UNIT I

1. Define cloud computing

Cloud computing is the delivery of computing as a service rather than a product, hereby shared resources, software, and information are provided to computers and other devices as a utility.

2. What is Distributed computing

This is a field of computer science/engineering that studies distributed systems. A distributed system consists of multiple autonomous computers, each having its own private memory, communicating through a computer network. Information exchange in a distributed system is accomplished through message passing. A computer program that runs in a distributed system is known as a distributed program. The process of writing distributed programs is referred to as distributed programming.

3. Difference between distributed and parallel computing.

Each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.	All processors may have access to a shared memory to exchange information between processors.
It is loosely coupled. An important goal and challenge of distributed systems is location transparency.	It is tightly coupled. Large problems can often be divided into smaller ones, which are then solved concurrently ("in parallel").

4. What is mean by service oriented architecture?

In grids/web services, Java, and CORBA, an entity is, respectively, a service, a Java object, and a CORBA distributed object in a variety of languages. These architectures build on the traditional seven Open Systems Interconnection (OSI) layers that provide the base networking abstractions. On top of this we have a base software environment, which would be .NET or Apache Axis for web services, the Java Virtual Machine for Java, and a broker network for CORBA.

5. What is High Performance Computing(HPC).

supercomputer sites and large data centers must provide high-performance computing services to huge numbers of Internet users concurrently.Because of this high demand, the Linpack Benchmark for high-performance computing (HPC) applications is no longer optimal for measuring system performance.
The emergence of computing clouds instead demands high-throughput computing (HTC) systems built with parallel and distributed computing technologies. We have to upgrade data centers using fast servers, storage systems, and high-bandwidth networks. The purpose is to advance network-based computing and web services with the emerging new technologies.

6. Define peer-to-peer network.

The P2P architecture offers a distributed model of networked systems. Every node acts as both a client and a server, providing part of the system resources. Peer machines are simply client computers connected to the Internet. All client machines act autonomously to join or leave the system freely. This implies that no master-slave relationship exists among the peers. No central coordination or central database is needed.

7. What are the Three New Computing Paradigms

Radio-frequency identification (RFID),
Global Positioning System (GPS),
Internet of Things (IoT).

8. What is degree of parallelism and types

The degree of parallelism (DOP) is a metric which indicates how many operations can be or are being simultaneously executed by a computer. It is especially useful for describing the performance of parallel programs and multi-processor systems.

- Bit-level parallelism (BLP)
- Instruction-level parallelism (ILP)
- VLIW (very long instruction word)
- Data-level parallelism (DLP)
- Multicore processors and chip multiprocessors (CMPs)
- Job-level parallelism (JLP)

9. What is Cyber-Physical Systems

A cyber-physical system (CPS) is the result of interaction between computational processes and the physical world. A CPS integrates ―cyber‖ (heterogeneous, asynchronous) with ―physical‖ (concurrent and information-dense) objects.

10. Define multi core CPU.

Advanced CPUs or microprocessor chips assume a multi-core architecture with dual, quad, six, or more processing cores. These processors exploit parallelism at ILP and TLP levels. CPU has reached its limit in terms of exploiting massive DLP due to the aforementioned memory wall problem

11. Define GPU.

A GPU is a graphics coprocessor or accelerator on a computer's graphics card or video card. A GPU offloads the CPU from tedious graphics tasks in video editing applications. The GPU chips can process a minimum of 10 million polygons per second. GPU's have a throughput architecture that exploits massive parallelism by executing many concurrent threads.

12. Clusters of Cooperative Computers

A computing cluster consists of interconnected stand-alone computers which work cooperatively as a single integrated computing resource.

13. What is single-system image (SSI)

An ideal cluster should merge multiple system images into a single-system image (SSI). Cluster designers desire a cluster operating system or some middleware to support SSI at various levels, including the sharing of CPUs, memory, and I/O across all cluster nodes.

14. What is Grid Computing

Grid computing is the collection of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. Grid computing is distinguished from conventional high performance computing systems such as cluster computing in that grid computers have each node set to perform a different task/application.

15. What is Computational Grids

A computing grid offers an infrastructure that couples computers,software/middleware, special instruments, and people and sensors together. The grid is often constructed across LAN, WAN, or Internet backbone networks at a regional, national, or global scale. Enterprises or organizations present grids as integrated computing resources.

16. What is Overlay Networks and its types

Overlay is a virtual network formed by mapping each physical machine with its ID, logically, through a virtual mapping . When a new peer joins the system, its peer ID is added as a node in the overlay network.

Two types of overlay networks:

1.Unstructured 2. Structured.

17. Write the any three Grid Applications.

- Schedulers
- Resource Broker
- Load Balancing

18. Difference between grid and cloud computing

Grid computing Grids enable access to shared computing power and storage capacity from your desktop	cloud computing Clouds enable access to leasedcomputing power and storage capacity from your desktop
In computing centres distributed across different sites, countries and continents	The cloud providers private data centres which are often centralised in a few locations with excellent network connections and cheap electrical power.
Grids were designed to handle large sets of limited duration jobs that produce or use large quantities of data (e.g. the LHC)	Clouds best support long term services and longer running jobs (E.g. facebook.com)

19. What are the derivatives of grid computing?

There are 8 derivatives of grid computing. They are as follows:

a)Compute grid
b) Data grid
c) Science grid
d) Access grid
e) Knowledge grid
f) Cluster grid
g) Terra grid
h) Commodity grid.

20. What is grid infrastructure?

Grid infrastructure forms the core foundation for successful grid applications. This infrastructure is a complex combination of number of capabilities and resources identified for the specific problem and environment being addressed.

PART – B

1) Explain in detail about virtual organization. (16)
2) Write about the scope of grid computing in business areas. (16)
3) Explain some of the grid application and their usage patterns. (16)
4) Write short notes on. (16)
   - Schedulers
   - Resource broker
   - Load balancing
   - Grid portals
5) What are the data and functional requirements of grid computing? (16)
6) Explain briefly about grid infrastructure. (16)
7) Describe in detail about the Technologies for network based systems? (16)

Unit-2

1.List the OGSA grid service interfaces?

Port Type	Operation
Grid service	Find service data, Termination time and Destroy
Notification source	Subscribe to notification topic
Notification sink	Deliver notification
Registry	Register service and Unregister service
Factory	Create service
Handle map	Find by handle

2. Define Endpoint References in WSRF

The WSRF service addressing mechanism is defined in the WS-addressing standard and uses a term called an endpoint reference (EPR), which is an XML document that contains various information about the service and resource. Specifically, the endpoint reference includes both the service address (URI) and resource identification called a key.

3. What are the specifications of WSRF

WSRF is actually a collection of four specifications (standards):
• WS-ResourceProperties — specifies how resource properties are defined and accessed
• WS-ResourceLifetime — specifies mechanisms to manage resource lifetimes
• WS-ServiceGroup — specifies how to group services or WS-Resources together
• WS-BaseFaults — specifies how to report faults

4. Define Globus 4 information services

Globus 4 information services collectively is called the Monitoring and Discovering System (MDS4 in GT 4) and consists of a set of three WSRF information components:
• Index service
• Trigger service • WebMDS
from which a framework can be constructed for collecting and using information. The three components are part of the full GT4 package.

5. Define WebMDS.

WebMDS (Web Monitoring and Discovering System) is a servlet that provides a Web-based interface to display XML-based information such as resource property information, and as such can be a front-end to index services.

6. Write about the strategies of replication

The strategies of replication can be classified into method types: dynamic and static. For the static method, the locations and number of replicas are determined in advance and will not be modified. Dynamic strategies can adjust locations and number of data replicas according to changes in conditions.

7. Define data grid? List the Grid Data Access Models

A data grid is a set of structured services that provides multiple services like the ability to access alter and transfer very large amounts of geographically separated data, especially for research and collaboration purposes.
1. Monadic model 2. Hierarchical model 3. Federation model 4. Hybrid model

8. Define grid data access Federation model

This model is better suited for designing a data grid with multiple sources of data supplies. Sometimes this model is also known as a mesh model. The data sources are distributed to many different locations. Although the data is shared, the data items are still owned and controlled by their original owners. According to predefined access policies, only authenticated users are authorized to request data from any data source.

9. Write about Parallel Data Transfer

Parallel data transfer opens multiple data streams for passing subdivided segments of a file simultaneously. Although the speed of each stream is the same as in sequential streaming, the total time to move data in all streams can be significantly reduced compared to FTP transfer.

10. Define Striped Data Transfer

Striped data transfer, a data object is partitioned into a number of sections, and each section is placed in an individual site in a data grid. When a user requests this piece of data, a data stream is created for each site, and all the sections of data objects are transferred simultaneously.

11. Write about Monadic access model

This is a centralized data repository model. All the data is saved in a central data repository. When users want to access some data they have to submit requests directly to the central repository. No data is replicated for preserving data locality. This model is the simplest to implement for a small grid.

12. Explain grid data access Hierarchical model

This is suitable for building a large data grid which has only one large data access directory. The data may be transferred from the source to a second-level center. Then some data in the regional center is transferred to the third-level center. After being forwarded several times, specific data objects are accessed directly by users.

13. List the basic functionality requirements of grid service

Discovery and brokering
Metering and accounting
Data sharing Deployment
Virtual organizations
Monitoring
Policy

14. What are the security requirements of grid service

Multiple security infrastructures
Perimeter security solutions
Authentication, Authorization, and Accounting
Encryption
Application and Network-Level Firewalls
Certification

15. List the System Properties Requirements of grid service

Fault tolerance
Disaster recovery Self-
healing capabilities
Strong monitoring
Legacy application management
Administration. Agreement-based
interaction Grouping/aggregation of
services

16. What are the objectives of OGSA?

Manage resources across distributed heterogeneous platforms
Support QoS-oriented Service Level Agreements (SLAs). Provide
a common base for autonomic management
Define open, published interfaces and protocols for the interoperability of diverse resources.

17. Define grid service instance

A grid service instance is a (potentially taransient) service that conforms to a set of conventions, expressed as WSDL interfaces, extensions, and behaviors, for such purposes as lifetime management, discovery of characteristics, and notification.

18. Define grid service handle (GSH)

A grid service handle (GSH) can be thought of as a permanent network pointer to a particular grid service instance. The GSH does not provide sufficient information to allow a client to access the service instance; the client needs to ―resolve‖ a GSH into a grid service reference (GSR).

19. Define grid service reference (GSR).

The GSR contains all the necessary information to access the service instance. The GSR is not a ―permanent‖ network pointer to the grid service instance because a GSR may become invalid for various reasons; for example, the grid service instance may be moved to a different server.

20.List the XML lifetime declaration properties

The three lifetime declaration properties are
1. ogsi:goodFrom
2. ogsi:goodUntil
3. ogsi:availableUntil

PART – B

1. Write short notes on Open Grid Service Architecture. (16)

2.Explain in detail, the functional requirements of OGSA. (16)

3.Explain Practical & Detailed view of OGSA/OGSI. (16)

4.Explain in detail, OGSA services.(16)

5.Describe about the relation of grid architecture with other distributed technologies.(16)

6.What are the third generation initiatives of grid computing?

7.Discuss briefly about organization building and using grid based solution to solve their computing data and network requirements.

UNIT-3

1. Define 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.

2. Define 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.

3. Define 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.

4. List the essential characteristics of cloud computing

1. On-demand capabilities 2. Broad network access 3. Resource pooling
4. Rapid elasticity 5. Measured service

5. List the design objectives of cloud computing.

Shifting Computing from Desktops to Datacenters
Service Provisioning and Cloud Economics Scalability
in Performance
Data Privacy Protection.
High Quality of Cloud Services.

6. Define anything-as-a-service.

Providing services to the client on the basis on meeting their demands at some pay per use cost such as data storage as a service, network as a service, communication as a service etc. it is generally denoted as anything as a service (XaaS).

7. What is mean by SaaS?

The software as a service refers to browser initiated application software over thousands of paid customer. The SaaS model applies to business process industry application, consumer relationship management (CRM), Enterprise resource Planning (ERP), Human Resources (HR) and collaborative application

8. What is mean by IaaS?

The Infrastructure as a Service model puts together the infrastructure demanded by the user namely servers, storage, network and the data center fabric. The user can deploy and run on multiple VM's running guest OS on specific application**.**

9. What is PaaS?

The Platform as a Service model enables the user to deploy user built applications onto a virtualized cloud platform. It includes middleware, database, development tools and some runtime support such as web2.0 and java. It includes both hardware and software integrated with specific programming interface.

10. What is mean by Virtualization?

Virtualization is a computer architecture technology by which multiple virtual machines (VMs) are multiplexed in the same hardware machine. The purpose of a VM is to enhance resource sharing by many users and improve computer performance in terms of resource utilization and application flexibility.

11. Define virtual machine monitor.

A traditional computer runs with a host operating system specially tailored for its hardware architecture, After virtualization, different user applications managed by their own operating systems (guest OS) can run on the same hardware, independent of the host OS. This is often done by adding additional software, called a virtualization layer. This virtualization layer is known as hypervisor or virtual machine monitor (VMM).

12. List the requirements of VMM.

- VMM should provide an environment for programs which is essentially identical to the original machine.
- Programs run in this environment should show, at worst, only minor decreases in speed. VMM should be in complete control of the system resources. Any program run under a VMM should exhibit a function identical to that which it runs on the original machine directly.

13. Define Host OS and Guest OS.

The guest OS, which has control ability, is called Domain 0, and the others are called Domain U. Domain 0 is a privileged guest OS of Xen. It is first loaded when Xen boots without any file system drivers being available. Domain 0 is designed to access hardware directly and manage devices.

14. What are the responsibilities of VMM?

- The VMM is responsible for allocating hardware resources for programs.
- It is not possible for a program to access any resource not explicitly allocated to it.
- It is possible under certain circumstances for a VMM to regain control of resources already allocated.

15. Define CPU virtualization.

CPU architecture is virtualizable if it supports the ability to run the VM's privileged and unprivileged instructions in the CPU's user mode while the VMM runs in supervisor mode. When the privileged instructions including control- and behavior-sensitive instructions of a VM are executed, they are trapped in the VMM. In this case, the VMM acts as a unified mediator for hardware access from different VMs to guarantee the correctness and stability of the whole system.

16. Define memory virtualization.

Virtual memory virtualization is similar to the virtual memory support provided by modern operating systems. In a traditional execution environment, the operating system maintains mappings of virtual memory to machine memory using page tables, which is a one-stage mapping from virtual memory to machine memory. All modern x86 CPUs include a memory management unit (MMU) and a translation look aside buffer (TLB) to optimize virtual memory performance.

17. What is mean by I/O virtualization?

I/O virtualization involves managing the routing of I/O requests between virtual devices and the shared physical hardware. There are three ways to implement I/O virtualization:

- full device emulation, Full device emulation is the first approach for I/O virtualization
- para-virtualization direct I/O.

18. Distinguish the physical and virtual cluster. (Jan.2014)

A physical cluster is a collection of servers (physical machines) connected by a physical network such as a LAN. Virtual clusters have different properties and potential applications. There are three critical design issues of virtual clusters: live migration of virtual machines (VMs), memory and file migrations, and dynamic deployment of virtual clusters.

19.What is memory migration?

Moving the memory instance of a VM from one physical host to another can be approached in any number of ways. Memory migration can be in a range of hundreds of megabytes to a few gigabytes in a typical system today, and it needs to be done in an efficient manner. The Internet Suspend-Resume (ISR) technique exploits temporal locality as memory states are likely to have considerable overlap in the suspended and the resumed instances of a VM.

20. What is mean by host based virtualization?

An alternative VM architecture is to install a virtualization layer on top of the host OS. This host OS is still responsible for managing the hardware. The guest OSes are installed and run on top of the virtualization layer. Dedicated applications may run on the VMs. Certainly, some other applications can also run with the host OS directly.

PART – B

1.Write short notes on cloud deployment model. (16)

2.Explain in detail, categories of cloud. (16)

3.Explain in detail, pros and cons of cloud. (8)

4.Explain in detail, different implementation level of virtualization? (16)

5.Write short notes on OS level virtualization. List the pros and cons of OS level virtualization.

(16)

6.Explain in detail, the virtualization of CPU, Memory and I/O devices. (16)

7.Write short notes on virtual clusters. (8)

8.Explain in detail, the virtualization for data center automation. (16)

UNIT-4

1. List out the grid middleware packages Package	Description
BOINC	Berkeley Open Infrastructure for Network Computing.
UNICORE	Middleware developed by the German grid computing community
Globus (GT4)	A middleware library jointly developed by Argonne National Lab.
CGSP in ChinaGrid	The CGSP (ChinaGrid Support Platform) is a middleware library developed by 20 top universities in China as part of the ChinaGrid Project
Condor-G	Originally developed at the Univ. of Wisconsin for general distributed computing, and later extended to Condor-G for grid job management.
Sun Grid Engine (SGE)	Developed by Sun Microsystems for business grid applications. Applied to private grids and local clusters within enterprises or campuses.

2. Define MapReduce.

The mapreduce software framework provides an abstraction layer with the data flow and flow of control of users and hides implementation of all data flow steps such as data partitioning mapping, synchronization, communication and scheduling. The data flow is such framework is predefined the abstraction layer provides two well defined interface in the form of two functions map and reduce**.**

3. What is the role of Map function?

Each Map function receives the input data split as a set of (key, value) pairs to process and produce the intermediated (key, value) pairs.

4. What is the role of Reduce function?

The reduce worker iterates over the grouped (key, value) pairs, and for each unique key, it sends the key and corresponding values to the Reduce function. Then this function processes its input data and stores the output results in predetermined files in the user's program.

5. List out the Hadoop core fundamental layers

The Hadoop core is divided into two fundamental layers: the MapReduce engine and HDFS. The MapReduce engine is the computation engine running on top of HDFS as its data storage manager. HDFS is a distributed file system inspired by GFS that organizes files and stores their data on a distributed computing system.

6. What are the features of HDFS?

HDFS is not a general-purpose file system, as it only executes specific types of applications, it does not need all the requirements of a general distributed file system. For example, security has never been supported for HDFS systems.

7. List the areas where HDFS cannot be used?

Low-latency data access
Lots of small files
Multiple writers, arbitrary file modifications

8. Why is a block in HDFS so large?

HDFS blocks are large compared to disk blocks, and the reason is to minimize the cost of seeks. By making a block large enough, the time to transfer the data from the disk can be made to be significantly larger than the time to seek to the start of the block. Thus the time to transfer a large file made of multiple blocks operates at the disk transfer rate.

9. Define Namenode in HDFS

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.

10. Define Datanode in HDFS

Datanodes are the work horses of the filesystem. They store and retrieve blocks when they are told to (by clients or the namenode), and they report back to the namenode periodically with lists of blocks that they are storing.

11. What are the permission models for files and directories in HDFS

There are three types of permission: the read permission (r), the write permission (w) and the execute permission (x). The read permission is required to read files or list the contents of a directory. The write permission is required to write a file, or for a directory, to create or delete files or directories in it. The execute permission is ignored for a file since you can't execute a file on HDFS (unlike POSIX), and for a directory it is required to access its children.

12. Define FUSE interface?

Filesystem in Userspace (FUSE) allows filesystems that are implemented in user space to be integrated as a Unix filesystem. Hadoop's Fuse-DFS contrib module allows any Hadoop filesystem (but typically HDFS) to be mounted as a standard filesystem. You can then use Unix utilities (such as ls and cat) to interact with the filesystem, as well as POSIX libraries to access the filesystem from any programming language. Fuse-DFS is implemented in C using *libhdfs* as the interface to HDFS.

13. Define globbing in HDFS?

It is a common requirement to process sets of files in a single operation.. To enumerate each file and directory to specify the input, it is convenient to use wildcard characters to match multiple files with a single expression, an operation that is known as *globbing*.

14. How to process globs in hadoop filesystem?

Hadoop provides two FileSystem methods for processing globs:
public FileStatus[] globStatus(Path pathPattern) throws IOException
public FileStatus[] globStatus(Path pathPattern, PathFilter filter) throws IOException
The globStatus() methods returns an array of FileStatus objects whose paths match the supplied pattern, sorted by path. An optional PathFilter can be specified to restrict the matches further

15. How to delete file or directory in hadoop filesystem?

Use the delete() method on FileSystem to permanently remove files or directories:
public boolean delete(Path f, boolean recursive) throws IOException
If *f* is a file or an empty directory, then the value of recursive is ignored. A nonempty directory is only deleted, along with its contents, if recursive is true (otherwise an IOException is thrown).

16. Define iterative MapReduce.

It is important to understand the performance of different runtime and in particular to compare MPI and map reduce. The two major sources of parallel overhead are load imbalance and communication. The communication overhead in mapreduce can be high for two reasons.

- Mapreduce read and writes files whereas MPI transfer information directly between nodes over the network.
- MPI does not transfer all data from node to node.

17. Define HDFS.

HDFS is a distributed file system inspired by GFS that organizes files and stores their data on a distributed computing system. The hadoop implementation of mapreduce uses the hadoop distributed file system as in underlying layer rather than GFS.

18. List the characteristics of HDFS.

- HDFS fault tolerance
- Block replication
- Relica placement
- Heartbeat and block report messages
- HDFS high throughput access to large dataset
- .

19. What are the operations of HDFS?

The control flow of HDFS operation such as read and write can properly highlights role of the name node and data node in the managing operations. The control flow of the main operations of HDFS on file is further described to manifest the interaction between the users**.**

20. Define block replication.

The reliably store data in HDFS is the file blocks, it is replicated in this system. HDFS store a file as a set of blocks and each block is replicated and distributed across the whole cluster.

21. Define heart beat in Hadoop. What are the advantages of heart beat?

The heart beat are periodic messages sent to the name node by each data node in the cluster. Receipt of a heartbeat implies that data mode is functioning properly while each block report contains list of all blocks in a data mode. The name node receives such messages because it is the sole decision maker of all replicas in the system.

PART -B

1. Explain the Globus Toolkit Architecture (GT4)
2. Explain MapReduce Model in detail
3. Explain Map & Reduce function?
4. Explain HDFS Concepts in detail?
5. Explain Anatomy of a File Read?
6. Explain Anatomy of a File write?

UNIT-5

1. What are the challenges of grid sites

The first challenge is integration with existing systems and technologies.
- The second challenge is interoperability with different hosting environments.
- The third challenge is to construct trust relationships among interacting hosting environments.

2. Define Reputation-Based Trust Model

In a reputation-based model, jobs are sent to a resource site only when the site is trustworthy to meet users' demands. The site trustworthiness is usually calculated from the following information: the defense capability, direct reputation, and recommendation trust.

3. Define direct reputation

Direct reputation is based on experiences of prior jobs previously submitted to the site. The reputation is measured by many factors such as prior job execution success rate, cumulative site utilization, job turnaround time, job slowdown ratio, and so on. A positive experience associated with a site will improve its reputation. On the contrary, a negative experience with a site will decrease its reputation.

4. What are the major authentication methods in the grid?

The major authentication methods in the grid include passwords, PKI, and Kerberos. The password is the simplest method to identify users, but the most vulnerable one to use. The PKI is the most popular method supported by GSI

5. List the types of authority in grid

The authority can be classified into three categories: attribute authorities, policy authorities, and identity authorities. Attribute authorities issue attribute assertions; policy authorities issue authorization policies; identity authorities issue certificates. The authorization server makes the final authorization decision.

6. Define grid security infrastructure

The Grid Security Infrastructure (GSI), formerly called the Globus Security Infrastructure, is a specification for secret, tamper-proof, delegatable communication between software in a grid computing environment. Secure, authenticatable communication is enabled using asymmetric encryption.

7. What are the functions present in GSI

GSI may be thought of as being composed of four distinct functions: message protection, authentication, delegation, and authorization.

8. List the protection mechanisms in GSI

GSI allows three additional protection mechanisms. The first is integrity protection, by which a receiver can verify that messages were not altered in transit from the sender. The second is encryption, by which messages can be protected to provide confidentiality. The third is replay prevention, by which a receiver can verify that it has not.

9. What is the primary information of GSI

GSI authentication, a certificate includes four primary pieces of information: (1) a subject name, which identifies the person or object that the certificate represents; (2) the public key belonging to the subject; (3) the identity of a CA that has signed the certificate to certify that the public key and the identity both belong to the subject; and (4) the digital signature of the named CA

10. Define blue pill

The blue pill is malware that executes as a hypervisor to gain control of computer resources. The hypervisor installs without requiring a restart and the computer functions normally, without degradation of speed or services, which makes detection difficult.

11. What are the host security threats in public IaaS

• Stealing keys used to access and manage hosts (e.g., SSH private keys)
• Attacking unpatched, vulnerable services listening on standard ports (e.g., FTP, SSH)
• Hijacking accounts that are not properly secured (i.e., no passwords for standard accounts)
• Attacking systems that are not properly secured by host firewalls
• Deploying Trojans embedded in the software component in the VM or within the VM image (the OS) itself

12. List the Public Cloud Security Limitations

There are limitations to the public cloud when it comes to support for custom security features. Security requirements such as an application firewall, SSL accelerator, cryptography, or rights management using a device that supports PKCS 12 are not supported in a public SaaS, PaaS, or IaaS cloud.

Any mitigation controls that require deploymentof an appliance or locally attached peripheral devices in the public IaaS/PaaS cloud are not feasible.

13. Define Data lineage

Data lineage is defined as a data life cycle that includes the data's origins and where it moves over time. It describes what happens to data as it goes through diverse processes. It helps provide visibility into the analytics pipeline and simplifies tracing errors back to their sources.

14. Define Data remanence

Data remanence is the residual representation of data that has been in some way nominally erased or removed.

15. What are the IAM processes operational activities.

Provisioning
Credential and attribute management
Entitlement management Compliance management
Identity federation management

16. What are the functions of Cloud identity administrative

Cloud identity administrative functions should focus on life cycle management of user identities in the cloud—provisioning, deprovisioning, identity federation, SSO, password or credentials management, profile management, and administrative management. Organizations that are not capable of supporting federation should explore cloud-based identity management services.

17.List the factors to manage the IaaS virtual infrastructure in the cloud

Availability of a CSP network, host, storage, and support application infrastructure.
Availability of your virtual servers and the attached storage (persistent and ephemeral) for compute services
- Availability of virtual storage that your users and virtual server depend on for storage Service
- Availability of your network connectivity to the Internet or virtual network connectivity to IaaS services.
- Availability of network services

18. What is meant by the terms data-in-transit

It is the process of the transfer of the data between all of the versions of the original file, especially when data may be in transit on the Internet. It is data that is exiting the network via email, web, or other Internet protocols.

19. List the IAM process business category

User management
- Authentication management
- Authorization management
- Access management
- Data management and provisioning
- Monitoring and auditing

20. What are the key components of IAM automation process?

- User Management, New Users
- User Management, User Modifications
- Authentication Management
- Authorization Management

PART B

1. Explain briefly authorization and authentication method
2. Describe grid security infrastructure
3. Describe cloud security infrastructure
4. Discuss in detail the aspects of data security
5. Explain in detail the Architecture of IAM
6. Explain IAM practice in cloud
7. Explain various trust model in grid
8. Explain cloud privacy and privacy by design