

1. Define cloud with example?

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.

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.

- 1) Large enterprises can offload some of their activities to cloud-based systems
- 2) Small enterprises and start-ups can afford to translate their ideas into business results more quickly, without excessive up-front costs.
- 3) System developers can concentrate on the business logic rather than dealing with the complexity of infrastructure management and scalability
- 4) End users can have their documents accessible from everywhere and any device

2. Which are the different cloud computing services

Infrastructure-as-a-Service (IaaS)- solutions deliver infrastructure on demand in the form of virtual hardware, storage, and networking,

Platform-as-a-Service (PaaS)- They deliver scalable and elastic runtime environments on demand and host the execution of applications. These services are backed by a core middleware platform that is responsible for creating the abstract environment where applications are deployed and executed

Software-as-a-Service (SaaS)- solutions provide applications and services on demand. Most of the common functionalities of desktop applications—such as office automation, document management, photo editing, and customer relationship management (CRM) software—are replicated on the provider's infrastructure and made more scalable and accessible through a browser on demand. These applications are shared across multiple users whose interaction is isolated from the other users.

3. Characteristics and benefits of cloud computing

- No up-front commitments
- On-demand access
- Nice pricing
- Simplified application acceleration and scalability
- Efficient resource allocation
- Energy efficiency
- Seamless creation and use of third-party services

4. Define Distributed systems

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

5. features of Distributed system

Distributed systems often exhibit other properties such as heterogeneity, openness, scalability, transparency, concurrency, continuous availability, and independent failures.

6. Give example for web 2.0

Examples of Web 2.0 applications are Google Documents, Google Maps, Flickr, Facebook, Twitter, YouTube, de.li.cious, Blogger, and Wikipedia.

7.define 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.

8. differentiate between parallel and distributed system

the term parallel computing refers to a model in which the computation is divided among several processors sharing the same memory. The term distributed computing encompasses any architecture or system that allows the computation to be broken down into units and executed concurrently on different computing elements, whether these are processors on different nodes, processors on the same computer, or cores within the same processor.

9. what is parallel system

Processing of multiple tasks simultaneously on multiple processors is called parallel processing. The parallel program consists of multiple active processes (tasks) simultaneously solving a given problem

10.define SISD

An SISD computing system is a uniprocessor machine capable of executing a single instruction, which operates on a single data stream

11. define SIMD

An SIMD computing system is a multiprocessor machine capable of executing the same instruction on all the CPUs but operating on different data streams

12.define MISD

An MISD computing system is a multiprocessor machine capable of executing different instructions on different PEs but all of them operating on the same data set

13. define MIMD

An MIMD computing system is a multiprocessor machine capable of executing multiple instructions on multiple data sets

14. define peer- to- peer model

The peer-to-peer model introduces a symmetric architecture in which all the components, called peers, play the same role and incorporate both client and server capabilities of the client/server model.

15. Define virtualization

Virtualization is a large umbrella of technologies and concepts that are meant to provide an abstract environment—whether virtual hardware or an operating system—to run applications.

In a virtualized environment there are three major components: guest, host, and virtualization layer. The guest represents the system component that interacts with the virtualization layer rather than with the host, as would normally happen. The host represents the original environment where the guest is supposed to be managed. The virtualization layer is responsible for recreating the same or a different environment where the guest will operate.

16. Advantages of virtualization

* **Increased security:** The ability to control the execution of a guest in a completely transparent manner opens new possibilities for delivering a secure, controlled execution environment.

* **Managed execution:** provides some other features such as sharing, aggregation, emulation, and isolation are the most relevant features.

Sharing. Virtualization allows the creation of a separate computing environments within the same host

Aggregation : A group of separate hosts can be tied together and represented to guests as a single virtual host.

Emulation: Guest programs are executed within an environment that is controlled by the virtualization layer, which ultimately is a program.

Isolation: Virtualization allows providing guests—whether they are operating systems, applications, or other entities—with a completely separate environment, in which they are executed.

* **Portability:** The concept of portability applies in different ways according to the specific type of virtualization considered. In the case of a hardware virtualization solution, the guest is packaged into a virtual image that, in most cases, can be safely moved and executed on top of different virtual machines.

17. disadvantages of virtualization

- Performance degradation

Since virtualization interposes an abstraction layer between the guest and the host, the guest can experience increased latencies.

- Inefficiency and degraded user experience
Virtualization can sometime lead to an inefficient use of the host.
- Security holes and new threats
Virtualization opens the door to a new and unexpected form of phishing. In the case of hardware virtualization, malicious programs can preload themselves before the operating system and act as a thin virtual machine manager toward it.

18. give example for virtualization technologies

- **Xen** - Xen is an open-source initiative implementing a virtualization platform based on paravirtualization. Xen-based technology is used for either desktop virtualization or server virtualization, and recently it has also been used to provide cloud computing solutions by means of Xen Cloud Platform (XCP).

Xen is the most popular implementation of paravirtualization, which, in contrast with full virtualization, allows high-performance execution of guest operating systems.

- **VMware**

VMware's technology is based on the concept of full virtualization, where the underlying hardware is replicated and made available to the guest operating system, which runs unaware of such abstraction layers and does not need to be modified. VMware implements desktop environment, by means of Type II hypervisors, or in the server environment, by means of Type I hypervisors.

- **Microsoft Hyper-V**

Hyper-V is an infrastructure virtualization solution developed by Microsoft for server virtualization. As the name recalls, it uses a hypervisor-based approach to hardware virtualization, which leverages several techniques to support a variety of guest operating systems. Hyper-V is currently shipped as a component of Windows Server 2008 R2 that installs the hypervisor as a role within the server

19. define Hypervisor

The hypervisor is the component that directly manages the underlying hardware (processors and memory).

20. Cloud Computing Services Classification

Category	Characteristics	Product Type	Vendors and Products
SaaS	Customers are provided with applications that are accessible anytime and from anywhere.	Web applications and services (Web 2.0)	SalesForce.com (CRM) Clarizen.com (project management) Google Apps
PaaS	Customers are provided with a platform for developing applications hosted in the cloud.	Programming APIs and frameworks Deployment systems	Google AppEngine Microsoft Azure Manjrasoft Aneka Data Synapse
IaaS/HaaS	Customers are provided with virtualized hardware and storage on top of which they can build their infrastructure.	Virtual machine management infrastructure Storage management Network management	Amazon EC2 and S3 GoGrid Nirvanix

21. different types of cloud

- **Public clouds.** The cloud is open to the wider public.
- **Private clouds.** The cloud is implemented within the private premises of an institution and generally made accessible to the members of the institution or a subset of them.
- **Hybrid or heterogeneous clouds.** The cloud is a combination of the two previous solutions and most likely identifies a private cloud that has been augmented with resources or services hosted in a public cloud.
- **Community clouds.** The cloud is characterized by a multi-administrative domain involving different deployment models (public, private, and hybrid), and it is specifically designed to address the needs of a specific industry.

22. application of community cloud

- Media industry. In the media industry, companies are looking for low-cost, agile, and simple solutions to improve the efficiency of content production.
- Healthcare industry- community clouds can provide a global platform on which to share information and knowledge without revealing sensitive data maintained within the private infrastructure.
- Energy and other core industries. In these sectors, community clouds can bundle the comprehensive set of solutions that together vertically address management, deployment, and orchestration of services and operations.
- Public sector. Legal and political restrictions in the public sector can limit the adoption of public cloud offerings
- Scientific research. Science clouds are an interesting example of community clouds. In this case, the common interest driving different organizations sharing a large distributed infrastructure is scientific computing

23. benefits of community cloud

- Openness. By removing the dependency on cloud vendors, community clouds are open systems in which fair competition between different solutions can happen.
- Community. Being based on a collective that provides resources and services, the infrastructure turns out to be more scalable because the system can grow simply by expanding its user base.
- Graceful failures. Since there is no single provider or vendor in control of the infrastructure, there is no single point of failure.
- Convenience and control. Within a community cloud there is no conflict between convenience and control because the cloud is shared and owned by the community, which makes all the decisions through a collective democratic process.
- Environmental sustainability. The community cloud is supposed to have a smaller carbon footprint because it harnesses underutilized resources. Moreover, these clouds tend to be more organic by growing and shrinking in a symbiotic relationship to support the demand of the community, which in turn sustains it.

24. Define Aneka

Aneka is a software platform for developing cloud computing applications. It allows harnessing of disparate computing resources and managing them into a unique virtual domain—the Aneka Cloud—in which applications are executed.

Aneka is a pure PaaS solution for cloud computing. Aneka is a cloud middleware product that can be deployed on a heterogeneous set of resources: a network of computers, a multicore server, datacenters, virtual cloud infrastructures, or a mixture of these. The framework provides both middleware for managing and scaling distributed applications and an extensible set of APIs for developing them.

25. services of aneka

- Elasticity and scaling. By means of the dynamic provisioning service, Aneka supports dynamically upsizing and downsizing of the infrastructure available for applications.
- Runtime

management. The runtime machinery is responsible for keeping the infrastructure up and running and serves as a hosting environment for services. It is primarily represented by the container and a collection of services that manage service membership and lookup, infrastructure maintenance, and profiling.

- Resource management. Aneka is an elastic infrastructure in which resources are added and removed dynamically according to application needs and user requirements. To provide QoS-based execution, the system not only allows dynamic provisioning but also provides capabilities for reserving nodes for exclusive use by specific applications.
- Application management. A specific subset of services is devoted to managing applications. These services include scheduling, execution, monitoring, and storage management.
- User management. Aneka is a multitenant distributed environment in which multiple applications, potentially belonging to different users, are executed. The framework provides an extensible user system via which it is possible to define users, groups, and permissions. The services devoted to user management build up the security infrastructure of the system and constitute a fundamental element for accounting management.
- QoS/SLA management and billing. Within a cloud environment, application execution is metered and billed. Aneka provides a collection of services that coordinate together to take into account the usage of resources by each application and to bill the owning user accordingly.

26. the services of Aneka container

The services installed in the Aneka container can be classified into three major categories:

- Fabric Services

Fabric Services define the lowest level of the software stack representing the Aneka Container. They provide access to the resource-provisioning subsystem and to the monitoring facilities implemented in Aneka. Fabric Services are fundamental services of the Aneka Cloud and define the basic infrastructure management features of the system.

- Foundation Services

Foundation Services are related to the logical management of the distributed system built on top of the infrastructure and provide supporting services for the execution of distributed applications.

- Application Service

Application Services manage the execution of applications and constitute a layer that differentiates according to the specific programming model used for developing distributed applications on top of Aneka.

27. what is heart beat service

Profiling and monitoring services are mostly exposed through the Heartbeat, Monitoring, and Reporting Services. The Heartbeat Service periodically collects the dynamic performance information about the node and publishes this information to the membership service in the Aneka Cloud. These data are collected by the index node of the Cloud, which makes them available for services such as reservations and scheduling in order to optimize the use of a heterogeneous infrastructure.

28. what is data intensive computing

Data-intensive computing focuses on a class of applications that deal with a large amount of data.

Data-intensive computing is concerned with production, manipulation, and analysis of large-scale data in the range of hundreds of megabytes (MB) to petabytes (PB) and beyond.

29. example for High-performance distributed file systems and storage clouds

Lustre. The Lustre file system is a massively parallel distributed file system that covers the needs of a small workgroup of clusters to a large-scale computing cluster.

IBM General Parallel File System (GPFS). GPFS is the high-performance distributed file system developed by IBM that provides support for the RS/6000 supercomputer and Linux computing clusters.

Google File System (GFS). GFS is the storage infrastructure that supports the execution of distributed applications in Google's computing cloud. The system has been designed to be a fault-tolerant, highly available, distributed file system built on commodity hardware and standard Linux operating systems.

Sector. Sector is the storage cloud that supports the execution of data-intensive applications defined according to the Sphere framework. It is a user space file system that can be deployed on commodity hardware across a wide-area network.

Amazon Simple Storage Service (S3). Amazon S3 is the online storage service provided by Amazon.

30. define MapReduce

MapReduce is a programming platform Google introduced for processing large quantities of data. It expresses the computational logic of an application in two simple functions: map and reduce. Data transfer and management are completely handled by the distributed storage infrastructure (i.e., the Google File System), which is in charge of providing access to data, replicating files, and eventually moving them where needed.

31. Alternatives to MapReduce

sphere. Sphere is the distributed processing engine that leverages the Sector Distributed File System (SDFS). Rather than being a variation of MapReduce, Sphere implements the stream processing model (Single Program, Multiple Data) and allows developers to express the computation in terms of user-defined functions (UDFs), which are run against the distributed infrastructure.

All-Pairs. All-Pairs [100] is an abstraction and a runtime environment for the optimized execution of data-intensive workloads. It provides a simple abstraction—in terms of the All-pairs function—that is common in many scientific computing domains:

All-pairs (A: set; B: set; F: function) → M: matrix

DryadLINQ. Dryad is a Microsoft Research project that investigates programming models for writing parallel and distributed programs to scale from a small cluster to a large datacenter.

32.give example for major cloud computing technologies(essay)

Amazon web services

Amazon Web Services (AWS) is a platform that allows the development of flexible applications by providing solutions for elastic infrastructure scalability, messaging, and data storage. The platform is accessible through SOAP or RESTful Web service interfaces and provides a Web-based console where users can handle administration and monitoring of the resources required, as well as their expenses computed on a pay-as-you-go basis

Google AppEngine

Google AppEngine is a PaaS implementation that provides services for developing and hosting scalable Web applications. AppEngine is essentially a distributed and scalable runtime environment that leverages Google's distributed infrastructure to scale out applications facing a large number of requests by allocating more computing resources to them and balancing the load among them. The runtime is completed by a collection of services that allow developers to design and implement applications that naturally scale on AppEngine.

Microsoft Windows Azure

Microsoft Windows Azure is a cloud operating system built on top of Microsoft datacenters' infrastructure and provides developers with a collection of services for building applications with cloud technology. Services range from compute, storage, and networking to application connectivity, access control, and business intelligence. Any application that is built on the Microsoft technology can be scaled using the Azure platform, which integrates the scalability features into the common Microsoft technologies such as Microsoft Windows Server 2008, SQL Server, and ASP.NET.

33. differentiate between web role, worker role and vm role

Compute services are the core components of Microsoft Windows Azure, and they are delivered by means of the abstraction of **roles**. A **role** is a runtime environment that is customized for a specific compute task. Roles are managed by the Azure operating system and instantiated on demand in order to address surges in application demand. Currently, there are three different roles: **Web role, Worker role, and Virtual Machine (VM) role**.

Web role

The Web role is designed to implement scalable Web applications. Web roles represent the units of deployment of Web applications within the Azure infrastructure. They are hosted on the IIS 7 Web Server, which is a component of the infrastructure that supports Azure. When Azure detects peak loads in the request made to a given application, it instantiates multiple Web roles for that application and distributes the load among them by means of a load balancer.

Worker role

Worker roles are designed to host general compute services on Azure. They can be used to quickly provide compute power or to host services that do not communicate with the external world through HTTP. A common practice for Worker roles is to use them to provide background processing for Web applications developed with Web roles.

Virtual machine role

The Virtual Machine role allows developers to fully control the computing stack of their compute service by defining a custom image of the Windows Server 2008 R2 operating system and all the service stack required by their applications. The Virtual Machine role is based on the Windows Hyper-V virtualization technology which is natively integrated in the Windows server technology at the base of Azure.

34. define Blobs

Azure allows storing large amount of data in the form of binary large objects (BLOBs) by means of the blobs service. This service is optimal to store large text or binary files. Two types of blobs are available:

- **Block blobs.** Block blobs are composed of blocks and are optimized for sequential access; therefore they are appropriate for media streaming. Currently, blocks are of 4 MB, and a single block blob can reach 200 GB in dimension.
- **Page blobs.** Page blobs are made of pages that are identified by an offset from the beginning of the blob. A page blob can be split into multiple pages or constituted of a single page. This type of blob is optimized for random access and can be used to host data different from streaming. Currently, the maximum dimension of a page blob can be 1 TB. Blobs storage provides users with the ability to describe the data by adding metadata. It is also possible to take snapshots of a blob for backup purposes. Moreover, to optimize its distribution, blobs storage can leverage the Windows Azure CDN so that blobs are kept close to users requesting them and can be served efficiently.

35.cloud computing applications

1. Scientific applications

Scientific applications are a sector that is increasingly using cloud computing systems and technologies. The immediate benefit seen by researchers and academics is the potentially infinite availability of computing resources and storage at sustainable prices compared to a complete in-house deployment.

Healthcare:

ECG analysis in the cloud Healthcare is a domain in which computer technology has found several and diverse applications: from supporting the business functions to assisting scientists in developing solutions to cure diseases.

Biology:

1. **protein structure prediction Applications** in biology often require high computing capabilities and often operate on large datasets that cause extensive I/O operations. Because of these requirements, biology applications have often made extensive use of supercomputing and cluster computing infrastructures. Similar capabilities can be leveraged on demand using cloud computing technologies in a more dynamic fashion, thus opening new opportunities for bioinformatics applications.
2. **gene expression data analysis for cancer diagnosis Gene expression profiling** is the measurement of the expression levels of thousands of genes at once. It is used to understand the biological processes that are triggered by medical treatment at a cellular

level. Together with protein structure prediction, this activity is a fundamental component of drug design, since it allows scientists to identify the effects of a specific treatment.

Geoscience:

satellite image processing Geoscience applications collect, produce, and analyse massive amounts of geospatial and nonspatial data. As the technology progresses and our planet becomes more instrumented (i.e., through the deployment of sensors and satellites for monitoring), the volume of data that needs to be processed increases significantly.

2. Business and consumer applications

CRM and ERP Customer relationship management (CRM) and enterprise resource planning (ERP) applications are market segments that are flourishing in the cloud, with CRM applications the more mature of the two. Cloud CRM applications constitute a great opportunity for small enterprises and start-ups to have fully functional CRM software without large up-front costs and by paying subscriptions.

Google docs

Google Docs is a SaaS application that delivers the basic office automation capabilities with support for collaborative editing over the Web. The application is executed on top of the Google distributed computing infrastructure, which allows the system to dynamically scale according to the number of users using the service.

Social networking

Social networking applications have grown considerably in the last few years to become the most active sites on the Web. To sustain their traffic and serve millions of users seamlessly, services such as Twitter and Facebook have leveraged cloud computing technologies.

Media applications

Media applications are a niche that has taken a considerable advantage from leveraging cloud computing technologies. In particular, video-processing operations, such as encoding, transcoding, composition, and rendering, are good candidates for a cloud-based environment. These are computationally intensive tasks that can be easily offloaded to cloud computing infrastructures.

Multiplayer online gaming

Online multiplayer gaming attracts millions of gamers around the world who share a common experience by playing together in a virtual environment that extends beyond the boundaries of a normal LAN. Online games support hundreds of players in the same session, made possible by the specific architecture used to forward interactions, which is based on game log processing.

36. what is salesforce.com

Salesforce.com is probably the most popular and developed CRM solution available today. As of today more than 100,000 customers have chosen Salesforce.com to implement their CRM solutions. The application provides customizable CRM solutions that can be integrated with additional features developed by third parties. Salesforce.com is based on the Force.com cloud development platform. This represents scalable and high-performance middleware executing all the operations of all Salesforce.com applications.

37. What is DataStore? What type of data can be stored in it?

DataStore is a service that allows developers to store semistructured data. The service is designed to scale and optimized to quickly access data. DataStore can be considered as a large object database in which to store objects that can be retrieved by a specified key. Both the type of the key and the structure of the object can vary.

DataStore provides high-level abstractions that simplify interaction with Bigtable. Developers define their data in terms of entity and properties, and these are persisted and maintained by the service into tables in Bigtable.

DataStore also provides facilities for creating indexes on data and to update data within the context of a transaction. Indexes are used to support and speed up queries. A query can return zero or more objects of the same kind or simply the corresponding keys.

38. define sandboxing

Sandboxing One of the major responsibilities of the runtime environment is to provide the application environment with an isolated and protected context in which it can execute without causing a threat to the server and without being influenced by other applications. In other words, it provides applications with a sandbox.

39. Define SimpleDB

Amazon SimpleDB Amazon SimpleDB is a lightweight, highly scalable, and flexible data storage solution for applications that do not require a fully relational model for their data. SimpleDB provides support for semistructured data, the model for which is based on the concept of domains, items, and attributes. With respect to the relational model, this model provides fewer constraints on the structure of data entries, thus obtaining improved performance in querying large quantities of data.

40. define ElastiCache

ElastiCache is an implementation of an elastic in-memory cache based on a cluster of EC2 instances. It provides fast data access from other EC2 instances through a Memcached-compatible protocol so that existing applications based on such technology do not need to be modified and can transparently migrate to ElastiCache.

41. explain Amazon Elastic Block Store (EBS)

The Amazon Elastic Block Store (EBS) allows AWS users to provide EC2 instances with persistent storage in the form of volumes that can be mounted at instance startup. They accommodate up to 1 TB of space and are accessed through a block device interface, thus allowing users to format them according to the needs of the instance they are connected to (raw storage, file system, or other).

42. define Buckets

A bucket is a container of objects. It can be thought of as a virtual drive hosted on the S3 distributed storage, which provides users with a flat store to which they can add objects. Buckets are toplevel elements of the S3 storage architecture and do not support nesting. That is, it is not possible to create “subbuckets” or other kinds of physical divisions. A bucket is located in a specific geographic location and eventually replicated for fault tolerance and better content distribution. Users can select the location at which to create buckets, which by default are created in Amazon’s U.S. datacenters. Once a bucket is created, all the objects that belong to the bucket will be stored in the same availability zone of the bucket.