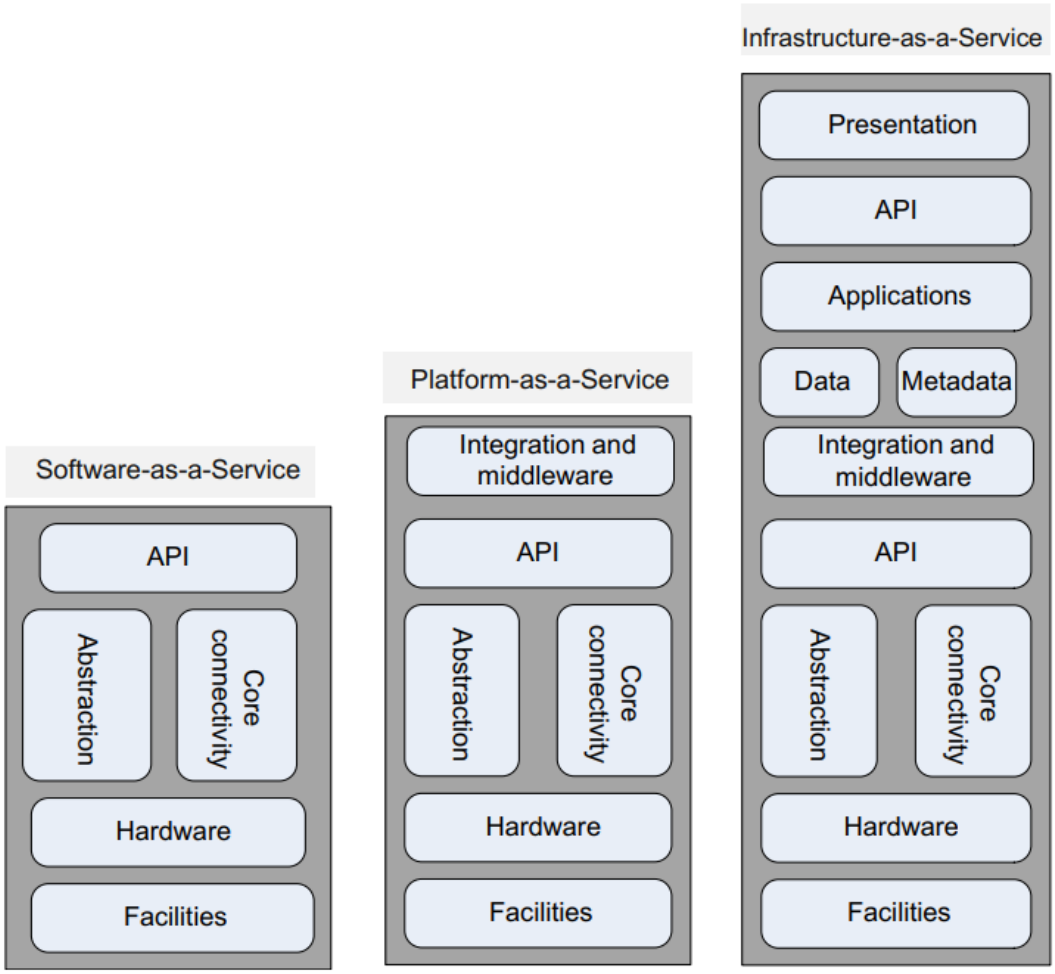
Unit – V

Introduction to Cloud Computing

Three Delivery Models:



Software-as-a-Service (SaaS):

- Software-as-a-Service (SaaS) gives the capability to use applications supplied by the service provider in a cloud infrastructure. The applications are accessible from various client devices through a thin-client interface such as a Web browser (e.g., Web-based email).
- The user does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- Services offered include:
  - Enterprise services such as workflow management, groupware and collaborative, supply chain, communications, digital signature, customer relationship management (CRM), desktop software, financial management, geo-spatial, and search [32].
  - Web 2.0 applications such as metadata management, social networking, blogs, wiki services, and portal services.
- The SaaS is not suitable for applications that require real-time response or those for which data is not allowed to be hosted externally.
- The most likely candidates for SaaS are applications for which:
  - Many competitors use the same product, such as email.
  - Periodically there is a significant peak in demand, such as billing and payroll.
  - There is a need for Web or mobile access, such as mobile sales management software.
  - There is only a short-term need, such as collaborative software for a project.

Platform-as-a-Service (PaaS):

- Platform-as-a-Service (PaaS) gives the capability to deploy consumer-created or acquired applications using programming languages and tools supported by the provider.
- The user does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, or storage.
- The user has control over the deployed applications and, possibly, over the application hosting environment configurations.
- Such services include session management, device integration, sandboxes, instrumentation and testing, contents management, knowledge management, and Universal Description, Discovery, and Integration (UDDI), a platform-independent Extensible Markup Language (XML)-based registry providing a mechanism to register and locate Web service applications.
- PaaS is not particularly useful when the application must be portable, when proprietary programming languages are used, or when the underlaying hardware and software must be customized to improve the performance of the application.
- The major PaaS application areas are in software development where multiple developers and users collaborate and the deployment and testing services should be automated.

Infrastructure-as-a-Service (IaaS):

- Infrastructure-as-a-Service (IaaS) is the capability to provision processing, storage, networks, and other fundamental computing resources;
- the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.
- The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of some networking components, such as host firewalls.
- Services offered by this delivery model include: server hosting, Web servers, storage, computing hardware, operating systems, virtual instances, load balancing, Internet access, and bandwidth provisioning.
- The IaaS cloud computing delivery model has a number of characteristics, such as the fact that the resources are distributed and support dynamic scaling, it is based on a utility pricing model and variable cost, and the hardware is shared among multiple users.
- This cloud computing model is particularly useful when the demand is volatile and a new business needs computing resources and does not want to invest in a computing infrastructure or when an organization is expanding rapidly.

A number of activities are necessary to support the three delivery models they include:

1. **Service Management.**
2. **Security Management.**
3. **Customer Services.**
4. **Integration Services.**

Ethical Issues in Cloud Computing:

- Cloud computing is based on a paradigm shift with profound implications for computing ethics. The main elements of this shift are:
  (i) the control is relinquished to third-party services;
  (ii) the data is stored on multiple sites administered by several organizations; and
  (iii) multiple services interoperate across the network.
- Unauthorized access, data corruption, infrastructure failure, and service unavailability are some of the risks related to relinquishing the control to third-party services;
- Systems can span the boundaries of multiple organizations and cross security borders, a process called *deperimeterization*. As a result of *deperimeterization*, "not only the border of the organization's IT infrastructure blurs, also the border of the accountability becomes less clear".
- The complex structure of cloud services can make it difficult to determine who is responsible in case something undesirable happens. In a complex chain of events or systems, many entities contribute to an action, with undesirable consequences. Some of them have the opportunity to prevent these consequences, and therefore no one can be held responsible – the so-called "problem of many hands."
- Cloud service providers have already collected petabytes of sensitive personal information stored in data centres around the world.
- Privacy is affected by cultural differences; though some cultures favor privacy, other cultures emphasize community, and this leads to an ambivalent attitude toward privacy on the Internet, which is a global system.
- Accountability is a necessary ingredient of cloud computing; adequate information about how data is handled within the cloud and about allocation of responsibility are key elements for enforcing ethics rules in cloud computing.
- Unwanted dependency on a cloud service provider, the so-called vendor lock-in, is a serious concern, and the current standardization efforts at NIST attempt to address this problem

Challenges Faced by Cloud Computing:

- Cloud computing inherits some of the challenges of parallel and distributed computing
- The specific challenges differ for the three cloud delivery models, but in all cases the difficulties are created by the very nature of utility computing, which is based on resource sharing and resource virtualization and requires a different trust model than the ubiquitous user-centric model we have been accustomed to for a very long time.
- The most significant challenge is security
- gaining the trust of a large user base is critical for the future of cloud computing. It is unrealistic to expect that a public cloud will provide a suitable environment for all applications.
- Many real-time applications will probably still be confined to private clouds. Some applications may be best served by a hybrid cloud setup; such applications could keep sensitive data on a private cloud and use a public cloud for some of the processing.
- The SaaS model faces similar challenges as other online services required to protect private information, such as financial or healthcare services.
- The IaaS model is by far the most challenging to defend against attacks. Indeed, an IaaS user has considerably more degrees of freedom than the other two cloud delivery models. An additional source of concern is that the considerable resources of a cloud could be used to initiate attacks against the network and the computing infrastructure.
- Virtualization is a critical design option for this model, but it exposes the system to new sources of attack. The trusted computing base (TCB) of a virtual environment includes not only the hardware and the hypervisor but also the management operating system.

- The next major challenge is related to resource management on a cloud. Any systematic rather than ad hoc resource management strategy requires the existence of controllers tasked to implement several classes of policies: admission control, capacity allocation, load balancing, energy optimization, and last but not least, to provide QoS guarantees.
- To implement these policies the controllers, need accurate information about the global state of the system. Determining the state of a complex system with 106 servers or more, distributed over a large geographic area, is not feasible.
- It seems reasonable to expect that such a complex system can only function based on self-management principles. But self-management and self-organization raise the bar for the implementation of logging and auditing procedures critical to the security and trust in a provider of cloud computing services.
- Under self-management it becomes next to impossible to identify the reasons that a certain action that resulted in a security breach was taken.
- The last major challenge we want to address is related to interoperability and standardization. Vendor lock-in, the fact that a user is tied to a particular cloud service provider, is a major concern for cloud users
- Standardization would support interoperability and thus alleviate some of the fears that a service critical for a large organization may not be available for an extended period of time. But imposing standards at a time when a technology is still evolving is not only challenging, it can be counterproductive because it may stifle innovation.

Different Types of Clouds:

1. **Private cloud:** The infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on or off the premises of the organization.
2. **Community cloud:** The 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 premises or off premises.
3. **Public cloud**: The infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
4. **Hybrid cloud**: The infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

Reasons for the Success of Cloud Computing:

- Cloud computing is in a better position to exploit recent advances in software, networking, storage, and processor technologies. Cloud computing is promoted by large IT companies where these new technological developments take place, and these companies have a vested interest in promoting the new technologies.
- A cloud consists of a homogeneous set of hardware and software resources in a single administrative domain. In this setup, security, resource management, fault tolerance, and quality of service are less challenging than in a heterogeneous environment with resources in multiple administrative domains.
- Cloud computing is focused on enterprise computing; its adoption by industrial organizations, financial institutions, healthcare organizations, and so on has a potentially huge impact on the economy.

- A cloud provides the illusion of infinite computing resources; its elasticity frees application designers from the confinement of a single system.
- A cloud eliminates the need for up-front financial commitment, and it is based on a pay-as-you-go approach. This has the potential to attract new applications and new users for existing applications, fomenting a new era of industrywide technological advancements.

Characteristics of Network centric computing and network centric content:

- Most applications are data-intensive.
- Virtually all applications are network-intensive.
- The systems are accessed using thin clients running on systems with limited resources.
- The infrastructure supports some form of workflow management.
- Computing and communication resources (CPU cycles, storage, network bandwidth) are shared and resources can be aggregated to support data-intensive applications. Multiplexing leads to a higher resource utilization.
- Data sharing facilitates collaborative activities.
- Cost reduction.
- User convenience and elasticity, that is the ability to accommodate workloads with very large peak to - average ratio.
- Storage technology has also evolved dramatically.