CS6703 GRID AND CLOUD COMPUTING

UNIT I
INTRODUCTION

1. What is Grid Computing?

Grid computing is a processor architecture that combines computer resources from various domains to reach a main objective. In **grid computing**, the **computers** on the network can work on a task together, thus functioning as a supercomputer.

2. What is QOS?

Grid computing system is the ability to provide the quality of service requirements necessary for the end-user community. QOS provided by the grid like performance, availability, management aspects, business value and flexibility in pricing.

3. 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

4. What are the features of data grids?

The ability to integrate multiple distributed, heterogeneous and independently managed data sources.
The ability to provide data catching and/or replication mechanisms to minimize network traffic.
The ability to provide necessary data discovery mechanisms, which allow the user to find data based on characteristics of the data.

5. Define – Cloud Computing.
Cloud computing, often referred to as simply "the cloud," is the delivery of on-demand computing resources—everything from applications to data centers—over the Internet on a pay-for-use basis. Storing and accessing data and programs over the Internet instead of your computer's hard drive

6. What is business on demand?

Business On Demand is not just about utility computing as it has a much broader set of ideas about the transformation of business practices, process transformation, and technology implementations.

The essential characteristics of on-demand businesses are responsiveness to the dynamics of business, adapting to variable cost structures, focusing on core business competency, and resiliency for consistent availability.

7. What are the facilities provided by virtual organization?

The formation of virtual task forces, or groups, to solve specific problems associated with the virtual organization.

The dynamic provisioning and management capabilities of the resource required meeting the SLA's.

8. What are the properties of Cloud Computing?

There are six key properties of cloud computing: Cloud computing is

user-centric
task-centric
powerful
accessible
intelligent
programmable

9. Differentiate Cluster computing and Grid Computing.

CHARACTERISTIC	CLUSTER	GRID
Population	Commodity Computers	Commodity and High-end computers
Ownership	Single	Multiple
Discovery	Membership Services	Centralized Index and Decentralized Info
User Management	Centralized	Decentralized

10. What are the types of Cloud service development?

Software as a Service
Platform as a Service
Web Services
On-Demand Computing

11. What is meant by scheduler?

Schedulers are types of applications responsible for the management of jobs, such as allocating resources needed for any specific job, partitioning of jobs to schedule parallel execution of tasks, data management, event correlation, and service-level management capabilities.

12. What is meant by resource broker?

Resource broker provides pairing services between the service requester and the service provider. This pairing enables the selection of best available resources from the service provider for the execution of a specific task.

13. What is load balancing?

Load balancing is concerned with the integrating the system in order to avoid processing delays and over-commitment of resources. It involves partitioning of jobs, identifying the resources and queuing the jobs.

14. 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.

15. Define – Distributed Computing.

Distributed computing is a field of **computer** science that studies **distributed** systems. 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.

UNIT II - GRID SERVICES

1. Define- OSGI.

Open Grid Services Architecture (OGSA) is a set of standards defining the way in which information is shared among diverse components of large, heterogeneous grid systems. In this context, a grid system is a scalable wide area network (WAN) that supports resource sharing and distribution. OGSA is a trademark of the Open Grid Forum.

2. Define – OSGA.

The **Open Grid Services Infrastructure** (OGSI) was published by the Global Grid Forum (GGF) as a proposed recommendation in June 2003.[1] It was intended to provide an infrastructure layer for the Open Grid Services Architecture (OGSA). OGSI takes the statelessness issues (along with others) into account by essentially extending Web services to accommodate grid computing resources that are both transient and stateful.

3. **Define – Peer to Peer Computing.**

Peer to Peer computing is a relatively new computing discipline in the realm of distributed computing. P2P system defines collaboration among a larger number of individuals and/or organizations, with a limited set of security requirements and a less complex resource- sharing topology.

4. **What is Dynamic Accounting System?**

DAS provides the following enhanced categories of accounting functionality to the IPG
commun
ity:

Allows a grid user to request access to a local resource via the presentation of grid credentials

Determines and grants the appropriate authorizations for a user to access a local resource without requiring a preexisting account on the resource to govern local authorizations.

5. **Define – SOA.**

A service-oriented architecture is intended to define loosely coupled and interoperable services/applications, and to define a process for integrating these interoperable components.

6.**What are the major goals of OSGA?**
Identify the use cases that can drive the OGSA platform components.
Identify and define the core OGSA platform components.
Define hosting and platform specific bindings.
Define resource models and resource profiles with interoperable solutions.

7. **What are the layers available in OGSA architectural organizations?**

Native platform services and transport mechanisms.
OGSA hosting environment.

OGSA transport and security.
OGSA infrastructure (OGSI).
OGSA basic services (meta-OS and domain services)

8. **What is meant by grid infrastructure?**

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

9. 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

10.Define - GRAM.

GRAM provides resource allocation, process creation, monitoring, and management services. The most common use of GRAM is the remote job submission and control facility. GRAM simplifies the use of remote systems.

11. What is the role of the grid computing organization?

- Organizations developing grid standards and best practices guidelines.
- Organizations developing grid computing toolkits, frameworks and middleware solutions.
- Organizations building and using grid - based solutions to solve their computing, data, and network requirements.
- Organizations working to adopt grid concepts into commercial products, via utility computing and business on demand computing.
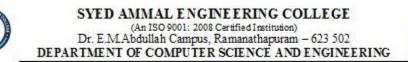
12. What are the different layers of grid architecture?

       Fabric Layer: Interface to local resources
       Connectivity Layer: Manages Communications
       Collective Layer: Coordinating Multiple Resources
       Application Layer: User Defined Application.

13. What are the fundamental components of SOAP specification?

       An envelope that defines a framework for describing message structure.
       A set of encoding rules for expressing instances of application defined data types
       A convention for representing remote procedure (RPC) and responses.
       A set of rules for using SOAP with HTTP.
       Message exchange patterns (MEP) such as request-response, one-way and peer-to- peer conversations.

14. Define - SOAP.

SOAP is a simple and lightweight XML-based mechanism for creating structured data packages that can be exchanged between network applications. SOAP provides a simple enveloping mechanism and is proven in being able to work with existing networking services technologies such as HHTP.SOAP is also flexible and extensible. SOAP is based on the fact that it builds upon the XML info set.

15. Define WSDL.

WSDL is an XML Info set based document, which provides a model and XML format for describe web services. This enables services to be described and enables the client to consume these services in a standard way without knowing much on the lower level protocol exchange binding including SOAP and HTTP. This high level abstraction on the service limits human interaction and enables the automatic generation of proxies for web services, and these proxies can be static or dynamic. It allows both document and RPC - oriented messages.

UNIT III

VIRTUALIZATION

1. **What is the working principle of Cloud Computing?**

    The cloud is a collection of computers and servers that are publicly accessible via the 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
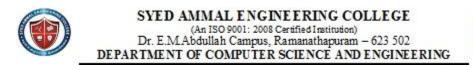
2. **What is Virtualization?**

    Virtualization is a foundational element of **cloud computing** and helps deliver on the value of **cloud computing**," Adams said. "**Cloud computing** is the delivery of shared **computing** resources, software or data — as a service and on-demand through the Internet.

3. **Define Cloud services with example.**
    Any web-based application or service offered via cloud computing is called a cloud
    Cloud services can include anything from calendar and contact applications to word processing and presentations.

4. **What are the types of Cloud service development?**

    Software as a Service

    Platform as a Service

    Infrastructure as a Service

5. **Define cloud provider and cloud broker?**

Cloud Provider: Is a company that offers some component of cloud computing typically infrastructure as a service, software as a Service or Platform as a Service. It is something referred as CSP.

Cloud Broker: It is a third party individual or business that act as an intermediary between the purchase of cloud computing service and sellers of that service.

6. **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.

7. **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.

8. **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.

9. **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).

10. 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.

11. 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**.**

12. Explain 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.

13. List out the advantages of Cloud Computing.

- • 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

14. List out the disadvantages of Cloud Computing.

- • 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

15. What is Hypervisor?

A **hypervisor** or virtual machine monitor (VMM) is a piece of computer software, firmware or hardware that creates and runs virtual machines. A computer on which a **hypervisor** is running one or more virtual machines is defined as a host machine. Each virtual machine is called a guest machine.

16. What are the types of hypervisor?

There are two types of hypervisors:

> Type 1 (bare-metal)

> Type 2 (hosted)

Type 1 hypervisors run directly on the system hardware. They are often referred to as a
"native" or "bare metal" or "embedded" hypervisors in vendor literature.

Type 2 hypervisors run on a host operating system. When the virtualization movement first began to take off, Type 2 hypervisors were most popular. Administrators could buy the software and install it on a server they already had.

UNIT IV

PROGRAMMING MODEL

1. What is Hadoop development?

Apache **Hadoop** is an open-source software **framework** written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.

2. Define- GT4.

Globus Toolkit 4 is a open – source toolkit developed to build grids. It provides full capabilities for sharing computing power and databases. Usage of Globus is extensive throughout the scientific community within NSF, DOE, DARPHA, IBM, NASA, and Microsoft projects.

3. Define- Map Reduce Computation.

MapReduce is designed to continue to work in the face of system failures. When a job is running, MapReduce monitors progress of each of the servers participating in the job. If one of them is slow in returning an answer or fails before completing its work, MapReduce automatically starts another instance of that task on another server that has a copy of the data. The complexity of the error handling mechanism is completely hidden from the programmer

4. What are the characteristics of Cloud Programming Model?

Cost model

Scalability

Fault-tolerance

Support for specific services

Control model

Data model

Synchronization mode

5. What are the phases in MapReduce Programming Model?

Map Phase:

Processes input key/value pair

Produces set of intermediate pair

map (in_key, in_value) -> list(out_key, interm_value)

Reduce Phase:

Combines all intermediate values for a given key

Produces a set of merged output values

reduce(out_key, list(interm_value)) ->

list(out_value)

6. Define- Hadoop Scheduler.

Job divided into several independent tasks executed in parallel

The input file is split into chunks of 64 / 128 MB

Each chunk is assigned to a map task

Reduce task aggregate the output of the map tasks

7. Define- HDFS.

Hadoop File System was developed using distributed file system design. It is run on commodity hardware. Unlike other distributed systems, HDFS is highly fault tolerant and designed using low-cost hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines.

8. What are the features of HDFS?

It is suitable for the distributed storage and processing.

Hadoop provides a command interface to interact with HDFS.

The built-in servers of name-node and data-node help users to easily check the status of cluster.

Streaming access to data in the file system.

HDFS provides file permissions and authentication.

9. What is Cloud Dataflow Programming Model?

The Dataflow programming model is designed to simplify the mechanics of large-scale data processing. When you program with a Dataflow SDK, you are essentially creating a data processing job to be executed by one of the Cloud Dataflow runner services. This model lets you concentrate on the logical composition of your data processing job, rather than the physical orchestration of parallel processing. You can focus on what you need your job to do instead of exactly how that job gets executed.

10. What is AIM?

The most-used instant messaging program is AOL Instant Messenger (www.aim.com), also known as AIM. AIM supports all manner of special features in addition to basic text messaging. The users get file sharing, RSS feeds, group chats, ability to text message to and from mobile phones, voice chat, video chat, and even a mobile client and can also enhance the basic AIM experience with a variety of official and user-created plug-ins.

11. Define- Multi-tenancy.

Multi-tenancy can be defined as a principle in software architecture, where a single instance of a vendor's offering runs on the vendor's servers, serving multiple client organizations (tenants). Often these tenants will pay a fee for this.

In practice, multi-tenancy allows a cloud provider to provide a service to organizations that have users of their own. Of course, in certain cases the tenant could have only 1 user; the important point is that the cloud provider has taken the tenant concept into account and provided e.g. access based on the tenant concept, billing based on the tenant concept, etc.

12. Define- GFS.

Google File System (**GFS** or GoogleFS) is a proprietary **distributed file system** developed by Google for its own use. It is designed to provide efficient, reliable access to data using large clusters of commodity hardware. A new version of the Google File System is codenamed Colossus which was released in 2010.

13.Define- OGF.

OGF is an open global community committed to driving the rapid evolution and adoption of modern advanced applied distributed computing, including cloud, grid and associated storage, networking and workflow methods. OGF is focused on developing and promoting innovative scalable techniques, applications and infrastructures to improve productivity in the enterprise and within the international research, science and business communities

14. Define failover controller.

The transition from the active namenode to the standby is managed by a new entity in the system called the *failover controller.*

15. List the functional modules in GT4 library.
Global Resource Allocation manager
Grid security Infrastructure
Communication
Monitor and Discovery service.

UNIT V
SECURITY

1. **What are the functions in Grid Security Model?**

    Multiple security mechanisms
    Dynamic creation of services
    Dynamic establishment of trust domains

2. **What are OGSA security services?**

    Credential processing service

    Authorization service

    Credential Conversion service

    Identity Mapping service

    Audit

3. **What are the high-level services included in Globas toolkit?**

Globus Resource Allocation Manager(GRAM)

Grid Security Infrastructure(GSI)

Information Services

4. **What are the most common Gt3 handlers?**

Authentication service Hanlder
WS Security Handler
Authorization Handler
X509 Sing Handler
GSS Handler

5. **Define- GSI.**

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.
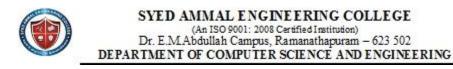
6. **What are the high level grid security requirement aspects?**

Authentication
Authorization
Delegation
Message integrity
Single logon
Confidentiality
Privacy
Policy exchange
Credential life span and renewal
Secure logging
Assurance
Manageability

7. **What is CISCO connected grid security principles?**

Cisco integrates security as a fundamental building block of any network architecture—whether for the field area network, transmission and substation network, or the intra-control center tier. The primary principles behind Cisco Connected Grid security include:

Access control
Data integrity, confidentiality, and privacy
Threat detection and mitigation
Device and platform integrity

8. What are the risks of storing data in the Cloud?

Reliability
Security
User error
Access problems

9. What are the factors to identify the threats in cloud?

Failures in Provider Security
Attacks by Other Customers
Availability and Reliability Issues
Legal and Regulatory Issues
Perimeter Security Model Broken
Integrating Provider and Customer Security Systems

10. What is PaaS Encryption?

Since PaaS is so diverse, the following list may not cover all potential options:

- **Client/application encryption:** Data is encrypted in the PaaS application, or the client accessing the platform.
- **Database encryption:** Data is encrypted in the database using encryption built in and supported by the database platform.
- **Proxy encryption:** Data passes through an encryption proxy before being sent to the platform.

11. Define- Database Activity Monitoring (DAM).

Database Activity Monitors capture and record, at a minimum, all Structured Query Language (SQL) activity in real time or near real time, including database administrator activity, across multiple database platforms; and can generate alerts on policy violations.