UNIT1

Introduction ,Cloud Computing at a Glance, The Vision of Cloud Computing, Defining a Cloud, A Closer Look, Cloud Computing Reference Model, Characteristics and Benefits, Challenges Ahead, Historical Developments, Distributed Systems, Virtualization, Web 2.0, Service-Oriented Computing, Utility-Oriented Computing, Building Cloud Computing Environments, Application Development, Infrastructure and System Development, Computing Platforms and Technologies. Virtualization : Introduction, Characteristics of Virtualized, Environments Taxonomy of Virtualization Techniques, Execution Virtualization, Other Types of Virtualization, Virtualization and Cloud Computing, Pros and Cons of Virtualization, Technology

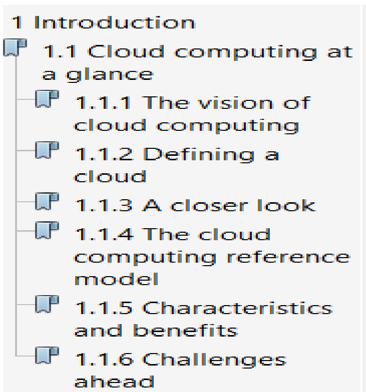
---

Introduction

Computing is being transformed into a model consisting of services that are commoditized and delivered in a manner similar to utilities such as water, electricity, gas, and telephony. In such a model, users access services based on their requirements, regardless of where the services are hosted.

Cloud computing is a technological advancement that focuses on the way we design computing systems, develop applications, and leverage existing services for building software. It is based on the concept of dynamic provisioning, which is applied not only to services but also to compute capability, storage, networking, and information technology (IT) infrastructure in general.

Cloud Computing at a Glance



In 1969, Leonard Kleinrock, one of the chief scientists of the original Advanced Research Projects Agency Network (ARPANET), which seeded the Internet, said:

"As of now, computer networks are still in their infancy, but as they grow up and become sophisti- cated, we will probably see the spread of 'computer utilities' which, like present electric and telephone utilities, will service individual homes and offices across the country."

Cloud computing allows renting infrastructure, runtime environments, and services on a pay-per- use basis. This principle finds several practical applications and then gives different images of cloud computing to different people. Chief information and technology officers of large enterprises see opportunities for scaling their infrastructure on demand and sizing it according to their business needs. End users leveraging cloud computing services can access their documents and data anytime,anywhere, and from any device connected to the Internet. Many other points of view exist.

One of the most diffuse views of cloud computing can be summarized as follows:

"I don't care where my servers are, who manages them, where my documents are stored, or where my applications are hosted. I just want them always available and access them from any device connected through Internet. And I am willing to pay for this service for as a long as I need it."

Vision of Cloud Computing:

Here vision of cloud computing means idea behind cloud computing.

1. Cloud computing provides the facility to provision virtual hardware, runtime environment and services to a person having money.
2. These all things can be used as long as they are needed by the user, there is no requirement for the upfront commitment.
3. The whole collection of computing system is transformed into a collection of utilities, which can be provisioned and composed together to deploy systems in hours rather than days, with no maintenance costs.
4. The long term vision of a cloud computing is that IT services are traded as utilities in an open market without technological and legal barriers.
5. In the near future we can imagine that it will be possible to find the solution that matches with our requirements by simply entering our request in a global digital market that trades with cloud computing services.
6. The existence of such market will enable the automation of the discovery process and its integration into its existing software systems.
7. Due to the existence of a global platform for trading cloud services will also help service providers to potentially increase their revenue.
8. A cloud provider can also become a consumer of a competitor service in order to fulfill its promises to customers.

Defining Cloud

The term cloud has historically been used in the telecommunications industry as an abstraction of the network in system diagrams. It then became the symbol of the most popular computer network, the Internet. This meaning also applies to cloud computing, which refers to an Internet-centric way of computing. The Internet plays a fundamental role in cloud computing, since it represents either the medium or the platform through which many cloud computing services are delivered and made accessible. This aspect is also reflected in the definition given by Armbrust et al.:

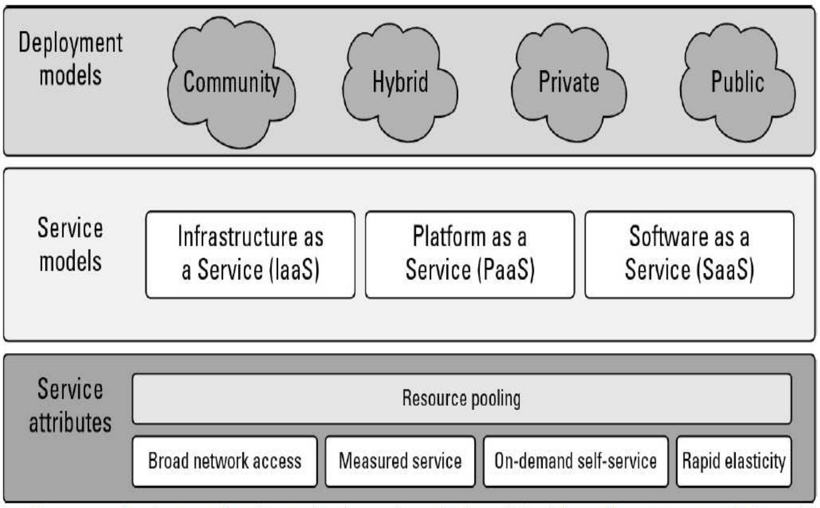
> *"Cloud computing refers to both the applications delivered as services over the Internet and the hardware and system software in the datacenters that provide those services."*

Forrester *defines cloud computing as:*

> *"A pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption."*

The notion of multiple parties using a shared cloud computing environment is highlighted in a definition proposed by the U.S. 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.*



Characteristics and benefits of Cloud Computing

Cloud computing has some interesting characteristics that bring benefits to both cloud service consumers (CSCs) and cloud service providers (CSPs). These characteristics are:

• **On-demand self-service**: A client can provision computer resources without the need for interaction with cloud service provider personnel.

***Broad network access**: Access to resources in the cloud is available over the network using standard methods in a manner that provides platform-independent access to clients of all types. This includes a mixture of heterogeneous operating systems, and thick and thin platforms such as laptops, mobile phones, and PDA.

• **Resource pooling**: A cloud service provider creates resources that are pooled together in a system that supports multi-tenant usage.

Physical and virtual systems are dynamically allocated or reallocated as needed. Intrinsic in this concept of pooling is the idea of abstraction that hides the location of resources such as virtual machines, processing, memory, storage, and network bandwidth and connectivity.

• **Rapid elasticity**: Resources can be rapidly and elastically provisioned.

The system can add resources by either scaling up systems (more powerful computers) or scaling out systems (more computers of the same kind), and scaling may be automatic or manual. From the standpoint of the client, cloud computing resources should look limitless and can be purchased at any time and in any quantity.

***Measured service**: The use of cloud system resources is measured, audited, and reported to the customer based on a metered system. A client can be charged based on a known metric such as amount of storage used, number of transactions, network I/O (Input/Output) or bandwidth, amount of processing power used, and so forth. A client is charged based on the level of services provided.

While these five core features of cloud computing are on almost anybody's list, you also should consider these additional advantages:

• **Lower costs:** Because cloud networks operate at higher efficiencies and with greater utilization, significant cost reductions are often encountered.

• **Ease of utilization**: Depending upon the type of service being offered, you may find that you do not require hardware or software licenses to implement your service.

• **Quality of Service**: The Quality of Service (QoS) is something that you can obtain under contract from your vendor.

• **Reliability:** The scale of cloud computing networks and their ability to provide load balancing and failover makes them highly reliable, often much more reliable than what you can achieve in a single organization.

• **Outsourced IT management**: A cloud computing deployment lets someone else manage your computing infrastructure while you manage your business. In most instances, you achieve considerable reductions in IT staffing costs.

• **Simplified maintenance and upgrade**: Because the system is centralized, you can easily apply patches and upgrades. This means your users always have access to the latest software versions.

• **Low Barrier to Entry**: In particular, upfront capital expenditures are dramatically reduced. In cloud computing, anyone can be a giant at any time.

Challenges Ahead

Despite its growing influence, concerns regarding cloud computing still remain. In our opinion, the benefits outweigh the drawbacks and the model is worth exploring. Some common challenges are:

1. Data Protection

Data Security is a crucial element that warrants scrutiny. Enterprises are reluctant to buy an assurance of business data security from vendors. They fear losing data to competition and the data confidentiality of consumers. In many instances, the actual storage location is not disclosed, adding onto the security concerns of enterprises. In the existing models, firewalls across data centers (owned by enterprises) protect this sensitive information. In the cloud model, Service providers are responsible for maintaining data security and enterprises would have to rely on them.

2. Data Recovery and Availability

All business applications have Service level agreements that are stringently followed. Operational teams play a key role in management of service level agreements and runtime governance of applications. In production environments, operational teams support Appropriate clustering and Fail over Data Replication ,System monitoring (Transactions monitoring, logs monitoring and others) ,Maintenance (Runtime Governance) ,Disaster recovery Capacity and performance management

If, any of the above mentioned services is under-served by a cloud provider, the damage & impact could be severe.

3. Management Capabilities

Despite there being multiple cloud providers, the management of platform and infrastructure is still in its infancy. Features like „Auto-scaling‟ for example, are a crucial requirement for many enterprises. There is huge potential to improve on the scalability and load balancing features provided today.

4. Regulatory and Compliance Restrictions

In some of the European countries, Government regulations do not allow customer's personal information and other sensitive information to be physically located outside the state or country. In order to meet such requirements, cloud providers need to setup a data center or a storage site exclusively within the country to comply with regulations. Having such an infrastructure may not always be feasible and is a big challenge for cloud providers.

5. Applications Not Ready

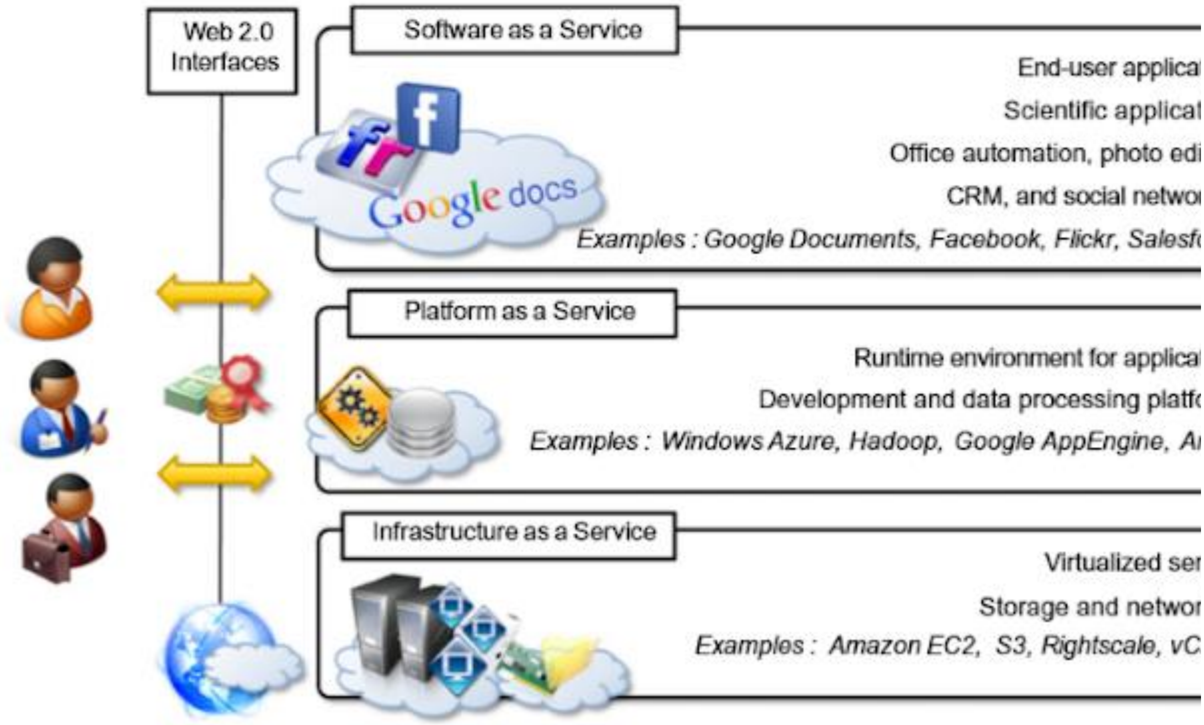
In some cases the applications themselves are not ready to be used on the cloud. They may have little quirks that prevent them from being used to their fullest abilities, or they may not work whatsoever.

First, the application might require a lot of bandwidth to communicate with users. Remember, since cloud computing is paid based on how much you use, it might turn out to be less expensive in the long run to simply house the application locally until it can be rewritten or otherwise modified to operate more efficiently.

Some applications may not be able to communicate securely across the Internet. If they cannot communicate securely or through a tunnel, then your data is at risk. In the event the application cannot communicate securely, you will need to host it locally where you can have other means of security to protect data as it is transported across networks.

Cloud Computing Reference Model:

The reference model for cloud computing is an abstract model that characterizes and standardizes a cloud computing environment by partitioning it into abstraction layers and cross-layer functions.



Cloud computing reference model

If we look in to the reference model as seen in above image we will find classification of Cloud Computing services:

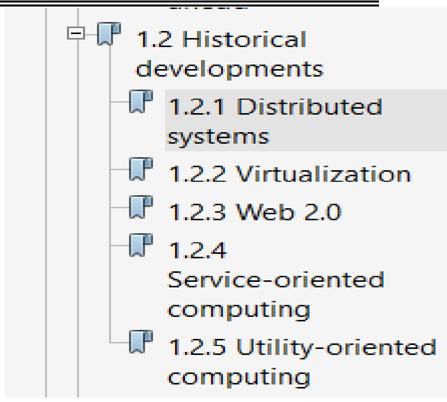
1. Infrastructure-as-a-Service (IaaS),
2. Platform-as-a-Service (PaaS), and
3. Software-as-a-Service (SaaS).
4. Web 2.0

1. Infrastructure as a service (IaaS) is a cloud computing offering in which a vendor provides users access to computing resources such as servers, storage and networking. To read more about IaaS click here.

2. Platform as a service (PaaS) is a cloud computing offering that provides users with a cloud environment in which they can develop, manage and deliver applications. To read more about PaaS click here.

3. Software as a service (SaaS) is a cloud computing offering that provides users with access to a vendor's cloud-based software. Users do not install applications on their local devices. Instead, the applications reside on a remote cloud network accessed through the web or an API. Through the application, users can store and analyze data and collaborate on projects. To read more about SaaS click here.

4. Web 2.0 is the term used to describe a variety of web sites and applications that allow anyone to create and share online information or material they have created. A key element of the technology is that it allows people to create, share, collaborate & communicate.



Historical Developments

The below Figure provides an overview of the evolution of the distributed computing technologies that have influenced cloud computing. In tracking the historical evolution, we briefly review five core technologies that played an important role in the realization of cloud computing.

These technologies are

- distributed systems,
- virtualization,
- Web 2.0,
- service orientation,
- utility computing

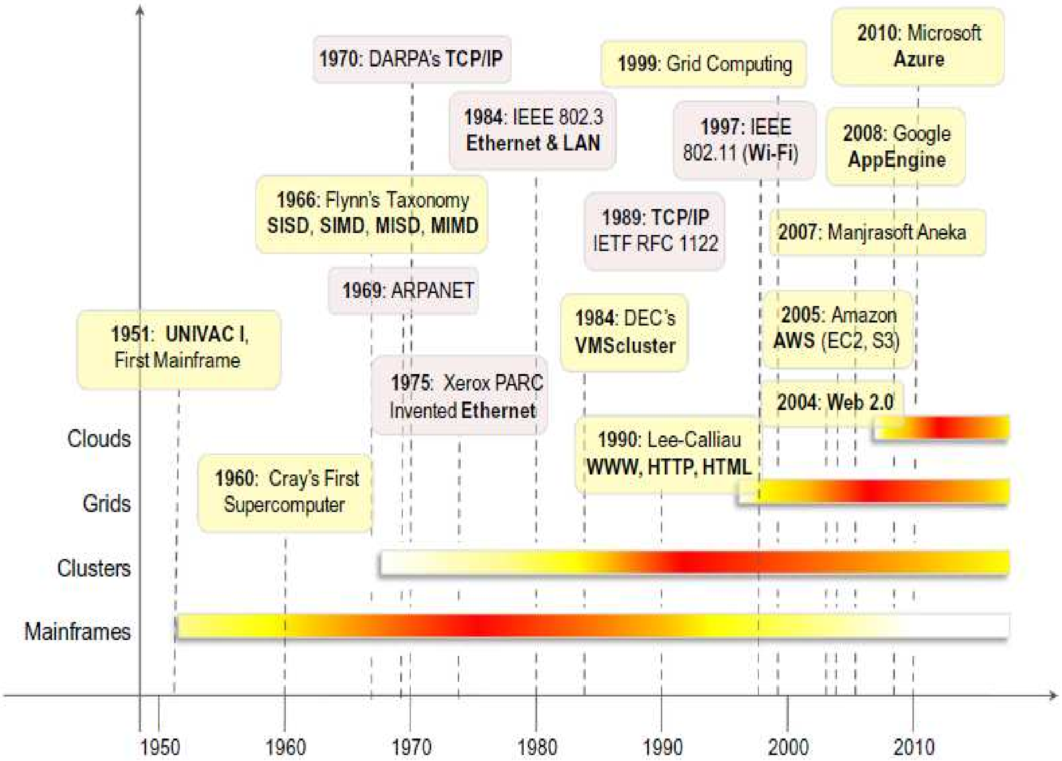


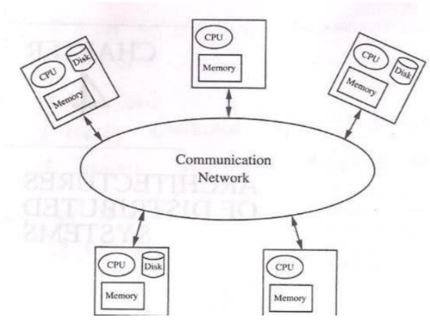
FIGURE 1.6

The evolution of distributed computing technologies, 1950s–2010s.

Distributed systems

- A distributed system is a collection of independent computers that appears to its users as a single coherent system.

A distributed system is a collection of computers that are separated and do not share a common memory or a clock. The processes executing on these computers communicates with one another by exchanging messages over communication channels. The messages are delivered after an arbitrary transmission delay.



The primary purpose of distributed systems is to share resources and utilize them better. This is also true in the case of cloud computing, where this concept is taken to the extreme and resources (infrastructure, runtime environments, and services) are rented to users. In fact, one of the driving factors of cloud computing has been the availability of the large computing facilities of IT giants (Amazon, Google) that found that offering their computing capabilities as a service provided opportunities to better utilize their infrastructure

Distributed systems often exhibit other properties such as heterogeneity, openness, scalability, transparency, concurrency, continuous availability, and independent failures. To some extent these also characterize clouds, especially in the context of scalability, concurrency , and continuous availability.

Three major milestones have led to cloud computing: mainframe computing, cluster computing,and grid computing

MAINFRAMES

- Mainframes were powerful, highly reliable computers specialized for large data movement and massive input/output (I/O) operations.

- They were mostly used by large organizations for bulk data processing tasks such as online transactions, enterprise resource planning, and other operations involving the processing of significant amounts of data.
- One of the most attractive features of mainframes was the ability to be highly reliable computers that were “always on” and capable of tolerating failures transparently.

CLUSTERS.

Cluster computing started as a low-cost alternative to the use of mainframes and supercomputers.

Cluster computing refers to the process of sharing the computation task to multiple computers of the cluster. The number of computers are connected on a network and they perform a single task by forming a Cluster of computers where the process of computing is called as cluster computing.

Benefits of Cluster Computing

Cluster computing offers a wide array of benefits. Some of these include the following.

Cost-Effectiveness – Compared with the mainframe systems, cluster computing is considered to be much more cost-effective. These computing systems offer enhanced performance with respect to the mainframe computer devices.
Processing Speed – The processing speed of cluster computing is justified with that of the mainframe systems and other supercomputers present in the world.
Expandability – Scalability and expandability are another set of advantages that cluster computing offers. Cluster computing represents an opportunity for adding any number of additional resources and systems to the existing computing network.
Increased Resource Availability – Availability plays a vital role in cluster computing systems. Failure of any connected active node can be easily passed on to other active nodes on the server, ensuring high availability.
Improved Flexibility – In cluster computing, superior specifications can be upgraded and extended by adding newer nodes to the existing server.

GRIDS

What is a Grid?"
In 1998, Ian Foster and Carl Kesselman provided an initial definition in their book *The Grid: Blueprint for a New Computing Infrastructure* : "A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities."
Grid computing is a collection of computing resources that appear to the end user as one large system and provide a single point of access for performing tasks. Grid computing allows for hundreds, if not thousands, of services and their requests to run concurrently, without being concerned about where they execute.
Grid computing gives enterprises the equivalent of a supercomputer without adding one to their data centers. The ability to use computational power not only within your organization but also from third parties further increases availability.
In an analogy to the power grid, grid computing proposed a new approach to access large computational power, huge storage facilities, and a variety of services. Users can "consume" resources in the same way as they use other utilities such as power, gas, and water. Grids initially developed as aggregations of geographically dispersed clusters by means of Internet connections. These clusters belonged to different organizations, and arrangements were made among them to share the computational power. Different from a "large cluster," a computing grid was a dynamic aggregation of heterogeneous computing nodes, and its scale was nation wide or even worldwide.

Cloud computing is often considered the successor of grid computing. In reality, it represents aspects of all these three major technologies. Computing clouds are deployed in large data centers hosted by a single organization that provides services to others. Clouds are characterized by thefact of having virtually infinite capacity, being tolerant to failures, and being always on, as in the case of mainframes.

Virtualization

Virtualization is the process of converting a physical IT resource into a virtual IT resource. Most types of IT resources can be virtualized, including:

- *Servers* – A physical server can be abstracted into a virtual server.

- *Storage* – A physical storage device can be abstracted into a virtual storage device or a virtual disk.
- *Network* – Physical routers and switches can be abstracted into logical network fabrics, such as VLANs.
- *Power* – A physical UPS and power distribution units can be abstracted into what are commonly referred to as virtual UPSs.

Virtualization has been around for more than40 years, but its application has always been limited by technologies that did not allow an efficient use of virtualization solutions. Today these limitations have been substantially overcome, and virtualization has become a fundamental element of cloud computing. This is particularly true for solutions that provide IT infrastructure on demand. Virtualization confers that degree of customization and control that makes cloud computing appealing for users and, at the same time, sustainable for cloud services providers.

Virtualization is the base technology that enables cloud computing solutions to deliver virtual servers on demand, such as Amazon EC2, RightScale,VMware vCloud, and others. Together with hardware virtualization, storage and network virtualization complete the range of technologies for the emulation of IT infrastructure.

Web 2.0

Web 2.0 is a term that was introduced in 2004 and refers to the second generation of the World Wide Web.

Web 2.0 technologies provide a level user interaction that was not available before. Websites have become much more dynamic and interconnected, producing "online communities" and making it even easier to share information on the Web. Because most Web 2.0 features are offered as free services, sites like Wikipedia and Facebook have grown at amazingly fast rates. As the sites continue to grow, more features are added, building off the technologies in place.

Some examples of features considered to be part of Web 2.0 are listed below:

- **Blogs** - also known as Web logs, these allow users to post thoughts and updates about their life on the Web.
- **Wikis** - sites like Wikipedia and others enable users from around the world to add and update online content.
- **Social networking** - sites like Facebook and MySpace allow users to build and customize their own profiles and communicate with friends.
- **Web applications** - a broad range of new applications make it possible for users to run programs directly in a Web browser.

Examples of Web 2.0 applications are Google Documents, Google Maps, Flickr, Facebook, Twitter, YouTube, de.li.cious, Blogger, and Wikipedia. In particular, social networking Websites take the biggest advantage of Web 2.0.

Service-oriented computing

Service orientation is the core reference model for cloud computing systems. This approach adopts the concept of services as the main building blocks of application and system development. Service-oriented computing (SOC)supports the development of rapid, low-cost, flexible, interoperable, and evolvable applications and systems

A service is an abstraction representing a self-describing and platform-agnostic component that can perform any function—anything from a simple function to a complex business process. Virtually any piece of code that performs a task can be turned into a service and expose its functionalities through a network-accessible protocol.

Services are composed and aggregated into a service-oriented architecture (SOA), which is a logical way of organizing software systems to provide end users or other entities distributed over the network with services through published and discoverable interfaces.

Utility-oriented computing

Utility computing is a vision of computing that defines a service-provisioning model for compute services in which resources such as storage, compute power, applications, and infrastructure are packaged and offered on a pay-per-use basis. The idea of providing computing as autility like natural gas, water, power, and telephone connection has a long history but has become a reality today with the advent of cloud computing

Properties of Utility Computing

The important properties of utility computing are its scalability, demand pricing, standardized utility computing services, utility computing on virtualization and automation.

- Scalability is an important metric that should be ensured in utility computing to provide sufficient IT resources available at any time. If the demand gets extended, the response time and quality should not get impacted.
- Demand pricing is scheduled effectively to pay for both hardware and software components as per the usage.
- The catalog is produced with standardized services with different service level agreements to the customers. So the consumer has no influence on the behind technology on the computer platform.
- The web services and other resources are shared by the pool of machines which is used in automation and virtualization technologies. It segregates the network into many logical resources instead of the available physical resources. An application is allotted with no specific predefined servers or storage space of any severs with more memory or free server runtime from the resource pool

Cloud computing and grid computing are based on the concept of utility computing.

Building cloud computing environments

The creation of cloud computing environments encompasses both the development of applications and systems that leverage cloud computing solutions and the creation of frameworks, platforms, and infrastructures delivering cloud computing services.

Application development

Applications that leverage cloud computing benefit from its capability to dynamically scale on demand. One class of applications that takes the biggest advantage of this

feature is that of Web applications. Their performance is mostly influenced by the workload generated by varying user demands. With the diffusion of Web 2.0 technologies, the Web has become a platform for developing rich and complex applications, including enterprise applications that now leverage the Internet as the preferred channel for service delivery and user interaction.
Another class of applications that can potentially gain considerable advantage by leveraging cloud computing is represented by resource-intensive applications. These can be either data-intensive or compute-intensive applications. In both cases, considerable amounts of resources are required to complete execution in a reasonable timeframe.
Cloud computing provides a solution for on-demand and dynamic scaling across the entire stack of computing.

This is achieved by
(a) providing methods for renting compute power, storage, and networking;
(b) offering runtime environments designed for scalability and dynamic sizing; and
(c) providing application services that mimic the behavior of desktop applications but that are completely hosted and managed on the provider side.

Infrastructure and system development

Distributed computing, virtualization, service orientation, and Web 2.0 form the core technologies enabling the provisioning of cloud services from anywhere on the globe. Developing applications and systems that leverage the cloud requires knowledge across all these technologies.
Distributed computing is a foundational model for cloud computing because cloud systems are distributed systems. Besides administrative tasks mostly connected to the accessibility of resources in the cloud, the extreme dynamism of cloud systems—where new nodes and services are provisioned on demand—constitutes the major challenge for engineers and developers.
Web 2.0 technologies constitute the interface through which cloud computing services are delivered, managed, and provisioned.
Cloud computing is often summarized with the acronym XaaS—Everything-as-a-Service—that clearly underlines the central role of service orientation.
Virtualization is another element that plays a fundamental role in cloud computing. This technology is a core feature of the infrastructure used by cloud providers.

Computing platforms and technologies

Development of a cloud computing application happens by leveraging platforms and frameworks that provide different types of services, from the bare-metal infrastructure to customizable applications serving specific purposes.
1. Amazon web services (AWS)
2 Google AppEngine
3 Microsoft Azure
4 Hadoop
5 Force.com and Salesforce.com
6 Manjrasoft Aneka

1 Amazon web services (AWS)
AWS offers comprehensive cloud IaaS services ranging from virtual compute, storage, and networking to complete computing stacks. AWS is mostly known for its compute

and storage-on- demand services, namely Elastic Compute Cloud (EC2) and Simple Storage Service (S3). EC2 provides users with customizable virtual hardware that can be used as the base infrastructure for deploying computing systems on the cloud. It is possible to choose from a large variety of virtual hardware configurations, including GPU and cluster instances. EC2 also provides the capability to save a specific running instance as an image, thus allowing users to create their own templates for deploying systems. These templates are stored into S3 that delivers persistent storage on demand. S3 is organized into buckets; these are containers of objects that are stored in binary form and can be enriched with attributes. Users can store objects of any size, from simple files to entire disk images, and have them accessible from everywhere.

2 Google AppEngine

Google AppEngine is a scalable runtime environment mostly devoted to executing Web applications. These take advantage of the large computing infrastructure of Google to dynamically scale as the demand varies over time. AppEngine provides both a secure execution environment and a collection of services that simplify the development of scalable and high-performance Web applications. These services include in-memory caching, scalable data store, job queues, messaging, and cron tasks.
Developers can build and test applications on their own machines using the AppEngine software development kit (SDK). Once development is complete, developers can easily migrate their application to AppEngine, set quotas to contain the costs generated, and make the application available to the world. The languages currently supported are Python, Java, and Go.

3 Microsoft Azure

Microsoft Azure is a cloud operating system and a platform for developing applications in the cloud. Applications in Azure are organized around the concept of roles, which identify a distribution unit for applications and embody the application's logic. Currently, there are three types ofrole: Web role, worker role, and virtual machine role. The Web role is designed to host a Web application, the
worker role is a more generic container of applications and can be used to perform workload processing, and the virtual machine role provides a virtual environment in which the computing stack can be fully customized, including the operating systems.

4 Hadoop

Apache Hadoop is an open-source framework that is suited for processing large data sets on commodity hardware. Hadoop is an implementation of MapReduce, an application programming model developed by Google, which provides two fundamental operations for data processing: map and reduce. The former transforms and synthesizes the input data provided by the user; the latter aggragates the output obtained by the map operations. Hadoop provides the runtime environment, and developers need only provide the input data and specify the map and reduce functions that need to be executed.

5 Force.com and Salesforce.com

Force.com is a cloud computing platform for developing social enterprise applications. The platform is the basis for SalesForce.com, a Software-as-a-Service solution for customer relationship management. Force.com allows developers to create applications by composing ready-to-use blocks; a complete set of components supporting all the activities of an enterprise are available. The platform provides complete support for developing applications, from the design of the data layout to the definition of business rules and workflows and the definition of the user interface.

6 Manjrasoft Aneka

Manjrasoft Aneka is a cloud application platform for rapid creation of scalable applications and their deployment on various types of clouds in a seamless and elastic manner. It supports a collection of programming abstractions for developing applications and a distributed runtime environment that can be deployed on heterogeneous hardware (clusters, networked desktop computers, and cloud resources).

Developers can choose different abstractions to design their application: tasks, distributed threads, and map-reduce. These applications are then executed on the distributed service-oriented runtime environment, which can dynamically integrate additional resource on demand.

************** end of part 1 **********************************

Part 2 is virtualization

****