Unit-III

- **Syllabus**

  Cloud Computing Architecture and Management:

  - Cloud architecture, Layer,
  - Anatomy of the Cloud,
  - Network Connectivity in Cloud Computing,
  - Applications on the Cloud

- **Managing the Cloud**
  - Managing the Cloud Infrastructure,
  - Managing the Cloud application.

- **Migrating Application to Cloud**
  - Phases of Cloud Migration,
  - Approaches for Cloud Migration.

Cloud architecture

- The cloud is a recent technology that is completely dependent on the Internet for its functioning.
- Figure 1 depicts the architecture. The cloud architecture can be divided into four layers based on the access of the cloud by the user.

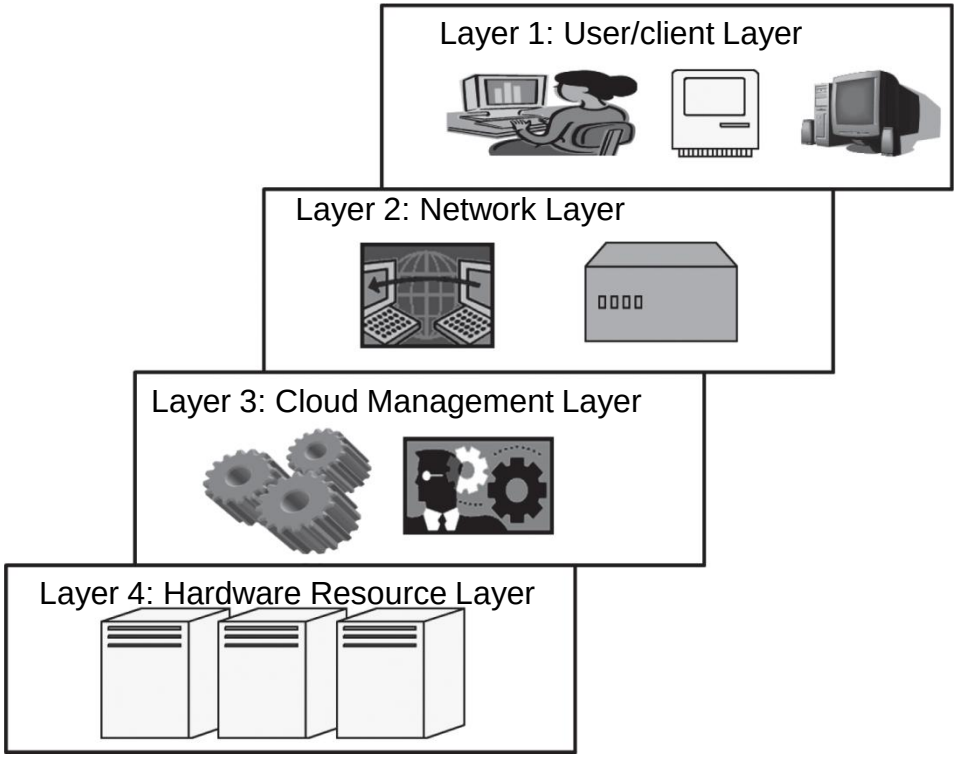


Figure1: Cloud Architecture

Cloud architecture

- 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.
- 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.
- 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,

Cloud architecture

- 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.

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.

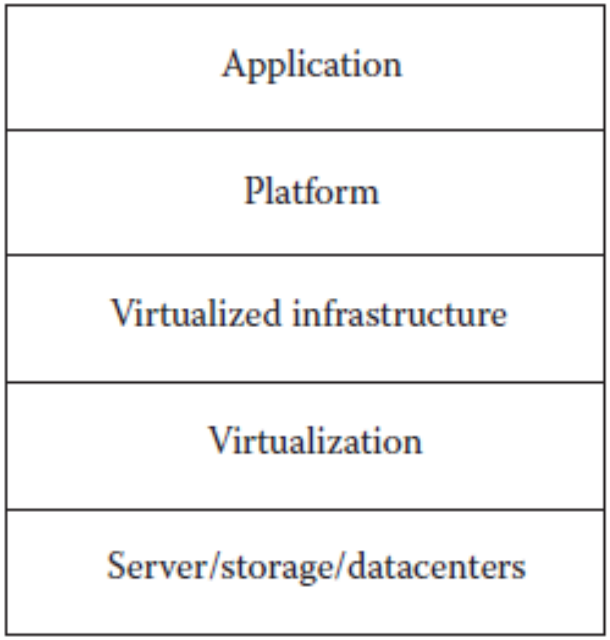


Figure2 : Cloud Structure

Anatomy 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.

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.

Network Connectivity in Cloud Computing

- There are 4 ways of network connectivity in cloud.
  These are

  1. Public Cloud Access Networking
  2. Private Cloud Access Networking
  3. Intracloud Networking for Public Cloud Services
  4. Private Intracloud Networking
     - New Facets in Private Networks
     - Path for Internet Traffic

Network Connectivity in Cloud Computing

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.
- 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])

Network Connectivity in Cloud Computing

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.

Network Connectivity in Cloud Computing

3. 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.

Network Connectivity in Cloud Computing

4. Private Intracloud Networking

- The most complicated issue for networking and connectivity in cloud computing is private intracloud networking.

- 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.*

Network Connectivity in Cloud Computing

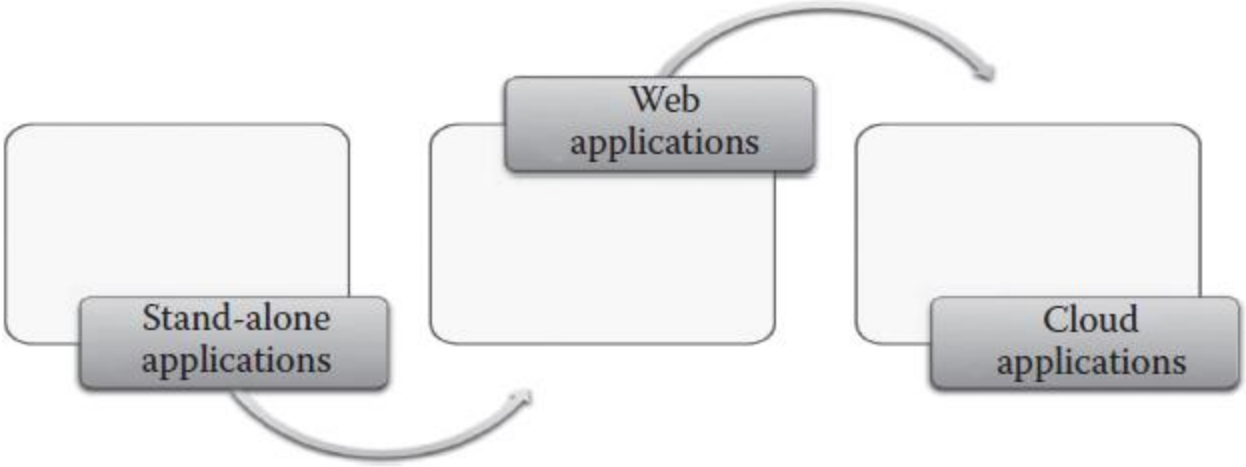
- New Facets in Private Networks
  - Conventional private networks have been architected for on-premise applications and maximum Internet security.
  - Typically, applications such as e-mail, file sharing, and *enterprise resource planning (ERP) systems are delivered to on-premise-based servers at each corporate data center.*
- 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.

Network Connectivity in Cloud Computing

- 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.*
- 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.

Applications on the Cloud

- 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



- 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.

Applications on the Cloud

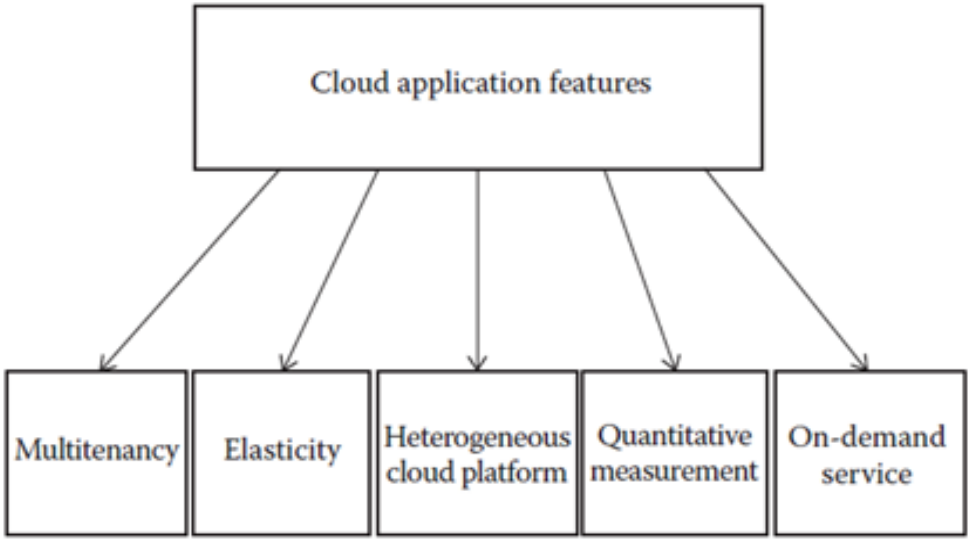
Though this application is much used, there are shortcomings of web applications as follows.

- 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.

Applications on the Cloud

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.



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.

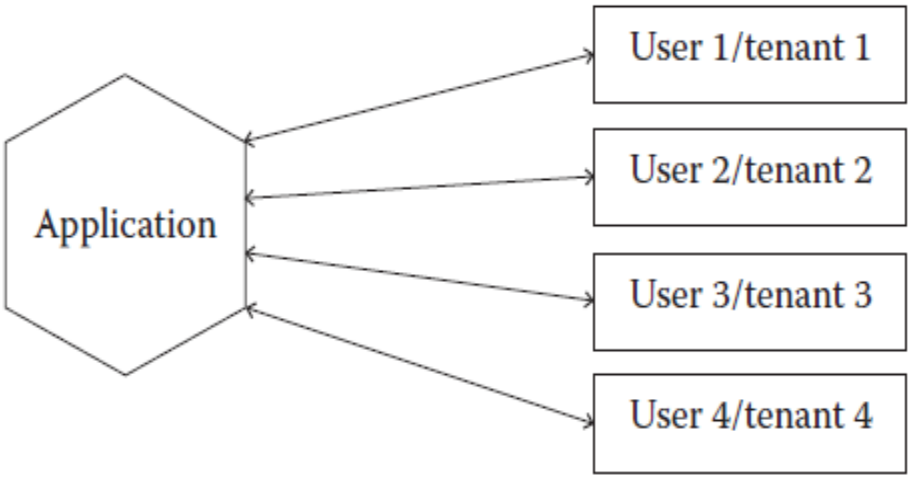


Fig: Multitenancy.

Elasticity

- Elasticity is also a unique property that enables the cloud to serve better 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.

Scalability Vs Elasticity

Increasing the capacity to meet the increasing workload	"Increasing or reducing" the capacity to meet the increasing or reducing workload
In a scaling environment, the available resources may exceed to meet the future demands	In elasticity environment the available resources matches the current demands
Scalability adapts only to the workload increase by provisioning the resources in an incremental manner	It adopts to both the workload increase & workload decrease in an automatic manner
Scalability enables a corporate to meet expected demands for services with long term strategic needs	Elasticity enables a corporate to meet the unexpected changes in the demand for service with "short-term", tactical needs

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.

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.

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.

1. Managing the Cloud Infrastructure

- Poor resource management may lead to several inefficiencies in terms of performance, functionality, and cost.

- 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.

- 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.

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

2. Managing the cloud application

- 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.

2.Managing the cloud application

- 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.

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
    2. Migration strategy
    3. Prototyping
    4. Provisioning
    5. Testing

  - Approaches for Cloud Migration

    1. Migrate existing applications
    2. Start from scratch
    3. Separate company
    4. Buy an existing cloud vendor:

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.

Phases of Cloud Migration

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:

Pre-migration optimizations identified are implemented. Cloud servers are provisioned for all the identified environments, necessary platform software's 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.

Phases of Cloud Migration

5. Testing:

Post-migration 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 re-architect 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.

Approaches for Cloud Migration

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.