**VIVA QUESTIONS AND ANSWERS**

1. **Define Cloud Computing with example.**

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.

### What is the working principle of Cloud Computing?

The cloud is a collection of computers and servers that are publicly accessible via the Internet. This hardware is typically owned and operated by a third party on a consolidated basis in one or more data center locations. The machines can run any combination of operating systems.

### What are the advantages and disadvantages of Cloud Computing?

Advantages

Lower-Cost Computers for Users Improved Performance

Lower IT Infrastructure Costs Fewer Maintenance Issues Lower Software Costs

Instant Software Updates Increased Computing Power Unlimited Storage Capacity Increased Data Safety

Improved Compatibility Between Operating Systems Improved Document Format Compatibility

Easier Group Collaboration Universal Access to Documents Latest Version Availability

Removes the Tether to Specific Devices Disadvantages

Requires a Constant Internet Connection

Doesn’t Work Well with Low-Speed Connections Can Be Slow

Features Might Be Limited Stored Data Might Not Be Secure

If the Cloud Loses Your Data, You’re Screwed

1. **What is distributed system?**

A *distributed system* is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal.

Three significant characteristics of distributed systems are:

* + Concurrency of components
  + Lack of a global clock
  + Independent failure of components
  + What is cluster?
  + A computing cluster consists of inter connected stand- alone computers which work cooperatively as a single integrated computing resource .In the past, clustered computer systems have demonstrated

### What is grid computing?

Grid Computing enables virtuals organizations to share geographically distributed resources as they pursue common goals, assuming the absence of central location, central control, omniscience, and an existing trust relationship.

(or)

* + Grid technology demands new distributed computing models, software/middle ware support, network protocols, and hardware infrastructures.
  + National grid projects are followed by industrial grid plat-

Form development by IBM, Microsoft, Sun, HP, Dell, Cisco, EMC, Platform Computing, and others. New grid service providers (zGSPs) and new grid applications have emerged rapidly, similar to the growth of Internet and webservices in the past two decades.

* + Grid systems are classified inessentially two categories: computational or data grids and P2Pgrids.

### What are the business areas needs in Grid computing?

* + Life Sciences
  + Financial services
  + Higher Education
  + Engineering Services
  + Government
  + Collaborative games

### List out the Grid Applications:

* + Application partitioning that involves breaking the problem into discrete pieces
  + Discovery and scheduling of tasks and workflow
  + Data communications distributing the problem data where and when it is required
  + Provisioning and distributing application codes to specific system nodes
  + Autonomic features such as self-configuration, self-optimization, self-recovery and self- management

### List some grid computing toolkits and frameworks?

* + Globus Toolkit Globus Resource Allocation Manager (GRAM)
  + Grid Security Infrastructure (GSI)
  + Information Services
  + Legion, Condor and Condor-G
  + NIMROD, UNICORE, NMI.

### What are Desktop Grids?

These are grids that leverage the compute resources of desktop computers.

Because of the true (but unfortunate) ubiquity of Microsoft® Windows® operating system in corporations, desktop grids are assumed to apply to the Windows environment. The Mac OS™ environment is supported by a limited number of vendors.

### What are Server Grids?

* + Some corporations, while adopting Grid Computing , keep it limited to server resources that are within the purview of the IT department.
  + Special servers, in some cases, are bought solely for the purpose of creating an internal “utility grid” with resources made available to various departments.
  + No desktops are included in server grids. These usually run some flavor of the Unix/Linux operating system.

### Define Opennebula.

OpenNebula is an open source management tool that helps virtualized data centers oversee private clouds, public clouds and hybrid clouds OpenNebula is vendor neutral, as well as platform- and API-agnostic. It

can use KVM, Xen or VMware hypervisors.

### Define Eclipse.

Eclipse is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment.

### Define Netbeans.

NetBeans is an open-source integrated development environment (IDE) for developing with Java, PHP, C++, and other programming languages. NetBeans is also referred to as a platform of modular components used for developing Java desktop applications.

### Define Apache Tomcat.

Apache Tomcat (or Jakarta Tomcat or simply Tomcat) is an open source servlet container developed by the Apache Software Foundation (ASF). Tomcat implements the Java Servlet and the JavaServer Pages

(JSP) specifications from Sun Microsystems, and provides a "pure Java" HTTP web server environment for Java code to run."

### What is private cloud?

The *private cloud* is built within the domain of an intranet owned by a single organization. Therefore, they are client owned and managed. Their access is limited to the owning clients and their partners. Their deployment was not meant to sell capacity over the Internet through publicly accessible interfaces. Private clouds give local users a flexible and agile private infrastructure to run service workloads within their administrative domains.

### What is public cloud?

A *public cloud* is built over the Internet, which can be accessed by any user who has paid for the service. Public clouds are owned by service providers. They are accessed by subscription. Many companies have built public clouds, namely Google App Engine, Amazon AWS, Microsoft Azure, IBM Blue Cloud, and Salesforce Force.com. These are commercial providers that offer a publicly accessible remote interface for creating and managing VM instances within their proprietary infrastructure.

### What is hybrid cloud?

A *hybrid cloud* is built with both public and private clouds, Private clouds can also support a *hybrid cloud* model by supplementing local infrastructure with computing capacity from an external public cloud. For example, the *research compute cloud* (RC2) is a private cloud built by IBM.

### What is a Community Cloud ?

A community cloud in [computing](https://en.wikipedia.org/wiki/Computing) is a collaborative effort in which infrastructure is shared between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally. This is controlled and used by a group of organizations that have shared interest. The costs are spread over fewer users than a public cloud (but more than a private cloud

### Define IaaS?

The IaaS layer offers storage and infrastructure resources that is needed to deliver the Cloud services. It only comprises of the infrastructure or physical resource. Top IaaS Cloud Computing Companies: Amazon (EC2), Rackspace, GoGrid, Microsoft, Terremark and Google.

### Define PaaS?

PaaS provides the combination of both, infrastructure and application. Hence, organisations using PaaS don’t have to worry for infrastructure nor for services. Top PaaS Cloud Computing Companies: Salesforce.com, Google, Concur Technologies, Ariba, Unisys and Cisco..

### Define SaaS?

In the SaaS layer, the Cloud service provider hosts the software upon their servers. It can be defined as a in model in which applications and softwares are hosted upon the server and made available to customers over a network. Top SaaS Cloud Computing Companies: Amazon Web Services, AppScale, CA Technologies, Engine Yard, Salesforce and Windows Azure.

### What is meant by virtualization?

Virtualizationisacomputerarchitecturetechnologybywhichmultiplevirtualmachines (VMs)are multipl exedin the same hardwar emachine.Theideaof VMs canbe dated back to the 1960s. The purpose of a VM is to enhance resource sharing by many users and improve computer performance interms of resource utilization and application flexibility.

1. **What are the implementation levels of virtualization?**

The virtualization types are following

1. OS-level virtualization
2. ISA level virtualization
3. User-ApplicationLevel virtualization
4. hardware level virtualization
5. library level virtualization

### List the requirements of VMM?

There are three requirements for a VMM.

First, a VMM should provide an environment for programs which is essentially identical to the original machine.

Second, programs run in this environment should show, at worst, only minor decreases in speed. Third, a VMM should be in complete control of the system resources.

### Explain Host OS and Guest OS?

A comparison of the differences between a host system, a guest system, and a virtual machine within a virtual infrastructure.

A host system (host operating system) would be the primary & first installed operating system. If you are using a bare metal Virtualization platform like Hyper-V or ESX, there really isn’t a host operating system besides the Hypervisor. If you are using a Type-2 Hypervisor like VMware Server or Virtual Server, the host operating system is whatever operating system those applications are installed into.

A guest system (guest operating system) is a virtual guest or virtual machine (VM) that is installed under the host operating system. The guests are the VMs that you run in your virtualization platform.

### Write the steps for live VM migration?

The five steps for live VM migration is

Stage 0: *Pre-Migration*

Active VM on Host A

Alternate physical host may be preselected for migration Block devices mirrored and free resources maintained Stage 1: *Reservation*

Initialize a container on the target host Stage 2: *Iterative pre-copy*

Enable shadow paging

Copy dirty pages in successive rounds. Stage 3: *Stop and copy*

Suspend VM on host A

Generate ARP to redirect traffic to Host B Synchronize all remaining VM state to Host B Stage 4: *Commitment*

VM state on Host A is released

Stage 5: *Activation*

VM starts on Host B Connects to local devices

Resumes normal operation

### 27..Define Globus Toolkit: Grid Computing Middleware

* Globus is open source grid software that addresses the most challenging problmes in distributed resources sharing.
* The Globus Toolkit includes software services and libraries for distributed security, resource management, monitoring and discovery, and data management.

### Define Blocks in HDFS

* A disk has a block size, which is the minimum amount of data that it can read or write. Filesystems for a single disk build on this by dealing with data in blocks, which are an integral multiple of the disk block size. Filesystem blocks are typically a few kilobytes in size, while disk blocks are normally 512 bytes. This is generally transparent to the filesystem user who is simply reading or writing a file—of whatever length.

### Define Namenodes and Datanodes

* + An HDFS cluster has two types of node operating in a master-worker pattern:
    - a *namenode* (the master) and
    - a number of *datanodes*(workers).
  + The namenode manages the filesystem namespace. It maintains the filesystem tree and the metadata for all the files and directories in the tree. This information is stored persistently on the local disk in the form of two files: the namespace image and the edit log.
* The namenode also knows the datanodes on which all the blocks for a given file are located, however, it does not store block locations persistently, since this information is reconstructed from datanodes when the system starts.

### Define HADOOP.

Hadoop is an open source, Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation.

### Define HDFS.

Hadoop Distributed File System (HDFS) is a Java-based file system that provides scalable and reliable data storage that is designed to span large clusters of commodity servers. HDFS, MapReduce, and YARN form the core of Apache™ Hadoop®.

### Write about HADOOP.

Hadoop was created by Doug Cutting and Mike Cafarella in 2005. Cutting, who was working at Yahoo! at the time, named it after his son's toy elephant. It was originally developed to support distribution for the Nutch search engine project.

1. **Definition of *Grid Portal*:**

A *Grid Portal* provides an efficient infrastructure to put Grid-empowered applications on corporate Intranet/Internet.

### Define GAE.

Google App Engine (often referred to as GAE or simply App Engine) is a [Platform as a Service](https://en.wikipedia.org/wiki/Platform_as_a_Service) and [cloud](https://en.wikipedia.org/wiki/Cloud_computing) [computing](https://en.wikipedia.org/wiki/Cloud_computing) platform for developing and hosting [web applications](https://en.wikipedia.org/wiki/Web_application) in Google-managed [data centers](https://en.wikipedia.org/wiki/Data_center).

Applications are [sandboxed](https://en.wikipedia.org/wiki/Sandbox_(computer_security)) and run across multiple servers. App Engine offers automatic scaling for web applications—as the number of requests increases for an application, App Engine automatically allocates more resources for the web application to handle the additional demand.

1. **What is Cloudsim?**

CloudSim is a simulation toolkit that supports the modeling and simulation of the core functionality of cloud, like job/task queue, processing of events, creation of cloud entities(datacenter, datacenter brokers,

etc), communication between different entities, implementation of broker policies, etc. This toolkit allows to:

* Test application services in a repeatable and controllable environment.
* Tune the system bottlenecks before deploying apps in an actual cloud.
* Experiment with different workload mix and resource performance scenarios on simulated infrastructure for developing and testing adaptive application provisioning techniques

### Core features of CloudSim are:

* The Support of modeling and simulation of large scale computing environment as federated cloud data centers, virtualized server hosts, with customizable policies for provisioning host resources to virtual machines and energy-aware computational resources
* It is a self-contained platform for modeling cloud’s service brokers, provisioning, and allocation policies.
* It supports the simulation of network connections among simulated system elements.
* Support for simulation of federated cloud environment, that inter-networks resources from both private and public domains.
* Availability of a virtualization engine that aids in the creation and management of multiple independent and co-hosted virtual services on a data center node.
* Flexibility to switch between space shared and time shared allocation of processing cores to virtualized services.

1. **Uses of Cloudsim.**

* Load Balancing of resources and tasks
* Task scheduling and its migrations
* Optimizing the Virtual machine allocation and placement policies
* Energy-aware Consolidations or Migrations of virtual machines
* Optimizing schemes for Network latencies for various cloud scenarios

### Define OpenStack.

OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed and provisioned through APIs with common authentication mechanisms.A dashboard is also available, giving administrators control while empowering their users to provision resources through a web interface.

### Define Trystack.

[TryStack](http://trystack.org/) is a great way to take OpenStack for a spin without having to commit to a full deployment.

This free service lets you test what the cloud can do for you, offering networking, storage and compute instances, without having to go all in with your own hardware.

It’s a labor of love spearheaded by three Red Hat OpenStack experts [Will Foster,](https://twitter.com/%40sadsfae) [Kambiz](https://twitter.com/%40kaghaiepour) [Aghaiepour](https://twitter.com/%40kaghaiepour) and [Dan Radez.](https://superuser.openstack.org/articles/openstack-essentials-a-book-to-get-you-grounded-in-the-fundamentals)

TryStack’s set-up must bear the load of anyone who wants to use it, but instead of an equally boundless budget and paid staff, it was originally powered by donated equipment and volunteers from Cisco, Dell, Equinix, NetApp, Rackspace and Red Hat who pulled together for this OpenStack Foundation project.

### Define Hadoop.

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.