**Introduction to Cloud Computing**

**Introduction**

The term ‘cloud’ that represents the internet, or various parts of it and ‘computing’ refers to the processing on those various resources of internet. The concept evolved during the middle of the first decade of the new millennium. The main fact of cloud computing is based on the following question,

**“Why we purchase resources if we can rent them?”**

‘Cloud computing’ is a relatively recent term, builds on decades of research in virtualization, distributed computing, utility computing, and more recently networking, web and software services. It implies a service oriented architecture, reduced information technology overhead for the end-user, great flexibility, reduced total cost of ownership, on-demand services and many other things. The approach is motivated by the idea that information processing can be done more efficiently on large farms of computing and storage systems accessible via the Internet.

|  |
| --- |
| **NIST definition of cloud computing :** *Cloud computing is a model for enabling 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*. |



Figure-1: Sharing resources among devices in a Cloud Environment

**WHY CLOUD COMPUTING?**

Traditional infrastructure provisioning model is inefficient and does not meet the requirements of the internet era. In this system centric model, once the need for a business application is identified, its infrastructure needs are identified and a request for infrastructure is placed with the IT infrastructure team that procures and provisions the infrastructure. The application is then developed, tested and deployed on that infrastructure.

**Characteristics of cloud computing**

* **Scalability and Elasticity**: It refers to the ability of dynamically acquiring computing resources and supporting a variable workload. A cloud provides the illusion of infinite computing resources; its elasticity frees the applications designers from the confinement of a single system. A cloud service provider maintains a massive infrastructure to support elastic services. That’s why cloud computing sometimes called as “Elastic Computing”.
* Cloud computing is **cost-effective** due to resource multiplexing; lower costs for the service provider are passed on to the cloud users.
* The maintenance and security are ensured by service providers.
* **Shared Infrastructure:** Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.
* 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 to promote the new technologies.
* **Centralized and Homogeneous:** 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.
* ***Virtualization*:** Virtualization is a technology that abstracts away the details of physical hardware and provides virtualized resources for high-level applications. A virtualized server is commonly called a virtual machine (VM). Virtualization forms the foundation of cloud computing, as it provides the capability of pooling computing resources from clusters of servers and dynamically assigning or reassigning virtual resources to applications on-demand.

**Cloud computing vs. Grid Computing**

A *computing grid* is a distributed system consisting of a large number of loosely coupled, heterogeneous, and geographically dispersed systems in different administrative domains. The vision of the grid movement was to give a user the illusion of a very large virtual supercomputer.

Two basic assumptions about the infrastructure prevented the grid movement from having the impact its supporters were hoping for. The first is the heterogeneity of the individual systems interconnected by the grid; the second is that systems in different administrative domain are expected to cooperate seamlessly. Indeed, the heterogeneity of the hardware and of system software poses significant challenges for application development and for application mobility. At the same, time critical areas of system management including scheduling, optimization of resource allocation, load balancing, and fault-tolerance are extremely difficult in a heterogeneous system. The fact that resources are in different administrative domains further complicates many, already difficult, problems related to security and resource management. While very popular in the science and the engineering community, the grid movement did not address the major concerns of enterprise computing community and did not make a noticeable impact for the IT industry.

Cloud computing is a technology largely viewed as the next big step in the development and deployment of an increasing number of distributed applications. The companies promoting

Cloud computing seem to have learned the most important lessons from the grid movement. Computer clouds are typically homogeneous. An entire cloud shares the same security, resource management, cost and other policies, and last, but not least, it targets enterprise computing. These are some of the reasons why several agencies of the US Government including the Health and Human Services, the Center for Disease Control (CDC), NASA etc.



Figure-1: Evolution of Computing

**Cloud Architecture**

The entire cloud architecture depends on the following models and entities,

**Delivery models**

. Software as a Service (SaaS)

. Platform as a Service (PaaS)

. Infrastructure as a Service (IaaS)

**Deployment models**

. Private cloud

. Community cloud

. Public cloud

. Hybrid cloud

**Resources**

Compute, storage servers, Network servers etc.

**Attributes**

Pay per usage, utility computing, Elasticity

Figure-2: Cloud architecture

The term “computer cloud” is overloaded as it covers infrastructures of different sizes, with different management, and a different user population. Several types of clouds are envisioned:

**Deployment Model**

*•* **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.

*•* **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.

*•* **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.

*•* **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).



Figure-3: Types of Cloud Deployment Model

**Delivery models**

According to the NIST reference model, the entities involved in cloud computing are,

***Service consumer***- Entity that maintains a business relationship with, and uses service from, service providers;

***Service provider***- Entity responsible for making a service available to service consumers; ***carrier* -** The intermediary that provides connectivity and transport of cloud services between providers and consumers;

***Broker* -** An entity that manages the use, performance and delivery of cloud services, and negotiates relationships between providers and consumers; *auditor* - a party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation. An *audit* is a systematic evaluation of a cloud system by measuring how well it conforms to a set of established criteria. For example, a security audit evaluates cloud security, a privacy-impact audit evaluates the cloud privacy assurance, while a performance audit evaluates the cloud performance.

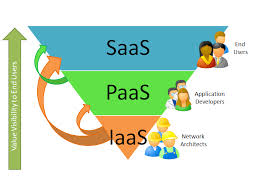


Figure-4: Types of Cloud Delivery Model

**Benefits of cloud computing**

* ***No up-front investment*:** Cloud computing uses a pay-as you- go pricing model. A service provider does not need to invest in the infrastructure to start gaining benefit from cloud computing. It simply rents resources from the cloud according to its own needs and pay for the usage.
* ***Lowering operating cost*:** Resources in a cloud environment can be rapidly allocated and de-allocated on demand. Hence, a service provider no longer needs to provision capacities according to the peak load. This provides huge savings since resources can be released to save on operating costs when service demand is low.
* ***Highly scalable*:** Infrastructure providers pool large amount of resources from data centers and make them easily accessible. A service provider can easily expand its service to large scales in order to handle rapid increase in service demands
* ***Easy access*:** Services hosted in the cloud are generally web-based. Therefore, they are easily accessible through a variety of devices with Internet connections. These devices not only include desktop and laptop computers, but also cell phones and PDAs. For example, in 2011 Apple announced the ***iCloud,***a network-centric alternative for content such as music, videos, movies, and personal information. Content previously confined to personal devices such as workstations, laptops, tablets, and smart phones need no longer be stored locally, can be shared by all these devices, and it is accessible whenever a device is connected to the Internet.
* ***Reducing business risks and maintenance expenses*:** By outsourcing the service infrastructure to the clouds, a service provider shifts its business risks (such as hardware failures) to infrastructure providers, who often have better expertise and are better equipped for managing these risks. In addition, a service provider can cut down the hardware maintenance and the staff training costs.

**Ethical issues and some challenges of cloud computing**



Figure-5: Partition of motivating factors for the adoption of cloud computing Cost saving (54%), Scalability (49%), efficiency (39%), Storage (11%), streamlined Administration (36%) (Approximate values)

* ***Security***

Security issue has played the most important role in hindering Cloud computing acceptance. Various security issues, possible in cloud computing are: availability, integrity, confidentiality, data access, data segregation, privacy, recovery, accountability, multi-tenancy issues and so on. Solution to various cloud security issues vary through cryptography, particularly public key infrastructure (PKI),use of multiple cloud providers, standardization of APIs, improving virtual machines support and legal support.

* ***Difficult to migrate***

It’s not very easy to move the applications from an enterprise to cloud computing environment or even within different cloud computing platforms because different cloud providers support different application architectures which are also dissimilar from enterprise application architectures.

* ***Internet dependency – performance and availability***

A cloud computing service relies fully on the availability, speed, quality and performance of internet as it works as carrier in between consumer and service provider.

* **Infrastructure failure**

Clouds are affected by malicious attacks and failures of the infrastructure, e.g., power failures or storage structure failure (Data Replication). Data centers are very large consumers of electric energy to keep the servers and the networking infrastructure running and for cooling.

* **Interoperability and Standardization**

The last major challenge we want to address is the 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.

* **Data transfer bottlenecks**

Many applications are data-intensive. A very important strategy is to store the data as close to the site where it is needed as possible. Transferring 1 TB of data on a 1 Mbps network takes 8 000 000 seconds or about 10 days; it is faster and cheaper to use courier service and send data recoded on some media than to send it over the network. Very high speed networks will alleviate this problem in the future, e.g., a 1 Gbps network would reduce this time to 8*,* 000 seconds, or slightly more than 2 hours.

* **Availability of service**

Availability of service is totally depends on the nature and capability of the cloud service provider.

**Developers**

Cloud computing is a path to utility computing embraced by major IT companies such as Amazon, Apple, Google, HP, IBM, Microsoft, Oracle, and others.



Figure-6: Types of Cloud Developers

**Conclusion and Future Scope**

Cloud computing is a technical and social reality and, at the same time, it is an emerging technology. “Cloud” computing builds on decades of research in virtualization, distributed computing, utility computing, and, more recently, networking, web and software services. It implies a service-oriented architecture, reduced information technology overhead for the end-user, great flexibility, reduced total cost of ownership, on-demand services and many other things.

The complexity of such systems is unquestionable and raises questions such as: How can we manage such systems? Do we have to consider radically new ideas, such as self management and self-repair for future clouds consisting of millions of servers?