systems can manipulate multiple light beams simultaneously, enabling the processing of multiple data streams in parallel.
- Wavelength Division Multiplexing(WDM) is a technique used in optical computing where multiple signals at different wavelengths are sent simultaneously through an optical medium. This allows for the transmission of multiple streams of data over the same optical fiber.
- Optical computing has the potential to reduce heat generation compared to electronic computing. Electronic devices can generate heat due to the resistance of materials, while photons, being massless, do not generate heat in the same way.
- Fiber optics, which involves the transmission of light through optical fibers, is a key application of optical computing. It is widely used in high-speed communication networks for data transmission over long distances.

11. Nano Computing

- Nano computing, also known as nanocomputing or molecular nanotechnology.
- Nanocomputing is a field of research and development that explores the use of nanoscale components and materials for building computational devices and systems.
- nano computing holds promise for revolutionizing computing and technology by enabling the development of smaller, faster, and more efficient devices and systems. Continued research and innovation in nanocomputing are expected to lead to breakthroughs in science, engineering, and medicine in the coming years.
- Nanocomputing aims to miniaturize electronic circuits and components to the nanometer scale, where individual atoms and molecules can be manipulated to perform computation

- Nanocomputing refers to the development and use of computing technology at the nanoscale, which involves structures and components at the scale of nanometers (one billionth of a meter).
- Nanocomputing involves the design, fabrication, and integration of electronic components and devices at the nanometer scale. This includes transistors, wires, switches, and other circuit elements that are built using nanomaterials and nanofabrication techniques.
- Nanocomputing involves the use of nanoscale materials and structures, such as nanowires, nanotubes, and nanodevices, to perform computational tasks.
- Nano-computing has the potential to significantly improve energy efficiency compared to traditional computing. The small size of nanoscale components allows for faster data transfer and reduced power consumption.
- Nanocomputing has potential applications in various domains, including electronics, information technology, medicine, energy, and materials science. For example, nano computing can lead to the development of faster and more energy-efficient electronic devices, advanced sensors and detectors, targeted drug delivery systems, and high-density data storage technologies.
- Nano computing faces several challenges also, including scalability, reliability, manufacturability, and cost-effectiveness. Fabricating nanoscale devices with high precision and yield, understanding and mitigating quantum effects, and integrating nano components into practical systems are areas of active research and development.

2. Cloud Computing Fundamentals

Modern computing with our laptop or desktop or even with tablets/smartphones using the Internet to access the data and details that we want, which are located/stored at remote places/computers, through the faces of applications like Facebook, e-mail, and YouTube, brings the actual power of information that we need instantaneously within no time.

Even if millions of users get connected in this manner, from anywhere in the world, these applications do serve what these users—customers want.

This phenomenon of supply of information or any other data and details to all the needy customers, as and when it is asked, is the conceptual understanding and working of what is known as cloud computing.

2.1 Motivation for Cloud Computing

The users who are in need of computing are expected to invest money on computing resources such as hardware, software, networking, and storage; this investment naturally costs a bulk currency to the users as they have to buy these computing resources, keep these in their premises, and maintain and make it operational—all these tasks would add cost. And, this is a particularly true and huge expenditure to the enterprises that require enormous computing power and resources, compared with classical academics and individuals.

On the other hand, it is easy and handy to get the required computing power and resources from some provider (or supplier) as and when it is needed and pay only for that usage. This would cost only a reasonable investment or spending, compared to the huge investment when buying the entire computing infrastructure. This phenomenon can be viewed as *capital expenditure* versus *operational expenditure*.

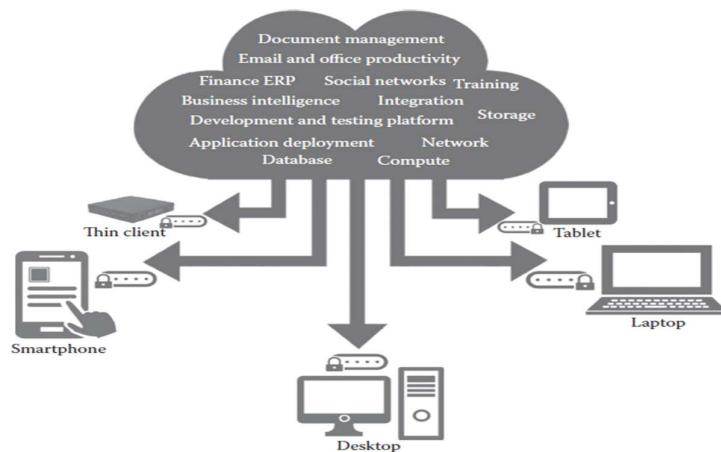
As one can easily assess the huge lump sum required for capital expenditure (whole investment and maintenance for computing infrastructure) and compare it with the moderate or smaller lump sum required for the hiring or getting the computing infrastructure only to the tune of required time, and rest of the time free from that.

Therefore, cloud computing is a mechanism of *bringing–hiring or getting the services of the computing power or infrastructure* to an organizational or individual level to the extent required and paying only for the consumed services.

One can compare this situation with the usage of electricity (its services) from its producer-cum-distributor (in India, it is the state-/government-owned electricity boards that give electricity supply to all residences and organizations) to houses or organizations; here, we do not generate electricity (comparable with electricity production-related tasks); rather, we use it only to tune up our requirements in our premises, such as for our lighting and usage of other electrical appliances, and pay as per the electricity meter reading value.

Therefore, cloud computing is needed in getting the services of computing resources. Thus, one can say as a one-line answer to the need for cloud computing that it eliminates a large computing investment without compromising the use of computing at the user level at an operational cost.

Cloud computing is very economical and saves a lot of money. A blind benefit of this computing is that even if we lose our laptop or due to some crisis our personal computer—and the desktop system—gets damaged, still our data and files will stay safe and secured as these are not in our local machine (but remotely located at the provider's place—machine).



The **cloud** represents the Internet-based computing resources, and the accessibility is through some secure support of connectivity. Cloud computing encompasses the subscription-based or pay-per-use service model of offering computing to end users or customers over the Internet and thereby extending the IT's existing capabilities.

The Need for Cloud Computing

The main reasons for the need and use of cloud computing are convenience and reliability. In the past, if we wanted to bring a file, we would have to save it to a Universal Serial Bus (USB) flash drive, external hard drive, or compact disc (CD) and bring that device to a different place.

Instead, saving a file to the cloud (e.g., use of cloud application Dropbox) ensures that we will be able to access it with any computer that has an Internet connection. The cloud also makes it much easier to share a file with friends, making it possible to collaborate over the web.

While using the cloud, losing our data/file is much less likely. However, just like anything online, there is always a risk that someone may try to gain access to our personal data, and therefore, it is important to choose an access control with a strong password and pay attention to any privacy settings for the cloud service that we are using.

3. Defining Cloud Computing

In the simplest terms, cloud computing means storing and accessing data and programs over the Internet from a remote location or computer instead of our computer's hard drive. This so called ***remote location*** has several properties such as scalability, elasticity etc., which is significantly different from a simple remote machine.

The cloud is just a metaphor for the Internet. When we store data on or run a program from the local computer's hard drive, that is called local storage and computing. For it to be considered *cloud computing*, we need to access our data or programs over the Internet. The end result is the same; however, with an online connection, cloud computing can be done anywhere, anytime, and by any device.

NIST Definition of Cloud Computing

The formal definition of cloud computing comes from the **National Institute of Standards and Technology (NIST)**: "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.

It means that the computing resource or infrastructure—be it server hardware, storage, network, or application software—all available from the cloud vendor or provider's site/premises, can be accessible over the Internet from any remote location and by any local computing device. In addition, the usage or accessibility is to cost only to the level of usage to the customers based on their needs and demands, also known as the *pay-as-you-go* or *pay-as-per-use* model.

To define and understand cloud computing from two other perspectives

- as a service
- as a platform

Cloud Computing Is a Service

The simplest thing that any computer does is allow us to store and retrieve information. We can store our family photographs, our favourite songs, or even save movies on it, which is also the most basic service offered by cloud computing.

The example of a popular application called **Flickr**. While Flickr started with an emphasis on sharing photos and images, it has emerged as a great place to store those images.

Cloud Computing Is a Platform

The World Wide Web (WWW) can be considered as the operating system for all our Internet-based applications. However, one has to understand that we will always need a local operating system in our computer to access web-based applications.

The basic meaning of the term **platform** is that it is the support on which applications run or give results to the users. For example, **Microsoft Windows is a platform**. But, a platform does not have to be an operating system. **Java** is a platform even though it is not an operating system.

Through cloud computing, the **web** is becoming a platform. With trends (applications) such as Office 2.0, more and more applications that were originally available on desktop computers are now being converted into web-cloud applications.

Word processors like Buzzword and office suites like Google Docs are now available in the cloud as their desktop counterparts. All these kinds of trends in providing applications via the cloud are turning cloud computing into a platform or to act as a platform.

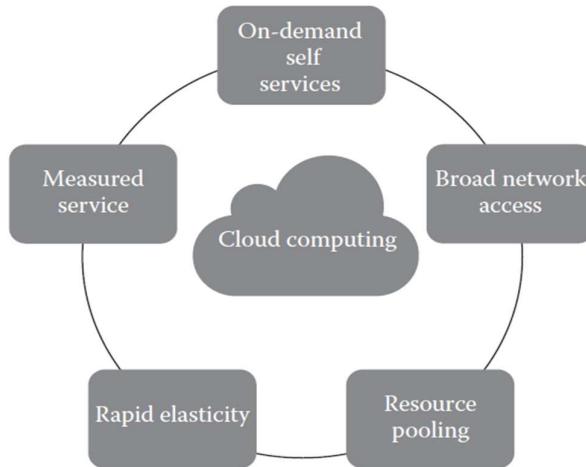
4. (5-4-3) Principles of Cloud computing

The 5-4-3 principles put forth by NIST describe (a) the five essential characteristic features that promote cloud computing, (b) the four deployment models that are used to narrate the cloud computing opportunities for customers while looking at architectural models, and (c) the three important and basic service offering models of cloud computing.

1 Five Essential Characteristics

Cloud computing has five essential characteristics, the word **essential**, which means that if any of these characteristics is missing, then it is not cloud computing:

- 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's 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, laptops, and personal digital assistants [PDAs]).



3. *Elastic resource pooling:* The provider's computing resources are pooled to serve multiple consumers using a multitenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify the location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, and network bandwidth.

4. *Rapid elasticity:* Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased 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.

2. Four Cloud Deployment Models

Deployment models describe the ways with which the cloud services can be deployed or made available to its customers, depending on the organizational structure and the provisioning location. One can understand it in this manner too: cloud (Internet)-based computing resources—that is, the locations where data and services are acquired and provisioned to its customers—can take various forms. Four deployment models are usually distinguished, namely, public, private, community, and hybrid cloud service usage:

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 some combination of them, and it may exist on or off premises.

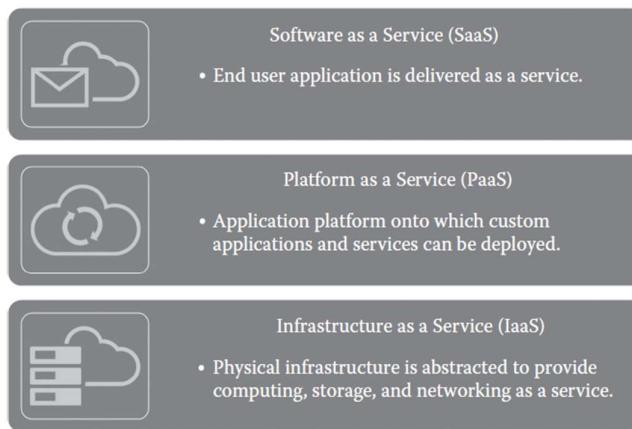
2. *Public cloud:* The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

3. Community cloud: 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). It may be managed by the organizations or a third party and may exist on premise or off premise.

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).

3. Three Service Offering Models

The three kinds of services with which the cloud-based computing resources are available to end customers are as follows: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). It is also known as the service–platform–infrastructure (SPI) model of the cloud.



- **SaaS** is a software distribution model in which applications (software, which is one of the most important computing resources) are hosted by a vendor or service provider and made available to customers over a network, typically the Internet. Typical applications offered as a service include customer relationship management (CRM), business intelligence analytics, and online accounting software.
- **PaaS** is a paradigm for delivering operating systems and associated services (e.g., computer aided software engineering [CASE] tools, integrated development environments [IDEs] for developing software solutions) over the Internet without downloads or installation. Examples of PaaS providers include Google App Engine and Microsoft Azure Services.
- IaaS involves outsourcing the equipment used to support operations, including storage, hardware, servers, networking components and other fundamental computing resources on a pay-per-use basis where he or she is able to deploy and run arbitrary software, which can include operating systems and applications.. Amazon Web Services (AWS) is a popular example of a large IaaS provider.

Benefits and Drawbacks

The benefits of cloud computing can be as follows

1. **Achieve economies of scale:** We can increase the volume output or productivity with fewer systems and thereby reduce the cost per unit of a project or product.
2. **Reduce spending on technology infrastructure:** It is easy to access data and information with minimal upfront spending in a *pay-as-you-go* approach, in the sense that the usage and payment are similar to an electricity meter reading in the house, which is based on demand.
3. **Globalize the workforce:** People worldwide can access the cloud with Internet connection.
4. **Streamline business processes:** It is possible to get more work done in less time with less resource.
5. **Reduce capital costs:** There is no need to spend huge money on hardware, software, or licensing fees.
6. **Pervasive accessibility:** Data and applications can be accessed anytime, anywhere, using any smart computing device, making our life so much easier.
7. **Monitor projects more effectively:** It is possible to confine within budgetary allocations and can be ahead of completion cycle times.
8. **Less personnel training is needed:** It takes fewer people to do more work on a cloud, with a minimal learning curve on hardware and software issues.
9. **Minimize maintenance and licensing software:** As there is no too much of on-premise computing resources, maintenance becomes simple and updates and renewals of software systems rely on the cloud vendor or provider.
10. **Improved flexibility:** It is possible to make fast changes in our work environment without serious issues at stake.

Drawbacks to cloud computing are obvious.

- The main point in this context is that if we lose our Internet connection, we have lost the link to the cloud and thereby to the data and applications. There is also a concern about security as our entire working with data and applications depend on other's (cloud vendor or providers) computing power.
- while cloud computing supports scalability (i.e., quickly scaling up and down computing resources depending on the need), it does not permit the control on these resources as these are not owned by the user or customer.
- Depending on the cloud vendor or provider, customers may face restrictions on the availability of applications, operating systems, and infrastructure options. And, sometimes, all development platforms may not be available in the cloud due to the fact that the cloud vendor may not aware of such solutions.
- A major barrier to cloud computing is the interoperability of applications, which is the ability of two or more applications that are required to support a business need to work together by sharing data and other business-related resources.
- Normally, this does not happen in the cloud as these applications may not be available with a single cloud vendor and two different vendors having these applications do not cooperate with each other.

5. Cloud Computing Architecture and Management

There are several processes and components of cloud computing are cloud computing Architecture, cloud anatomy, network connectivity in the cloud, managing the cloud.

- ❖ **Cloud architecture** consists of a hierarchical set of components that collectively describe the way the cloud works. Architecture is the hierarchical view of describing a technology. This usually includes the components over which the existing technology is built and the components that are dependent on the technology.
- ❖ **Anatomy** describes the core structure of the cloud. Once the structure of the cloud is clear, the network connections in the cloud and the details about the cloud application need to be known. This is important as the cloud is a completely Internet-dependent technology.
- ❖ **cloud management** describes the way an application and infrastructure in the cloud are managed. Management is important because of the quality of service (QoS) factors that are involved in the cloud. These QoS factors form the basis for cloud computing. All the services are given based on these QoS factors
- ❖ Similarly, **application migration** to the cloud also plays a very important role. Not all applications can be directly deployed to the cloud. An application needs to be properly migrated to the cloud to be considered a proper cloud application that will have all the properties of the cloud.

1. Cloud Architecture

Any technological model consists of an architecture based on which the model functions, which is a hierarchical view of describing the technology. The cloud also has an architecture that describes its working mechanism. It includes the dependencies on which it works and the components that work over it. The cloud is a recent technology that is completely dependent on the Internet for its functioning. The cloud architecture can be divided into four layers based on the access of the cloud by the user.

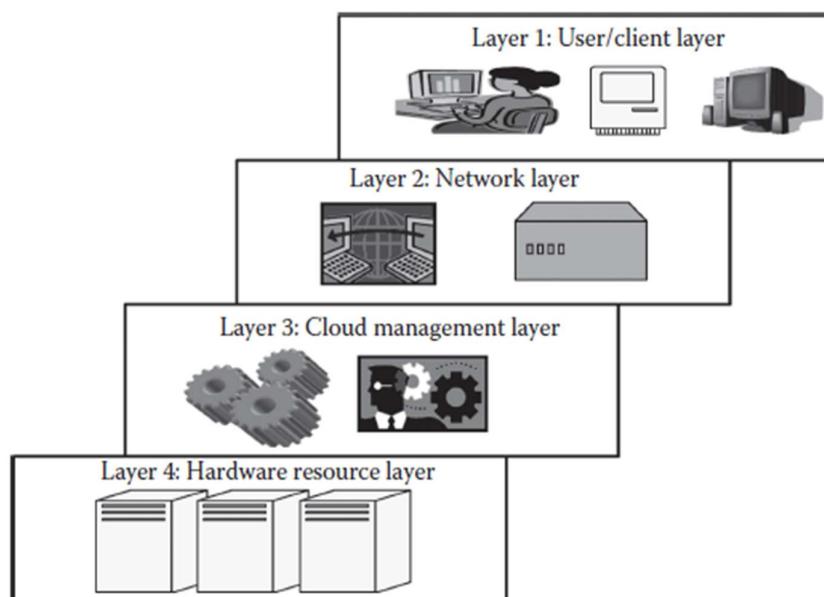


Fig. Cloud architecture.

They are as follows.

1. Layer 1 (User/Client Layer)
2. Layer 2 (Network Layer)
3. Layer 3 (Cloud Management Layer)
4. Layer 4 (Hardware Resource Layer)

1. Layer 1 (User/Client Layer):

This layer is the lowest layer in the cloud architecture. All the users or client belong to this layer. This is the place where the client/user initiates the connection to the cloud. The client can be any device such as a thin client, thick client, or mobile or any handheld device that would support basic functionalities to access a web application.

The thin client here refers to a device that is completely dependent on some other system for its complete functionality. In simple terms, they have very low processing capability.

Similarly, thick clients are general computers that have adequate processing capability. They have sufficient capability for independent work.

Usually, a cloud application can be accessed in the same way as a web application. But internally, the properties of cloud applications are significantly different. Thus, this layer consists of client devices.

2. Layer 2 (Network Layer):

This layer allows the users to connect to the cloud. The whole cloud infrastructure is dependent on this connection where the services are offered to the customers. This is primarily the Internet in the case of a public cloud.

- The public cloud usually exists in a specific location and the user would not know the location as it is abstract. And, the public cloud can be accessed all over the world.
- In the case of a private cloud, the connectivity may be provided by a local area network (LAN). Even in this case, the cloud completely depends on the network that is used.

Usually, when accessing the public or private cloud, the users require minimum bandwidth, which is sometimes defined by the cloud providers. This layer does not come under the purview of service-level agreements (SLAs), that is, SLAs do not take into account the Internet connection between the user and cloud for quality of service (QoS).

3. Layer 3 (Cloud Management Layer)

This layer consists of softwares that are used in managing the cloud. The softwares can be a cloud operating system (OS), a software that acts as an interface between the data center (actual resources) and the user, or a management software that allows managing resources. These softwares usually **allow resource management** (scheduling, provisioning, etc.), **optimization** (server consolidation, storage workload consolidation), **and internal cloud governance**.

This layer comes under the purview of SLAs, that is, the operations taking place in this layer would affect the SLAs that are being decided upon between the users and the service providers. Any delay in processing or any discrepancy in service provisioning may lead to an SLA violation.

As per rules, any SLA violation would result in a penalty to be given by the service provider. These SLAs are for both private and public clouds. Popular service providers are Amazon Web Services (AWS) and Microsoft Azure for public cloud. Similarly, OpenStack and Eucalyptus allow private cloud creation, deployment, and management.

4. Layer 4 (Hardware Resource Layer)

Layer 4 consists of provisions for actual hardware resources. Usually, in the case of a public cloud, a data center is used in the back end. Similarly, in a private cloud, it can be a data center, which is a huge collection of hardware resources interconnected to each other that is present in a specific location or a high configuration system.

This layer comes under the purview of SLAs. This is the most important layer that governs the SLAs. This layer affects the SLAs most in the case of data centers. Whenever a user accesses the cloud, it should be available to the users as quickly as possible and should be within the time that is defined by the SLAs.

As mentioned, if there is any discrepancy in provisioning the resources or application, the service provider has to pay the penalty. Hence, the data center consists of a high-speed network connection and a highly efficient algorithm to transfer the data from the data center to the manager. There can be a number of data centers for a cloud, and similarly, a number of clouds can share a data center.

Thus, this is the architecture of a cloud. The layering is strict, and for any cloud application, this is followed. There can be a little loose isolation between layer 3 and layer 4 depending on the way the cloud is deployed.

2. Anatomy of the Cloud

Cloud anatomy can be simply defined as the structure of the cloud. Cloud anatomy cannot be considered the same as cloud architecture. It may not include any dependency on which or over which the technology works,

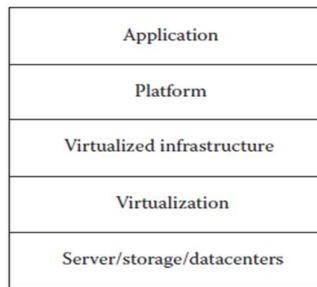


Fig. Cloud structure.

whereas architecture wholly defines and describes the technology over which it is working. Architecture is a hierarchical structural view that defines the technology as well as the technology over which it is dependent or/and the technology that are dependent on it. Thus, anatomy can be considered as a part of architecture.

There are basically five components of the cloud:

1. **Application:** The upper layer is the application layer. In this layer, any applications are executed.
2. **Platform:** This component consists of platforms that are responsible for the execution of the application. This platform is between the infrastructure and the application.
3. **Infrastructure:** The infrastructure consists of resources over which the other components work. This provides computational capability to the user.
4. **Virtualization:** Virtualization is the process of making logical components of resources over the existing physical resources. The logical components are isolated and independent, which form the infrastructure.
5. **Physical hardware:** The physical hardware is provided by server and storage units.

3. Network Connectivity in Cloud Computing

Cloud computing is a technique of resource sharing where servers, storage, and other computing infrastructure in multiple locations are connected by networks. In the cloud, when an application is submitted for its execution, needy and suitable resources are allocated from this collection of resources; as these resources are connected via the Internet, the users get their required results.

For many cloud computing applications, network performance will be the key issue to cloud computing performance. Since cloud computing has various deployment options, we now consider the important aspects related to the cloud deployment models and their accessibility from the viewpoint of network connectivity.

1. Public Cloud Access Networking

In this option, the connectivity is often through the Internet, though some cloud providers may be able to support virtual private networks (VPNs) for customers. Accessing public cloud services will always create issues related to security, which in turn is related to performance. One of the possible approaches toward the support of security is to promote connectivity through encrypted tunnels, so that the information may be sent via secure pipes on the Internet. This procedure will be an overhead in the connectivity, and using it will certainly increase delay and may impact performance.

If we want to reduce the delay without compromising security, then we have to select a suitable routing method such as the one reducing the delay by minimizing transit *hops* in the end-to-end connectivity between the cloud provider and cloud consumer. Since the end-to-end connectivity support is via the Internet, which is a complex federation of interconnected providers (known as Internet service providers [ISPs]), one has to look at the options of selecting the path.

2. Private Cloud Access Networking

In the private cloud deployment model, since the cloud is part of an organizational network, the technology and approaches are local to the in-house network structure. This may include an Internet VPN or VPN service from a network operator. If the application access was properly done with an

organizational network—connectivity in a *precloud* configuration—transition to private cloud computing will not affect the access performance.

3. Intracloud Networking for Public Cloud Services

Another network connectivity consideration in cloud computing is intracloud networking for public cloud services. Here, the resources of the cloud provider and thus the cloud service to the customer are based on the resources that are geographically apart from each other but still connected via the Internet. Public cloud computing networks are internal to the service provider and thus not visible to the user/customer; however, the security aspects of connectivity and the access mechanisms of the resources are important. Another issue to look for is the QoS in the connected resources worldwide. Most of the performance issues and violations from these are addressed in the SLAs commercially.

4. Private Intracloud Networking

The most complicated issue for networking and connectivity in cloud computing is private intracloud networking. What makes this particular issue so complex is that it depends on how much intracloud connectivity is associated with the applications being executed in this environment. Private intracloud networking is usually supported over connectivity between the major data center sites owned by the company. At a minimum, all cloud computing implementations will rely on intracloud networking to link users with the resource to which their application was assigned. Once the resource linkage is made, the extent to which intracloud networking is used depends on whether the application is componentized based on *service-oriented architecture (SOA)* or not, among multiple systems. If the principle of SOA is followed, then traffic may move between components of the application, as well as between the application and the user. The performance of those connections will then impact cloud computing performance overall. Here too, the impact of cloud computing performance is the differences that exist between the current application and the network relationships with the application.

There are reasons to consider the networks and connectivity in cloud computing with newer approaches as globalization and changing network requirements, especially those related to increased Internet usage, are demanding more flexibility in the network architectures of today's enterprises.

5. New Facets in Private Networks

Conventional private networks have been architected for on-premise applications and maximum Internet security. Typically, applications such as e-mail, file sharing, and *enterprise resource planning (ERP)* systems are delivered to on-premise-based servers at each corporate data center. Increasingly today, software vendors are offering Software as a Service (SaaS) as an alternative for their software support to the corporate offices, which brings more challenges in the access and usage mechanisms of software from data center servers and in the connectivity of network architectures. The traditional network architecture for these global enterprises was not designed to optimize performance for cloud applications, now that many applications including mission-critical applications are transitioning (moving) from on-premise based to cloud based, wherein the network

6. Path for Internet Traffic

The traditional Internet traffic through a limited set of Internet gateways poses performance and availability issues for end users who are using cloud-based applications. It can be improved if a more widely distributed Internet gateway infrastructure and connectivity are being supported for accessing applications, as they will provide lower-latency access to their cloud applications. As the volume of traffic to cloud applications grows, the percentage of the legacy network's capacity in terms of traffic to regional gateways increases. Applications such as video conferencing would hog more bandwidth while mission-critical applications such as ERP will consume less bandwidth, and hence, one has to plan a correct connectivity and path between providers and consumers.

4. Applications on the Cloud

The power of a computer is realized through the applications. There are several types of applications.

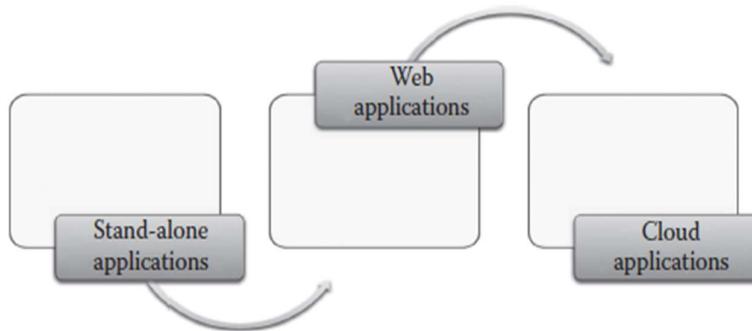


Fig. Computer application evolution.

The first type of applications that was developed and used was a **stand-alone application**.

- A stand-alone application is developed to be run on a single system that does not use network for its functioning.
- These stand-alone systems use only the machine in which they are installed.
- The functioning of these kinds of systems is totally dependent on the resources or features available within the system.
- These systems do not need the data or processing power of other systems; they are self-sustaining. But as the time passed, the requirements of the users changed and certain applications were required, which could be accessed by other users away from the systems.
- This led to the inception of web application.

The **web applications** were different from the stand-alone applications in many aspects.

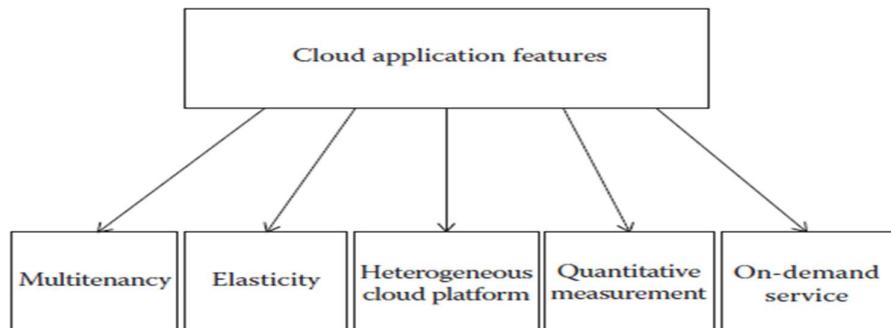
- The main difference was the client server architecture that was followed by the web application. Unlike stand-alone applications, these systems were totally dependent on the network for its working.
- Here, there are basically two components, called as the client and the server.
- The server is a high-end machine that consists of the web application installed. This web application is accessed from other client systems.

- The client can reside anywhere in the network. It can access the web application through the Internet.
- This type of application was very useful, and this is extensively used from its inception and now has become an important part of day-to-day life.
 - ✓ The web application is not elastic and cannot handle very heavy loads, that is, it cannot serve highly varying loads.
 - ✓ The web application is not multitenant.
 - ✓ The web application does not provide a quantitative measurement of the services that are given to the users, though they can monitor the user.
 - ✓ The web applications are usually in one particular platform.
 - ✓ The web applications are not provided on a pay-as-you-go basis; thus, a particular service is given to the user for permanent or trial use and usually the timings of user access cannot be monitored.
 - ✓ Due to its nonelastic nature, peak load transactions cannot be handled.

Primarily to solve the previously mentioned problem, the **cloud applications** were developed.

The cloud as mentioned can be classified into three broad access or service models, Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Cloud application in general refers to a SaaS application.

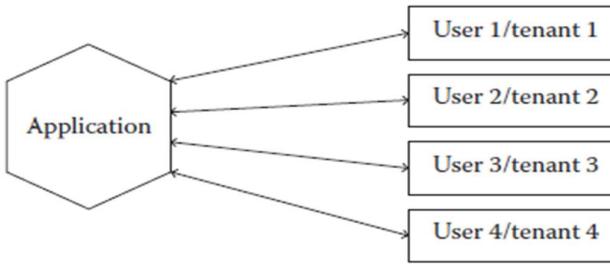
- A cloud application is different from other applications; they have unique features.
- A cloud application usually can be accessed as a web application but its properties differ.



1. **Multitenancy:** Multitenancy is one of the important properties of cloud that make it different from other types of application in which the software can be shared by different users with full independence. Here, independence refers to logical independence.

- Each user will have a separate application instance and the changes in one application would not affect the other.
- Physically, the software is shared and is not independent. The degree of physical isolation is very less.
- The logical independence is what is guaranteed. There are no restrictions in the number of applications being shared.

- The difficulty in providing logical isolation depends on the physical isolation to a certain extent.
- If an application is physically too close, then it becomes difficult to provide multitenancy. Web application and cloud application are similar as the users use the same way to access both.



- a multitenant application where several users share the same application.

2. Elasticity: Elasticity is also a unique property that enables the cloud to serve better. According to Herbst et al. [4], elasticity can be defined as the degree to which a system is able to adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner such that at each point in time, the available resources match the current demand as closely as possible. Elasticity allows the cloud providers to efficiently handle the number of users, from one to several hundreds of users at a time. In addition to this, it supports the rapid fluctuation of loads, that is, the increase or decrease in the number of users and their usage can rapidly change.

3. Heterogeneous cloud platform: The cloud platform supports heterogeneity, wherein any type of application can be deployed in the cloud. Because of this property, the cloud is flexible for the developers, which facilitates deployment. The applications that are usually deployed can be accessed by the users using a web browser.

4. Quantitative measurement: The services provided can be quantitatively measured. The user is usually offered services based on certain charges. Here, the application or resources are given as a utility on a pay-per-use basis. Thus, the use can be monitored and measured. Not only the services are measureable, but also the link usage and several other parameters that support cloud applications can be measured. This property of measuring the usage is usually not available in a web application and is a unique feature for cloud-based applications.

5. On-demand service: The cloud applications offer service to the user, on demand, that is, whenever the user requires it. The cloud service would allow the users to access web applications usually without any restrictions on time, duration, and type of device used.

5. Managing the Cloud

Cloud management is aimed at efficiently managing the cloud so as to maintain the QoS. It is one of the prime jobs to be considered. The whole cloud is dependent on the way it is managed.

Cloud management can be divided into two parts:

1. Managing the infrastructure of the cloud
2. Managing the cloud application

1. Managing the Cloud Infrastructure

The infrastructure of the cloud is considered to be the backbone of the cloud. This component is mainly responsible for the QoS factor. If the infrastructure is not properly managed, then the whole cloud can fail and QoS would be adversely affected. The core of cloud management is resource management. Resource management involves several internal tasks such as resource scheduling, provisioning, and load balancing. These tasks are mainly managed by the cloud service provider's core software capabilities such as the cloud OS that is responsible for providing services to the cloud and that internally controls the cloud. A cloud infrastructure is a very complex system that consists of a lot of resources. These resources are usually shared by several users.

Poor resource management may lead to several inefficiencies in terms of performance, functionality, and cost. If a resource is not efficiently managed, the performance of the whole system is affected. Performance is the most important aspect of the cloud, because everything in the cloud is dependent on the SLAs and the SLAs can be satisfied only if performance is good. Similarly, the basic functionality of the cloud should always be provided and considered at any cost. Even if there is a small discrepancy in providing the functionality, the whole purpose of maintaining the cloud is futile. A partially functional cloud would not satisfy the SLAs.

Lastly, the reason for which the cloud was developed was cost. The cost is a very important criterion as far as the business prospects of the cloud are concerned. On the part of the service providers, if they incur less cost for managing the cloud, then they would try to reduce the cost so as to get a strong user base. Hence, a lot of users would use the services, improving their profit margin. Similarly, if the cost of resource management is high, then definitely the cost of accessing the resources would be high and there is never a lossy business from any organization and so the service provider would not bear the cost and hence the users have to pay more. Similarly, this would prove costly for service providers as they have a high chance of losing a wide user base, leading to only a marginal growth in the industry. And, competing with its industry rivals would become a big issue. Hence, efficient management with less cost is required.

At a higher level, other than these three issues, there are few more issues that depend on resource management. These are power consumption and optimization of multiple objectives to further reduce the cost. To accomplish these tasks, there are several approaches followed, namely, consolidation of server and storage workloads. Consolidation would reduce the energy consumption and in some cases would increase the performance of the cloud. According to Margaret Rouse [5], server consolidation by definition is an approach

to the efficient usage of computer server resources in order to reduce the total number of servers or server locations that an organization requires.

The previously discussed prospects are mostly suitable for IaaS. Similarly, there are different management methods that are followed for different types of service delivery models. Each of the type has its own way of management. All the management methodologies are based on load fluctuation. Load fluctuation is the point where the workload of the system changes continuously. This is one of the important criteria and issues that should be considered for cloud applications. Load fluctuation can be divided into two types: predictable and unpredictable. Predictable load fluctuations are easy to handle. The cloud can be preconfigured for handling such kind of fluctuations. Whereas unpredictable load fluctuations are difficult to handle, ironically this is one of the reasons why cloud is preferred by several users.

This is as far as cloud management is concerned. Cloud governance is another topic that is closely related to cloud management. Cloud governance is different from cloud management. Governance in general is a term in the corporate world that generally involves the process of creating value to an organization by creating strategic objectives that will lead to the growth of the company and would maintain a certain level of control over the company. Similar to that, here cloud organization is involved.

2. Managing the Cloud Application

Business companies are increasingly looking to move or build their corporate applications on cloud platforms to improve agility or to meet dynamic requirements that exist in the globalization of businesses and responsiveness to market demands. But, this shift or moving the applications to the cloud environment brings new complexities. Applications become more composite and complex, which requires leveraging not only capabilities like storage and database offered by the cloud providers but also third-party SaaS capabilities like e-mail and messaging. So, understanding the availability of an application requires inspecting the infrastructure, the services it consumes, and the upkeep of the application. The composite nature of cloud applications requires visibility into all the services to determine the overall availability and uptime.

Cloud application management is to address these issues and propose solutions to make it possible to have insight into the application that runs in the cloud, as well as implement or enforce enterprise policies like governance and auditing and environment management while the application is deployed in the cloud. These cloud-based monitoring and management services can collect a multitude of events, analyze them, and identify critical information that requires additional remedial actions like adjusting capacity or provisioning new services. Additionally, application management has to be supported with tools and processes required for managing other environments that might coexist, enabling efficient operations.

6. Migrating Application to Cloud

Cloud migration encompasses moving one or more enterprise applications and their IT environments from the traditional hosting type to the cloud environment, either public, private, or hybrid. Cloud migration presents an opportunity to significantly reduce costs incurred on applications. This activity comprises, of different phases like evaluation, migration strategy, prototyping, provisioning, and testing.

Phases of Cloud Migration

1. **Evaluation:** Evaluation is carried out for all the components like current infrastructure and application architecture, environment in terms of compute, storage, monitoring, and management, SLAs, operational processes, financial considerations, risk, security, compliance, and licensing needs are identified to build a business case for moving to the cloud.
2. **Migration strategy:** Based on the evaluation, a migration strategy is drawn—a hotplug strategy is used where the applications and their data and interface dependencies are isolated and these applications can be operationalized all at once. A fusion strategy is used where the applications can be partially migrated; but for a portion of it, there are dependencies based on existing licenses, specialized server requirements like mainframes, or extensive interconnections with other applications.
3. **Prototyping:** Migration activity is preceded by a prototyping activity to validate and ensure that a small portion of the applications are tested on the cloud environment with test data setup.
4. **Provisioning:** Premigration optimizations identified are implemented. Cloud servers are provisioned for all the identified environments, necessary platform softwares and applications are deployed, configurations are tuned to match the new environment sizing, and databases and files are replicated. All internal and external integration points are properly configured. Web services, batch jobs, and operation and management software are set up in the new environments.
5. **Testing:** Postmigration tests are conducted to ensure that migration has been successful. Performance and load testing, failure and recovery testing, and scale-out testing are conducted against the expected traffic load and resource utilization levels.

Approaches for Cloud Migration

The following are the four broad approaches for cloud migration that have been adopted effectively by vendors:

1. **Migrate existing applications:** Rebuild or rearchitect some or all the applications, taking advantage of some of the virtualization technologies around to accelerate the work. But, it requires top engineers to develop new functionality. This can be achieved over the course of several releases with the timing determined by customer demand.
2. **Start from scratch:** Rather than cannibalize sales, confuse customers with choice, and tie up engineers trying to rebuild existing application, it may be easier to start again. Many of the R&D decisions will be different now, and with some of the more sophisticated development environments, one can achieve more even with a small focused working team.

3. ***Separate company:*** One may want to create a whole new company with separate brand, management, R&D, and sales. The investment and internet protocol (IP) may come from the existing company, but many of the conflicts disappear once a new *born in the cloud* company is established. The separate company may even be a subsidiary of the existing company. What is important is that the new company can act, operate, and behave like a cloud-based start-up.
4. ***Buy an existing cloud vendor:*** For a large established vendor, buying a cloud-based competitor achieves two things. Firstly, it removes a competitor, and secondly, it enables the vendor to hit the ground running in the cloud space. The risk of course is that the innovation, drive, and operational approach of the cloud-based company are destroyed as it is merged into the larger acquirer.