**Cloud Computing**

**Unit I**

**Introduction**

Cloud computing is a way of delivering computing services like storage, servers, and applications over the Internet, like utilities like electricity or water. Users can access these services as needed, without worrying about where they are hosted.

It allows resources to be provided dynamically, such as computing power, storage, and IT infrastructure, on a pay-per-use basis. This means anyone can quickly set up servers for their applications and adjust them based on demand, paying only for what they use.

This concept has evolved from ideas like grid computing and focuses on designing better systems, building applications, and using existing services more efficiently.

This chapter covers the basics of cloud computing, its features, key technologies, and the advancements that made it possible.

**Cloud Computing at a Glance**

In 1969, Leonard Kleinrock, one of the pioneers of ARPANET (the precursor to the internet), envisioned computing as a utility, much like electricity or water. This vision has materialized into what we now call **cloud computing**—a model where computing services are provisioned on demand, enabling users to access resources without concern for their physical location or infrastructure maintenance.

Cloud computing eliminates the need for heavy investments in IT infrastructure, offering services like hardware, software, and storage as utilities. Users only pay for what they use, fostering cost efficiency and accessibility.

Cloud computing employs **virtualization** to enable resource consolidation, scalability, and flexibility, making it a cornerstone of modern IT systems.

**The Vision of Cloud Computing**

Cloud computing embodies the transformation of IT resources into utilities that can be provisioned dynamically and on demand. This model allows:

1. **Flexible Service Usage**: Users can access virtual hardware, runtime environments, and services as needed, with no upfront commitments.
2. **Rapid Deployment**: Systems can be composed and deployed in hours rather than days, reducing maintenance efforts.
3. **Global Accessibility**: End users can access applications and data anytime, from any device connected to the internet.
4. **Open Marketplaces**: A future where cloud services can be traded as utilities in an open market, overcoming vendor lock-in and enabling automated service discovery.

By concentrating resources in centralized data centres, cloud computing promotes **greener IT** through optimized resource usage, reducing energy waste and carbon emissions.

**Defining a Cloud**

The term **cloud** originates from the telecommunications industry, symbolizing network abstraction in system diagrams. Today, it signifies an **internet-centric computing model** that delivers IT resources as services.

Key elements of cloud computing include:

1. **Everything as a Service (XaaS)**: From infrastructure to software, all components can be delivered as services, measured, and priced accordingly.
2. **Utility-Oriented Approach**: Services follow a **pay-per-use** model, avoiding upfront costs and enabling flexible scaling.
3. **Web-Based Accessibility**: Services are accessible via web browsers or APIs, making them platform-agnostic.

**NIST Definition:**

The National Institute of Standards and Technology (NIST) defines cloud computing as:

"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 definition highlights the **shared, scalable, and accessible nature** of cloud resources.

**A Closer Look**

Cloud computing is revolutionizing enterprises, governments, institutions, and organizations by enabling demand-driven, cost-efficient computing. Accessing cloud resources is now as seamless as completing a credit card transaction online. Some examples of cloud-powered solutions include:

* **Large Enterprises**: The *New York Times* converted its digital library into a web-friendly format in 36 hours using Amazon EC2 and S3, avoiding infrastructure costs.
* **Small Businesses**: *Animoto*, a video creation service, scales its processing power on Amazon Web Services to handle dynamic workloads without owning servers.
* **Developers**: *Little Fluffy Toys* launched a widget for bike rental services in one week using Google AppEngine, focusing solely on business logic.
* **End Users**: Apple’s *iCloud* allows users to access and sync data seamlessly across devices, eliminating the need for manual setup or connectivity.

These examples illustrate how cloud computing delivers IT services on a **pay-as-you-go basis**, fostering scalability and efficiency.

**Deployment Models**

* **Public Cloud**: Hosted by third-party providers, offering subscription-based resources.
* **Private Cloud**: In-house solutions for organizations needing high security and control, such as the U.S. federal government’s cloud initiative.
* **Hybrid Cloud**: Combines public and private clouds to meet quality-of-service requirements while addressing specific organizational needs.

**The Cloud Computing Reference Model**

The **Cloud Computing Reference Model** organizes services into three layers, offering diverse IT capabilities:

1. **Infrastructure-as-a-Service (IaaS)**:
   * Provides virtual hardware, storage, and networking on demand.
   * Examples: Virtual machine instances, persistent storage, and virtual networking services.
   * Ideal for building scalable systems with custom software stacks.
2. **Platform-as-a-Service (PaaS)**:
   * Delivers scalable runtime environments for hosting applications.
   * Enables developers to focus on application logic, while providers manage scalability and fault tolerance.
   * Suitable for building new systems without worrying about infrastructure complexity.
3. **Software-as-a-Service (SaaS)**:
   * Offers end-user applications like CRM, document management, and social networking platforms via browsers.
   * Targets users needing scalable, ready-made solutions with minimal customization.

Each layer addresses unique user needs, from developers building scalable systems to businesses leveraging ready-to-use applications.

**Characteristics and Benefits of Cloud Computing**

1. **No Up-Front Commitments**
   * Pay-as-you-go model reduces initial capital costs.
   * Enables organizations, especially startups, to grow without heavy investments.
2. **On-Demand Access**
   * IT resources are available when needed, enhancing agility and reducing the need for capacity planning.
   * Supports dynamic and flexible software composition.
3. **Simplified Scalability and Efficiency**
   * Easily scale infrastructure and applications to handle spikes in demand.
   * Multitenancy ensures optimal resource utilization and energy efficiency.
4. **Cost-Efficiency**
   * Reduced maintenance and operational costs by offloading these responsibilities to providers.
   * IT infrastructure transforms into a utility, improving the company’s net gain.
5. **Universal Accessibility**
   * Access data and applications anytime, anywhere, on multiple devices.
   * Web-based interfaces offer consistent user experiences without substantial software investments.
6. **Eco-Friendly**
   * Large datacentres achieve energy optimization, reducing environmental impact.

**Challenges in Cloud Computing**

1. **Dynamic Resource Provisioning**
   * Deciding optimal resources and duration for provisioning to maximize benefits.
   * Technical issues with virtualization and large-scale infrastructure management.
2. **Security Concerns**
   * Confidentiality of sensitive data in shared infrastructure.
   * Risk of data exposure due to vulnerabilities in memory during processing.
3. **Legal and Regulatory Issues**
   * Varying privacy laws across countries complicate data management.
   * Potential conflicts arise with cross-border data storage and government access.
4. **Integration of Real and Virtual Infrastructure**
   * Balancing real and virtual components poses technical and security challenges.
5. **Service Assurance**
   * Need for compliance with security standards and guarantees for confidentiality.

**Historical Developments in Cloud Computing**

**Overview:**  
The concept of renting computing services dates back to the 1950s with mainframes. Over time, technologies like distributed systems, virtualization, Web 2.0, service orientation, and utility computing laid the foundation for cloud computing.

**Key Technologies in Cloud Computing Evolution:**

**Distributed Systems:**

* + Definition: A collection of independent computers that appear as one system to users (Tanenbaum).
  + Characteristics: Scalability, transparency, concurrency, and availability.
  + Cloud Connection: Resources (infrastructure, services) are shared and rented to users via distributed systems, often hosted by IT giants like Amazon and Google.

1. **Three Milestones Leading to Clouds:**
   * **Mainframes:**
     + Large, reliable, multi-processor systems used for tasks like transaction processing.
     + Offered high reliability and fault tolerance.
     + Still in use for applications like online banking and government services.
   * **Cluster Computing:**
     + Emerged in the 1980s as a cheaper alternative to mainframes.
     + Commodity machines connected via high-bandwidth networks.
     + Used for high-performance computing, making it more accessible to universities and small labs.
   * **Grid Computing:**
     + Evolved from clusters in the 1990s, offering resources (computing power, storage) as utilities like electricity.
     + Large-scale networks of geographically dispersed clusters enabled through the Internet.

**Web 2.0 and Its Role in Cloud Computing:**

1. **Definition:**
   * Interactive, user-centered technologies transforming the Web into a platform for dynamic applications.
2. **Technologies Used:**
   * XML, AJAX, Web Services.
3. **Key Features:**
   * Enhanced interactivity, collaboration, and continuous updates without user-side deployment.
   * Lightweight and flexible application development models.
   * Applications composed by integrating existing services (loose coupling).
4. **Examples:**
   * Google Docs, Facebook, Twitter, YouTube, and Wikipedia.
   * Social networking sites leverage Web 2.0 tools (e.g., AJAX, RSS) for dynamic user interaction and content creation.
5. **Impact on Cloud Computing:**
   * Web 2.0 provides the interface for cloud services, supporting dynamism and user needs.

**Service-Oriented Computing (SOC)**  
SOC is the foundation of cloud computing, where services are the core components of system development. A service is a platform-independent, self-describing unit of functionality accessible via a network. Key concepts in SOC include:

* **Quality of Service (QoS)**: Metrics like performance, security, and reliability that define service behavior, often specified through Service Level Agreements (SLAs).
* **Software-as-a-Service (SaaS)**: A delivery model where software is provided over the internet, eliminating maintenance costs for clients and enabling multitenancy.
* **Web Services**: Implement SOC by making services accessible over the web using standards like SOAP and WSDL. This allows services to be reused across platforms.

SOC's major contribution to cloud computing is the ability to create distributed systems composed of services that can be easily integrated into applications.

**Utility-Oriented Computing**  
Utility computing envisions computing resources (e.g., storage, compute power) as services offered on a pay-per-use basis. This model was first hinted at by John McCarthy in 1961 and evolved through mainframe, cluster, and grid computing.

* **Early examples**: Mainframe providers offered computing power to organizations on-demand.
* **Grid computing**: Advanced utility computing by enabling resource access over the internet, leading to market-oriented services.
* **Cloud computing**: Fully realized utility computing, offering infrastructure, applications, and services for pay-per-use consumption.