JEPPIAAR ENGINEERING COLLEGE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CS6703 – GRID AND CLOUD COMPUTING

QUESTION BANK

DEPARTMENT OF CSE
JEPPIAAR ENGINEERING COLLEGE

VISION

To build Jeppiaar Engineering College as an institution of academic excellence in technology and management education, leading to become a world class university.

MISSION

- To excel in teaching and learning, research and innovation by promoting the principles of scientific analysis and creative thinking
- To participate in the production, development and dissemination of knowledge and interact with national and international communities
- To equip students with values, ethics and life skills needed to enrich their lives and enable them to contribute for the progress of society
- To prepare students for higher studies and lifelong learning, enrich them with the practical skills necessary to excel as future professionals and entrepreneurs for the benefit of Nation's economy

PROGRAM OUTCOMES

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and

write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

VISION OF THE DEPARTMENT

To educate and nurture the upcoming professionals through excellence in scientific and knowledge based education to yield globally competitive and self-disciplined computer engineers.

MISSION OF THE DEPARTMENT

- To create computer professionals, capable of doing research, build innovative ideas and creative solutions for betterment of industries.
- To stimulate and build academic team to cater the ever increasing demand of student community, train them to take uphill challenges through interactions with globally renowned organizations.
- To attain ethical and value added personality that would revamp students life to participate in technology transfer.
- To ignite registrants towards the aptitude of learning every dynamic progress through higher level studies, provide a platform for employment and self-employment to succeed and support the nation.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO1 - Develop Computer Engineers to understand collaborative projects by strengthening problem solving skills, Core computing skills, which offer opportunities for long term interaction with academic and industry.

PEO2 - Establish design, research, product execution and services in the field of Computer Science and Engineering through strong technical, communication and entrepreneurial skills

PEO3 - Support Society by engaging to scrutinize issues of national relevance as well as of global concern.

PEO4 - Contribute to life-long learning through the successful completion of advanced degrees, continuing education, certifications and/or other professional developments.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1-An ability to understand the basic concepts in computer science and engineering and to apply them in various areas like Fundamentals of programming, Data structures ,computer architecture, Theory of computing ,Database management system, computer networks, operating system, ,Software engineering etc in the design and implementation of complex system.

PSO2 - Ability to execute computer science and engineering problem using modern hardware and software tools along with analytical skills to arrive cost effective and appropriate solution**.**

PSO3 - An understanding social awareness and environmental wisdom along with ethical responsibility to have a successful carrier to sustain passion as an entrepreneur. Familiarity and practical proficiency with a broad area of programming concepts and provide new ideas and innovations towards research.

BLOOM TAXANOMY LEVEL (BTL)

BTL6: Creating
BTL 5: Evaluating
BTL 4: Analyzing
BTL 3: Applying
BTL 2: Understanding
BTL 1: Remembering

CS6703 - GRID AND CLOUD COMPUTING SYLLABUS

OBJECTIVES:

The student should be made to:

- Understand how Grid computing helps in solving large scale scientific problems.
- Gain knowledge on the concept of virtualization that is fundamental to cloud computing.
- Learn how to program the grid and the cloud.
- Understand the security issues in the grid and the cloud environment.

UNIT I - INTRODUCTION (9)

Evolution of Distributed computing: Scalable computing over the Internet – Technologies for network based systems – clusters of cooperative computers - Grid computing Infrastructures – cloud computing - service oriented architecture – Introduction to Grid Architecture and standards – Elements of Grid – Overview of Grid Architecture.

UNIT II - GRID SERVICES (9)

Introduction to Open Grid Services Architecture (OGSA) – Motivation – Functionality Requirements –Practical & Detailed view of OGSA/OGSI – Data intensive grid service models – OGSA services.

UNIT III - VIRTUALIZATION (9)

Cloud deployment models: public, private, hybrid, community – Categories of cloud computing: Everything as a service: Infrastructure, platform, software - Pros and Cons of cloud computing – Implementation levels of virtualization – virtualization structure – virtualization of CPU, Memory and I/O devices – virtual clusters and Resource Management – Virtualization for data center automation.

UNIT IV - PROGRAMMING MODEL (9)

Open source grid middleware packages – Globus Toolkit (GT4) Architecture , Configuration – Usage of Globus – Main components and Programming model - Introduction to Hadoop Framework - Mapreduce, Input splitting, map and reduce functions, specifying input and output parameters, configuring and running a job – Design of Hadoop file system, HDFS concepts, command line and java interface, dataflow of File read & File write.

UNIT V - SECURITY (9)

Trust models for Grid security environment – Authentication and Authorization methods – Grid security infrastructure – Cloud Infrastructure security: network, host and application level – aspects of data security, provider data and its security, Identity and access management architecture, IAM practices in the cloud, SaaS, PaaS, IaaS availability in the cloud, Key privacy issues in the cloud.

TOTAL: 45 PERIODS

OUTCOMES:

At the end of the course, the student should be able to:

- Apply grid computing techniques to solve large scale scientific problems.
- Apply the concept of virtualization.
- Use the grid and cloud tool kits.

- Apply the security models in the grid and the cloud environment.

TEXT BOOK

- Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", First Edition, Morgan Kaufman Publisher, an Imprint of Elsevier, 2012.

REFERENCES

1. Jason Venner, "Pro Hadoop- Build Scalable, Distributed Applications in the Cloud", A Press, 2009
2. Tom White, "Hadoop The Definitive Guide", First Edition. O'Reilly, 2009
3. Bart Jacob (Editor), "Introduction to Grid Computing", IBM Red Books, Vervante, 2005
4. Ian Foster, Carl Kesselman, "The Grid: Blueprint for a New Computing Infrastructure", 2nd Edition, Morgan Kaufmann
5. Frederic Magoules and Jie Pan, "Introduction to Grid Computing" CRC Press, 2009
6. Daniel Minoli, "A Networking Approach to Grid Computing", John Wiley Publication, 2005
7. Barry Wilkinson, "Grid Computing: Techniques and Applications", Chapman and Hall, CRC, Taylor and Francis Group, 2010

COURSE OUTCOME

C403.1	Understand the traditional computing architecture and Recent Technologies
C403.2	Elaborate the open standard services for Grid Architecture.
C403.3	Apply the concept of virtualization.
C403.4	Utilize the Grid and Cloud Tool Kit to program on it.
C403.5	Apply the security model in Grid and Cloud Environment

UNIT – I – INTRODUCTION

Evolution of Distributed computing: Scalable computing over the Internet – Technologies for network based systems – clusters of cooperative computers - Grid computing Infrastructures – cloud computing - service oriented architecture – Introduction to Grid Architecture and standards – Elements of Grid – Overview of Grid Architecture

PART A

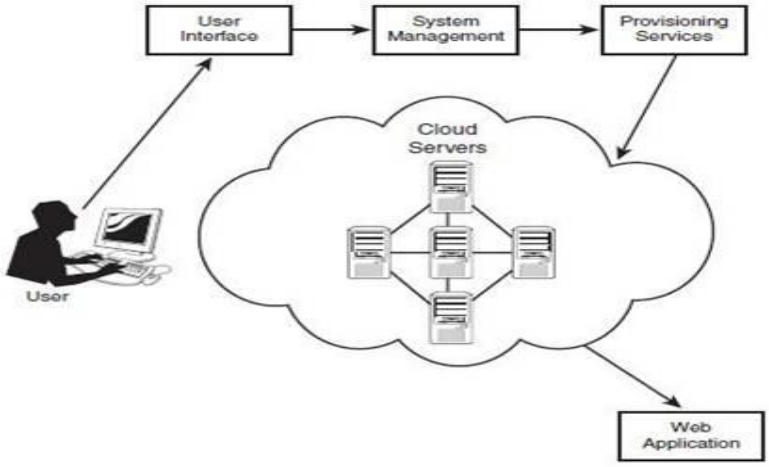
<table>
<tr><th>S. No.</th><th>Question</th><th>Course Outcome</th><th>Blooms Taxonomy Level</th></tr>
<tr><td>1</td><td>Bring out the difference between private cloud and public cloud. (ND2016)
<table>
<tr><th>Public Cloud</th><th>Private Cloud</th></tr>
<tr><td>Multiple Clients</td><td>Single Clients</td></tr>
<tr><td>Hosted at providers location</td><td>Hosted at providers / organization location</td></tr>
<tr><td>Shared infrastructure</td><td>Private Infrastructure</td></tr>
<tr><td>Access over Internet</td><td>Access over Internet / Private network</td></tr>
<tr><td>Low cost</td><td>High cost</td></tr>
<tr><td>Less Secure</td><td>More Secure</td></tr>
</table></td><td>C403.1</td><td>BTL 2</td></tr>
<tr><td>2</td><td>Why is Cloud computing important? (ND2016)

There are many implications of cloud technology, for both developers and end users.
For developers, cloud computing provides
● Increased amounts of storage
● Increased processing power
● Enables new ways to access information, process and analyze data
● Connect people and resources from any location anywhere in the world.
For users,
● Documents hosted in the cloud always exist, no matter what happens to the user’s machine.
● Users from around the world can collaborate on the same documents, applications, and projects, in real time.
Cloud computing does all this at lower costs, because the cloud enables more efficient sharing of resources than does traditional network computing</td><td>C403.1</td><td>BTL 1</td></tr>
<tr><td>3</td><td>What is Grid Computing?
Grid computing is the concept of distributed computing technologies for computing resource sharing among participants in a virtualized collection</td><td>C403.1</td><td>BTL 1</td></tr>
</table>

	of organization.		
4	**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.	C403.1	BTL 1
5	**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	C403.1	BTL 1
6	**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	C403.1	BTL 1
7	**What are the features of computational grids?** • The ability to allow for independent management of computing resources • Failure detection and failover mechanisms	C403.1	BTL 1
8	**What is virtual organization?** Virtual organization is nothing but coordinating resource sharing and problem sharing and dynamic multi institution organization.	C403.1	BTL 1
9	**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.	C403.1	BTL 1
10	**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.	C403.1	BTL 1

	• The dynamic provisioning and management capabilities of the resource required meeting the SLA's.		
11	**What is the definition of grid computing concept given by Foster?** A computational grid is a combination of hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high end-user computational capabilities.	C403.1	BTL 1
12	**What are the business benefits in grid computing?** • Acceleration of implementation time frames in order to intersect with the anticipated business end results. • Robust and infinitely flexible and resilient operational infrastructures. • Avoiding common pitfalls of over provisioning and incurring excess costs	C403.1	BTL 1
13	**What are the examples of major business areas in grid computing?** • Life sciences for analyzing and decoding strings of biological and chemical information. • Financial services for running long, complex financial models and arriving at more accurate decisions. • Higher education for enabling advanced, data and computation intensive research	C403.1	BTL 1
14	**What are the grid computing 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.	C403.1	BTL 1
15	**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.	C403.1	BTL 1
16	**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.	C403.1	BTL 1
17	**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.	C403.1	BTL 1
18	**What are grid portals? Give example.** Grid portals are similar to web portals, in the sense they provide uniform	C403.1	BTL 1

	access to grid resources. **Eg:** Grid portals provide capabilities for the GC resource authentication, remote resource access, scheduling capabilities and monitoring status information.		
19	**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.	C403.1	BTL 1
20	**Give the example of software application ASP.** • Weather Predication • Math Modeling Application	C403.1	BTL 1
21	**Give the examples of Hardware service provider.** • Computer Cluster • Computer System • Linux on Demand • Network Bandwidth • Blades	C403.1	BTL 1
22	**List out any three Grid Applications.** • Schedulers • Resource Broker • Load Balancing	C403.1	BTL 1
23	**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. For example, Google hosts a cloud that consists of both smallish PCs and larger servers. Google's cloud is a private one (that is, Google owns it) that is publicly accessible (by Google's users).	C403.1	BTL 1
24	**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	C403.1	BTL 1
25	**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	C403.1	BTL 1

	operating systems.		
26	**Draw the architecture of Cloud** 	C403.1	BTL 6
27	**Define Cloud services with example.** Any web-based application or service offered via cloud computing is called a cloud service. Cloud services can include anything from calendar and contact applications to word processing and presentations	C403.1	BTL 1
28	**What are the advantages of cloud services?** • If the user's PC crashes host application and document both remain unaffected in the cloud. • An individual user can access applications and documents from any location on any PC. • Because documents are hosted in the cloud, multiple users can collaborate on the same document in real time, using any available Internet connection. • Documents are not machine-centric	C403.1	BTL 1
29	**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	C403.1	BTL 1

	• 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		
30	**Who get benefits from Cloud Computing?** • Collaborators • Road Warriors • Cost-Conscious Users • Cost-Conscious IT Departments • Users with Increasing Needs	C403.1	BTL 1
31	**Who shouldn't be using Cloud Computing?** • The Internet-Impaired • Offline Workers • The Security Conscious • Anyone Married to Existing Applications	C403.1	BTL 1
32	**List the advantages and disadvantages of cloud service deployment.** **Advantages** • Economy of scale • Offer better, cheaper, and more reliable applications • Utilization of the full resources • Less up-front investment • Rapid provisioning • Automatic • Scaling **Disadvantages** • Security • Need Redundancy Tool • No physical backup	C403.1	BTL 1
33	**What are the types of Cloud service development?** • Software as a Service • Platform as a Service • Web Services • On-Demand Computing	C403.1	BTL 1
34	**List the companies who offer cloud service development?** • Amazon • Google App Engine	C403.1	BTL 1

	• IBM • Salesforce.com		
35	**What are the features of robust Cloud development? Who it offers?** • Dynamic web serving • Full support for all common web technologies • Persistent storage with queries, sorting, and transactions • Automatic scaling and load balancing • APIs for authenticating users and sending email using Google Accounts	C403.1	BTL 1
36	**What are the other Cloud service development tools.** • 3tera • 10gen • Cohesive Flexible Technologies • Joyent • Mosso • Nirvanix • Skytap • StrikeIron	C403.1	BTL 1
37	**Define the term web service with example.** A web service is an application that operates over a network typically, over the Internet. Most typically, a web service is an API that can be accessed over the Internet. The service is then executed on a remote system that hosts the requested services. A good example of web services are the "mashups" created by users of the Google Maps API. With these custom apps, the data that feeds the map is provided by the developer, where the engine that creates the map itself is provided by Google.	C403.1	BTL 1
38	**What are the issues in web based applications?** • Technical issues • Business model issues • Internet issues • Security issues • Compatibility issues • Social issues	C403.1	BTL 1
39	**Tabulate the difference between high performance computing and high throughput computing (Apr/May 2017)** • **High-Performance Computing** HPC systems emphasize the raw speed performance. The speed of HPC systems has increased from Gflops in the early 1990s to now Pflops in 2010. This improvement was driven mainly by the demands from scientific, engineering, and manufacturing communities. For example, the Top 500 most powerful computer systems in the world are measured by floating-point speed in Linpack benchmark results. However, the number of supercomputer users is limited to less than 10% of all computer users. Today, the majority of computer users are using desktop computers or large servers when they conduct	C403.1	BTL 2

<table>
<tr><td></td><td>Internet searches and market-driven computing tasks.

• High-Throughput Computing
The development of market-oriented high-end computing systems is undergoing a strategic change from an HPC paradigm to an HTC paradigm. This HTC paradigm pays more attention to high-flux computing. The main application for high-flux computing is in Internet searches and web services by millions or more users simultaneously. The performance goal thus shifts to measure high throughput or the number of tasks completed per unit of time. HTC technology needs to not only improve in terms of batch processing speed, but also address the acute problems of cost, energy savings, security, and reliability at many data and enterprise computing centers.</td><td></td><td></td></tr>
<tr><td>40</td><td>Give basic operations of VM (Apr/May 2017)
• First, the VMs can be multiplexed between hardware machines
• Second, a VM can be suspended and stored in stable storage
• Third, a suspended VM can be resumed or provisioned to a new hardware platform
• Finally, a VM can be migrated from one hardware platform to another</td><td>C403.1</td><td>BTL 2</td></tr>
<tr><td>41</td><td>“Grid inherits the features of P2P and cluster computing systems”. Is the statement true? Validate your answer (Nov/Dec 2017)</td><td>C403.1</td><td>BTL 5</td></tr>
<tr><td>42</td><td>Differentiate grid and cloud computing (Nov/Dec 2017)
<table>
<tr><th>Basis for comparison</th><th>Cloud computing</th><th>Grid computing</th></tr>
<tr><td>Application focus</td><td>business and web-based applications.</td><td>Collaborative purposes.</td></tr>
<tr><td>Architecture used</td><td>Client-server</td><td>Distributed computing</td></tr>
<tr><td>Management</td><td>Centralized</td><td>Decentralized</td></tr>
<tr><td>Business model</td><td>Pay per use</td><td>No defined business model</td></tr>
<tr><td>Accessibility of services</td><td>High because it is real-time</td><td>Low because of scheduled services.</td></tr>
<tr><td>Programming models</td><td>Eucalyptus, Open Nebula, Open stack etc, for Iaas but no middleware exists.</td><td>Different middlewares are available such as Globus gLite, Unicore, etc.</td></tr>
<tr><td>Resource usage patterns</td><td>Centralized manner</td><td>Collaborative manner</td></tr>
<tr><td>Flexibility</td><td>High</td><td>Low</td></tr>
<tr><td>Interoperability</td><td>Vendor lock-in and integration</td><td>Easily deals with interoperability</td></tr>
</table></td><td>C403.1</td><td>BTL 2</td></tr>
</table>

	are some issues between providers.		
43	**"Networks are backbones of grid computing" – Justify your answer (Apr/May 2018)** A grid computing interconnects various pieces of network, providing a path for the exchange of information between different nodes.	C403.1	BTL 5
44	**Differentiate GRIS and GIIS with an illustration (Apr/May 2018)** **Grid Resource Information Service (GRIS)** • Associated with each resource. • Answers queries from client/user about the particular resource. • Accesses an "information provider" deployed on that resource for requested information. **Grid Index Information Service (GIIS)** • A directory service that collects ('pulls") information for GRIS's. • A "caching" service. • Provides indexing and searching functions.	C403.1	BTL 2

PART B

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	Illustrate the architecture of Virtual Machine and brief out the operations.(16) **(ND2016)** Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 149-152	C403.1	BTL 2
2	Explain in detail about distributed computing. (16) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 04-10	C403.1	BTL 2
3	Discuss about virtualization (16) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 22-24	C403.1	BTL 6
4	Explain in detail about clusters of cloud computing. (8) **(ND2016)** Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 28-29	C403.1	BTL 2
5	Discuss short notes on Service Oriented Architecture. (8) **(ND2016)** Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 56-59	C403.1	BTL 6
6	Explain briefly about grid infrastructure. (16) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet",	C403.1	BTL 2

	Page No: 29-31		
7	What are the data and functional requirements of grid computing? (16) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 17-20	C403.1	BTL 1
8	Explain the architecture of Cloud computing in detail. Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 34-36	C403.1	BTL 2
9	Explain the Cloud service development. Refer: Bart Jacob (Editor), "Introduction to Grid Computing", IBM Red Books, Vervante, 2005, Page No: 45-50	C403.1	BTL 2
10	Brief the interaction between GPU and CPU in performing parallel execution of operations (Apr/May 2017) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet",	C403.1	BTL 2
11	Illustrate with the neat sketch, the grid computing infrastructure (Apr/May 2017) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Page No: 29-31	C403.1	BTL 2
12	i) Describe the infrastructure requirements for grid computing (ND2017) ii) What are the issues in Cluster design? How can they be resolved? Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Pg no:29-31, 69-71	C403.1	BTL 1
13	i) Describe layered grid architecture. How does it map onto internet protocol architecture? (ND2017) ii) Describe the architecture of clusters with suitable illustrations Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Pg no: 36-44, 75-86	C403.1	BTL 2
14	Explain in detail the layered architecture of grid environment and functionalities of grid server (Apr/May 2018) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet", Pg no: 36-44	C403.1	BTL 2
15	Discuss the evolution path of cloud computing. Also, express the difference between grid and distributed computing (Apr/May 2018) Refer: Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, "Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet",Pg no:192-205	C403.1	BTL 6

UNIT – II – GRID SERVICES

Introduction to Open Grid Services Architecture (OGSA) – Motivation – Functionality Requirements –Practical & Detailed view of OGSA/OGSI – Data intensive grid service models – OGSA services.

PART A

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	**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.	C403.2	BTL 1
2	**What are the derivatives of grid computing?** There are 8 derivatives of grid computing. They are as follows: • Compute grid • Data grid • Science grid • Access grid • Knowledge grid • Cluster grid • Terra grid • Commodity grid	C403.2	BTL 1
3	**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.	C403.2	BTL 1
4	**What are the features of computational grids?** • The ability to allow for independent management of computing resources • Failure detection and failover mechanisms	C403.2	BTL 1
5	**List the requirements of resource sharing in a grid. (ND 2016)** Grid computing is the concept of distributed computing technologies for computing resource sharing among participants in a virtualized collection of organization. Through resource sharing and cooperation among participating organizations Computational grid or data grid provide • Computing utility • Data services • Information services.	C403.2	BTL 1
6	**What are the security concerns associated with grid? (ND 2016)** The major security problems with grid computing include:	C403.2	BTL 1

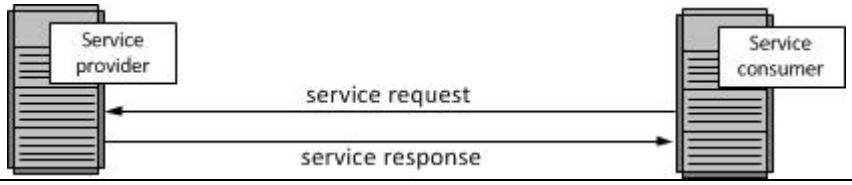
	• **Impact on Local Host:** Grid computing involves running an alien code in the host system. This external code can hamper jobs running locally, and compromise local data security. • **Vulnerable Hosts**: Clients using the grid remain in danger from the local hosts. The major vulnerabilities include the local hosts shutting down resulting in denial of service, viruses, or other malware in the local host affecting the entire process, and local hosts compromising client data integrity and confidentiality. • **Interception**: One major security risk with grid computing is an attacker intercepting the resources and data in the grid. The attack can take various forms such as a distributed denial-of-service (DDOS) attack and the like.		
7	**What is virtual organization?** Virtual organization is nothing but coordinating resource sharing and problem sharing and dynamic multi institution organization.	C403.2	BTL 1
8	**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.	C403.2	BTL 1
9	**What are the grid computing 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.	C403.2	BTL 1
10	**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.	C403.2	BTL 1
11	**Give the example of software application ASP.** • Weather Predication • Math Modeling Application	C403.2	BTL 1
12	**What are the three Grid Applications.** • Schedulers • Resource Broker • Load Balancing	C403.2	BTL 1
13	**What are the collective services available in grid computing?** • Discovery services • Co allocation, scheduling, and brokering services • Monitoring and diagnostic services • Data replication services	C403.2	BTL 1

	• Grid-enabled programming systems • Software discovery services		
14	**What are the basic principles of autonomous computing?** • Self-configuring (able to adapt to the changes in the system) • Self-optimizing (able to improve performance) • Self-healing (able to recover from mistakes) • Self-protecting (able to anticipate and cure intrusions)	C403.2	BTL 1
15	**Name the classification of GC organization based on their Functional role.** • Organizations developing grid standards and best practices guidelines. • Organizations developing GC 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.	C403.2	BTL 1
16	**What are the basic goals of GGF?** • Create an open process for the development of the grid agreements and specifications. • Create grid specifications, architecture documents and best practice guidelines. • Manage and version controls the documents and specifications. • Handle intellectual property policies. • Provide a forum for information exchange and collaboration. • Improve collaboration among the people involved in the grid research, grid deployment and grid users. • Create best practice guidelines from the experience of the technologies associated with GC. • Educate on advances in the grid technologies and share experiences among the people of interest.	C403.2	BTL 1
17	**What are the major works of GGF?** • Application and programming environments • Data • Architecture • Information system and performance • Peer to peer: desktop grids • Scheduling and resource management • Security	C403.2	BTL 1
18	**What are the high level services including in existing globus tool kit?** • GRAM (Globus Resource Allocation Manager) • GSI (Grid Security Infrastructure) • Information services.	C403.2	BTL 1

19	**Mention the important characteristic of legion system** • Everything is an object • Classes manage their own instance, users can provide their own classes	C403.2	BTL 1
20	**What are the core objects defined by legion system?** **Host objects:** Abstractions of processing resources which may represent a single processor or multi host and processors. **Value objects:** Provide persistent storage for scalable persistence of the objects. **Binding objects:** Maps the object ID's to the physical addresses **Implementation objects:** Allow objects to run as processes.	C403.2	BTL 1
21	**Name the components available in Nimrod architecture?** **Nimrod-G clients**: This can provide tools for creating parameter sweep applications, steering and control monitors, and customized end-user applications and GUI's. **Nimrod-G resource broker**: it consists of a Task Farming Engine (TFE), a scheduler that performs resource discovery, trading and scheduling features, a dispatcher and actuator, and agents for managing the jobs on the resource.	C403.2	BTL 1
22	**What are the scheduling algorithms used in Nimrod G?** • Cost optimization- uses the cheapest resource. • Time optimization- results in parallel execution of the job. • Cost-time optimization-similar to cost optimization but if there are multiple jobs with the same cost, then the time factor is taken into consideration. • Conservative time strategy- similar to time optimization, but guarantees that each unprocessed job has a minimum budget per job.	C403.2	BTL 1
23	**What are the major objectives of Euro grid project?** • To establish a European GRID network of leading high performance computing centers from different European countries. • To operate and support the EUROGRID software infrastructure. • To develop important GRID software components and to integrate them into EUROGRID • To demonstrate distributed simulated codes from different application areas • To contribute to the international GRID development and work with the leading international GRID projects.	C403.2	BTL 1
24	**What is the application specific work packages identified for the Euro grid?** • Bio-Grid • Metro Grid • Computer Aided Engineering (CAE) Grid	C403.2	BTL 1

	• High performance center (HPC) research Grid.		
25	**Define dynamic accounting system.** • DAS provides the following enhanced categories of accounting functionality to the IPG community: • 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. • Provides resource pricing information on the grid.	C403.2	BTL 1
26	**Mention the characteristic of connectivity layer?** • Single sign-on • Delegation • Integration with local resource specific security solutions • User- based trust relationships • Data security	C403.2	BTL 1
27	**What are the two primary classes of resource layer protocols?** The resource protocols are the key to operations and integrity of any single resource. These protocols are as follows: • Information protocols • Management protocols	C403.2	BTL 1
28	**What are the collective services available in grid computing?** • Discovery services • Co allocation, scheduling, and brokering services • Monitoring and diagnostic services • Data replication services • Grid-enabled programming systems • Software discovery services	C403.2	BTL 1
29	**What are the basic principles of autonomous computing?** • Self-configuring (able to adapt to the changes in the system) • Self-optimizing (able to improve performance) • Self-healing (able to recover from mistakes) • Self-protecting (able to anticipate and cure intrusions)	C403.2	BTL 1
30	**What are the four essential characteristics of on demand business?** • **Responsive:** Business On Demand has to be responsive to dynamic, unpredictable changes in demand, supply, pricing, labor, and competition. • **Variable:** Business on Demand has to be flexible in adapting to variable cost structure and processes associated with productivity, capital, and finance. • **Focused:** Business On Demand has to focus on their core competency, its differentiating tasks and assets along with closer integration with its partners.	C403.2	BTL 1

	• **Resilient:** A Business On Demand company has to be capable of managing changes and competitive threats with consistent availability and security.		
31	**What are the essential capabilities provided by on demand business?** • Integrate • Virtualization • Automation • Open standards	C403.2	BTL 1
32	**What are the two most important technologies for building semantic webs?** • XML • Resource Description Framework(RDF)	C403.2	BTL 1
33	**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. • Both P2P and distributed computing are focused on resource sharing	C403.2	BTL 1
34	**Write the combination of Globus GT3 toolkit?** • GT3 core • Base services • User- defined services	C403.2	BTL 1
35	**What is a GT3 core?** • It provides a framework to host the high-level services. • The core consists of OGSI reference implementation, security infrastructure, and System level services.	C403.2	BTL 1
36	**What do you understand by the term 'data intensive'? (Apr/May 2017)** Data intensive refers to using a lot of data. Data-intensive computing is a class of parallel computing applications which use a data parallel approach to process large volumes of data typically terabytes or petabytes in size and typically referred to as big data. Computing applications which devote most of their execution time to computational requirements are deemed compute-intensive, whereas computing applications which require large volumes of data and devote most of their processing time to I/O and manipulation of data are deemed data-intensive	C403.2	BTL 1
37	**Define OGSA (Apr/May 2017)** 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.	C403.2	BTL 1

38	**Compare GSH with GSR (Nov/Dec 2017)** GSH & GSR • GSH: Grid Service Handle (URI) - Unique - Shows the location of the service • GSR: Grid Service Reference - Describes how to communicate with the service - As Web Service use SOAP, our GSR is a WSDL file	C403.2	BTL 2
39	**What is the purpose of Grid service description (Nov/Dec 2017)** A grid service description describes how a client interacts with service instances.This description is independent of any particular instance. Within a WSDL document, the grid service description is embodied in the most derived portType of the instance, along with its associated portTypes, bindings, messages, and types definitions.	C403.2	BTL 1
40	**Justify that web and web architecture are SOA based (Apr/May 2018)** The technology of Web Services is the most likely connection technology of service-oriented architectures. The following figure illustrates a basic service-oriented architecture. It shows a service consumer at the right sending a service request message to a service provider at the left. The service provider returns a response message to the service consumer. The request and subsequent response connections are defined in some way that is understandable to both the service consumer and service provider. A service provider can also be a service consumer.  	C403.2	BTL 1
41	**List the services provided by grid infrastructure (Apr/May 2018)** OGSA SERVICES: • Common Management Model (CMM) • Service domains • Distributed data access and replication. • Policy, security • Provisioning and resource management.	C403.2	BTL 1

PART- B

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	Explain about Open Grid Services Architecture (OGSA). (TB Pg.422-425) • Infrastructure services • Execution Management services • Data services	C403.2	BTL 1

	• Resource Management services • Security services • Self-management services • Information services		
2	Explain about the Motivation in OGSA. (Ref Book6 – Pg.109-113) • Data management • Dispatch management • Information services • Scheduling • Security • Work unit management • Increased effective computing capacity • Interoperability of resources • Speed of application development	C403.2	BTL 1
3	Explain about the functionality requirements in OGSA. (Ref Book6 – Pg.177-183) • Basic Functionality Requirements • Security Requirements • Resource Management Requirements • System Properties Requirements • Other Functionality Requirements	C403.2	BTL 1
4	Write about the detailed view of OGSA. (Ref Book6 – Pg.139-151) **(ND 2016)** • Setting the Context • The Grid Service • WSDL Extensions and Conventions • Service Data • Core Grid Service Properties	C403.2	BTL 1
5	Write in detail about OGSA Services. (Ref Book6 – Pg.164-173)/ (TB –Pg. 283-287) • Infrastructure services • Execution Management services • Data services • Resource Management services • Security services • Self-management services • Information services	C403.2	BTL 1
6	Explain in detail about data intensive grid service models with suitable diagrams. (TB –Pg. 425-427) (ND2016)	C403.2	BTL 1
7	Write a detailed notes on OGSA security models (Apr/May 2017) (TB – Pg. 422-424)	C403.2	BTL 1
8	Explain how migrations of grid services are handled? (TB –Pg. 283)	C403.2	BTL 1
9	"Data produced by a large Hadron Collider may exceed several petabyts". What type of grid service model(s) will you suggest for such an application? Illustrate with diagrams. (Nov/Dec 2017) (Apr/May 2017)	C403.2	BTL 1

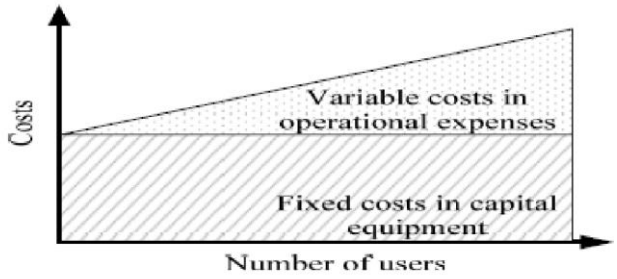
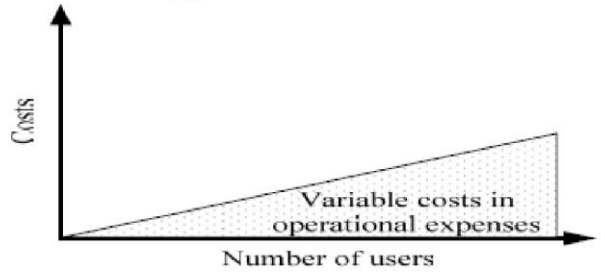
	(TB –Pg. 425-426)		
10	What is OGSA? Explain open grid service architecture in detail with the functionalities of the components. (Nov/Dec 2017) (TB –Pg. 283-286)	C403.2	BTL 1
11	Explain in detail the OGSA security architecture and security services. (Apr/May 2018) (TB –Pg. 283-286)	C403.2	BTL 1
12	What is the purpose of OGSI? Describe the ports and interfaces defined in OGSI along with its inheritance hierarchy (Apr/May 2018) (TB – Pg no: 283)	C403.2	BTL 1

UNIT – III – VIRTUALIZATION

Cloud deployment models: public, private, hybrid, community – Categories of cloud computing: Everything as a service: Infrastructure, platform, software - Pros and Cons of cloud computing – Implementation levels of virtualization – virtualization structure – virtualization of CPU, Memory and I/O devices – virtual clusters and Resource Management – Virtualization for data center automation.

PART A

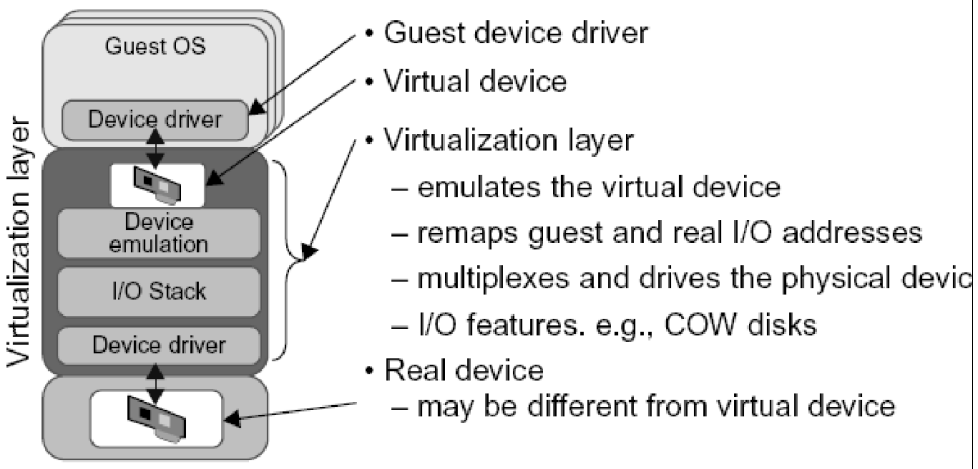
S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	**What is meant by Service Level Agreement (SLA)?** A service-level agreement (SLA) is a part of a standardized service contract where a service is formally defined. Particular aspects of the service includes scope, quality, responsibilities which are agreed between the service provider and the service user.	C403.3	BTL 1
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. Eg. Google App Engine GAE, Amazon Web Services AWS, Microsoft Azure, IBM Blue Cloud etc.	C403.3	BTL 1
3	**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.	C403.3	BTL 1
4	**Define Hybrid Cloud. (ND 2016)** Hybrid cloud is a cloud computing environment which uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between the two platforms. For example, an enterprise can deploy an on-premises private cloud to host sensitive or critical workloads, but use a third-party public cloud provider, such as Google_Computer Engine, to host less-critical resources, such as test and development workloads.	C403.3	BTL 1
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 • New Standards and Interfaces	C403.3	BTL 1
6	**Compare traditional IT cost model and Cloud computing cost model.** In traditional IT computing, user must acquire their own computer and peripheral equipment as capital expenses. In addition, they have to face operational expenditure in operating and maintaining the computer systems, including the human and service costs. The operational costs may	C403.3	BTL 1

	increases sharply with larger number of users. Therefore, the total cost escalates quickly with massive number of users. On the other hand, Cloud computing applies a pay-per-use business model. User jobs are outsourced to the datacenters. To use cloud, there is no out-front costs in acquiring heavy machines. Only variable costs are experienced by cloud users. Overall, cloud computing will reduce the computing costs significantly for both small users and large enterprises.  (a) Traditional IT cost model  (b) Cloud computing cost model		
7	**List the different types of cloud service models?** • Infrastructure as a Service IaaS • Platform as a Service PaaS • Software (application) as services SaaS	C403.3	BTL 1
8	**Write short notes on Infrastructure as a Service (Iaas)?** IaaS model allows users to rent processing, storage, networks, and other resources. The user can deploy and run the guest OS and applications. The user does not manage or control the underlying cloud infrastructure but has control over OS, storage, deployed applications, and possibly select networking components. This IaaS model encompasses the storage as a service, computation resource as a service, and communication resource as a service. Example for this kind of service is: Amazon-S3 for storage, Amazon-EC2 for computation resources, and Amazon-SQS for communication resources.	C403.3	BTL 1
9	**Write short notes on Platform as a Service (PaaS)?** Platform as a service (PaaS) is a category of cloud computing services that provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app. The PaaS model provides the user to deploy user-built applications on top of the cloud infrastructure, that are built using the programming languages	C403.3	BTL 1

	and software tools supported by the provider (e.g., Java, python, .Net).		
10	**Write short notes on Software as a Service (SaaS)?** Software as a Service (SaaS) is a software delivery method that provides access to software and its functions remotely as a Web-based service. SaaS model provides the software applications as a service. As a result, on the customer side, there is no upfront investment in servers or software licensing. On the provider side, costs are rather low, compared with conventional hosting of user applications. The customer data is stored in the cloud that is either vendor proprietary or a publically hosted cloud supporting the PaaS and IaaS.	C403.3	BTL 1
11	**List some of the advantages of cloud computing?** • Cost Efficient • Unlimited Storage • Backup and Recovery • Automatic Software Integration • Easy Access to Information • Quick Deployment	C403.3	BTL 1
12	**List some of the disadvantages of cloud computing?** • **Technical Issues -** technology is always prone to outages and other technical issues. • **Security -** need to make absolutely sure that you choose the most reliable service provider, who will keep your information totally secure • **Prone to Attack -** Storing information in the cloud could make your company vulnerable to external hack attacks and threats. • **Limited Control -** customer can only control and manage the applications, data and services operated on top of that, not the backend infrastructure itself	C403.3	BTL 1
13	**Define Virtual Machine / Role of VM (ND 2016)** A virtual machine is a software computer that, like a physical computer, runs an operating system and applications. The virtual machine is comprised of a set of specification and configuration files and is backed by the physical resources of a host. 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.	C403.3	BTL 1
14	**Define Virtualization.** Virtualization is a computer architecture technology by which multiple virtual machines (VMs) are multiplexed in the same hardware machine. The idea is to separate the hardware from the software to yield better system efficiency. For example, computer users gained access to much enlarged memory space when the concept of virtual memory was introduced. Similarly, virtualization techniques can be applied to enhance the use of compute engines, networks, and storage.	C403.3	BTL 1
15	**List down the various levels of virtualization?** • Application level	C403.3	BTL 1

	• Library (user-level API) level • Operating system level • Hardware abstraction layer (HAL) level • Instruction set architecture (ISA) level		
16	**List the requirements for 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	C403.3	BTL 1
17	**Define OS-Level Virtualization?** Operating-system-level virtualization is a server virtualization method in which the kernel of an operating system allows the existence of multiple isolated user-space instances, instead of just one. Operating system virtualization inserts a virtualization layer inside an operating system to partition a machine's physical resources. It enables multiple isolated VMs within a single operating system kernel. This kind of VM is often called a virtual execution environment (VE), Virtual Private System (VPS), or simply container.	C403.3	BTL 1
18	**List the advantages of OS Extensions?** • VMs at the operating system level have minimal startup/shutdown costs, low resource requirements, and high scalability • For an OS-level VM, it is possible for a VM and its host environment to synchronize state changes when necessary.	C403.3	BTL 1
19	**Write down the disadvantages of OS Extensions?** The main disadvantage of OS extensions is that all the VMs at operating system level on a single container must have the same kind of guest operating system. That is, although different OS-level VMs may have different operating system distributions, they must pertain to the same operating system family.	C403.3	BTL 1
20	**What is paravirtualization or OS Assisted Virtualization?** Paravirtualization is virtualization in which the guest operating system (the one being virtualized) is aware that it is a guest and accordingly has drivers that, instead of issuing hardware commands, simply issue commands directly to the host operating system. This also includes memory and thread management as well, which usually require unavailable privileged instructions in the processor. Para-virtualization attempts to reduce the virtualization overhead, and thus improve performance by modifying only the guest OS kernel. **Eg.** KVM (Kernel-Based VM) - Linux para-virtualization system	C403.3	BTL 1
21	**What is full virtualization?** **Full Virtualization** is virtualization in which the guest operating system is unaware that it is in a virtualized environment, and therefore hardware is virtualized by the host operating system so that the guest can issue commands to what it thinks is actual hardware, but really are just simulated hardware devices created by the host. With full virtualization, noncritical	C403.3	BTL 1

	instructions run on the hardware directly while critical instructions are discovered and replaced with traps into the VMM to be emulated by software. Both the hypervisor and VMM approaches are considered full virtualization.		
22	**What is Host-Based Virtualization?** In Host-Based Virtualization a virtualization layer installed on the 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.	C403.3	BTL 1
23	**What are the advantages of Host-Based Virtualization?** • The user can install this VM architecture without modifying the host OS. The virtualizing software can rely on the host OS to provide device drivers and other low-level services. This will simplify the VM design and ease its deployment. • The host-based approach appeals to many host machine configurations.	C403.3	BTL 1
24	**What is Hardware Assisted Virtualization?** **Hardware Assisted Virtualization** is a type of Full Virtualization where the microprocessor architecture has special instructions to aid the virtualization of hardware. These instructions might allow a virtual context to be setup so that the guest can execute privileged instructions directly on the processor without affecting the host. Such a feature set is often called a **Hypervisor**. In this way, the VMM and guest OS run in different modes and all sensitive instructions of the guest OS and its applications are trapped in the VMM. To save processor states, mode switching is completed by hardware.	C403.3	BTL 1
25	**What is Hybrid Virtualization?** Hybrid Virtualization is a combination of Para Virtualization and Full Virtualization where parts of the guest operating system use para virtualization for certain hardware drivers, and the host uses full virtualization for other features.	C403.3	BTL 1
26	**Define CPU Virtualization.** CPU virtualization involves a single CPU acting as if it were two separate CPUs. In effect, this is like running two separate computers on a single physical machine. Perhaps the most common reason for doing this is to run two different operating systems on one machine. The aim of CPU virtualization is to make a CPU run in the same way that two separate CPUs would run. A 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.	C403.3	BTL 1
27	**What is meant by memory virtualization?** Virtual memory virtualization involves sharing the physical system memory in RAM and dynamically allocating it to the physical memory of the VMs. That means a two-stage mapping process should be maintained by the guest OS and the VMM, respectively: virtual memory to physical	C403.3	BTL 1

	memory and physical memory to machine memory.		
28	**What is meant by I/O Virtualization?** I/O virtualization involves managing the routing of I/O requests between virtual devices and the shared physical hardware. I/O virtualization technology allows a single physical adapter to be visualized as multiple virtual network interface cards (vNICs) and virtual host bus adapters (vHBAs).	C403.3	BTL 1
29	**Explain full device emulation method of I/O Virtualization?**  All the functions of a device or bus infrastructure, such as device enumeration, identification, interrupts, and DMA, are replicated in software. This software is located in the VMM and acts as a virtual device. The I/O access requests of the guest OS are trapped in the VMM which interacts with the I/O devices.	C403.3	BTL 2
30	**Explain Para-virtualization method of I/O Virtualization?** The para-virtualization method of I/O virtualization is typically used in Xen. It is also known as the split driver model consisting of a frontend driver and a backend driver. The frontend driver is running in Domain U and the backend driver is running in Domain 0. They interact with each other via a block of shared memory. The frontend driver manages the I/O requests of the guest OSes and the backend driver is responsible for managing the real I/O devices and multiplexing the I/O data of different VMs.	C403.3	BTL 2
31	**Explain Direct I/O method of I/O Virtualization?** Direct I/O virtualization lets the VM access devices directly. It can achieve close-to-native performance without high CPU costs.	C403.3	BTL 2
32	**Explain self-virtualized I/O (SV-IO)?** The key idea of SV-IO is to harness the rich resources of a multicore processor. All tasks associated with virtualizing an I/O device are encapsulated in SV-IO. It provides virtual devices and an associated access API to VMs and a management API to the VMM. SV-IO defines one virtual interface (VIF) for every kind of virtualized I/O device, such as virtual network interfaces, virtual block devices (disk), virtual camera	C403.3	BTL 2

	devices, and others. The guest OS interacts with the VIFs via VIF device drivers. Each VIF consists of two message queues. One is for outgoing messages to the devices and the other is for incoming messages from the devices. In addition, each VIF has a unique ID for identifying it in SV-IO		
33	**Define Virtual Hierarchy?** A virtual hierarchy is a cache hierarchy that can adapt to fit the workload or mix of workloads. The hierarchy's first level locates data blocks close to the cores needing them for faster access, establishes a shared-cache domain, and establishes a point of coherence for faster communication. When a miss leaves a tile, it first attempts to locate the block (or sharers) within the first level. The first level can also provide isolation between independent workloads. A miss at the L1 cache can invoke the L2 access.	C403.3	BTL 1
34	**Write short notes on virtual clusters?** Virtual clusters are built with virtual machines (VMs) installed at distributed servers from one or more physical clusters. The VMs in a virtual cluster are interconnected logically by a virtual network across several physical networks.	C403.3	BTL 1
35	**List the steps to deploy a group of VMs onto a target cluster?** • Prepare the disk image. • Configure the VMs . • Choose destination nodes. • Execute the VM deployment command on every host.	C403.3	BTL 1
36	**Define Data Center Automation?** Data-center automation means that huge volumes of hardware, software, and database resources in these data centers can be allocated dynamically to millions of Internet users simultaneously, with guaranteed QoS and cost-effectiveness. This automation process is triggered by the growth of virtualization products and cloud computing services.	C403.3	BTL 1
37	**List the benefits of server virtualization?** • Consolidation enhances hardware utilization. • This approach enables more agile provisioning and deployment of resources • The total cost of ownership is reduced • This approach improves availability and business continuity.	C403.3	BTL 1
38	**Write short notes on Virtualization-based intrusion detection?** Virtualization-based intrusion detection can isolate guest VMs on the same hardware platform. Even some VMs can be invaded successfully; they never influence other VMs, which is similar to the way in which a NIDS operates. Furthermore, a VMM monitors and audits access requests for hardware and system software. This can avoid fake actions and possess the merit of a HIDS.	C403.3	BTL 1
39	**Mention the characteristic features of the cloud (Apr/May 2017)**	C403.3	BTL 1
40	**Summarize the difference between PaaS and SaaS (Apr/May 2017)**	C403.3	BTL 1
41	**List the requirements of VMM (Nov/Dec 2017)**	C403.3	BTL 1

42	**Distinguish between physical and virtual clusters (Nov/Dec 2017)**	C403.3	BTL 1
43	**How does performance enhances by virtualizing the data center? (Apr/May 2018)**	C403.3	BTL 1
44	**"Although virtualization is widely accepted today, it does have its limits." Comment on the statement. (Apr/May 2018)**	C403.3	BTL 1

PART B

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	List the Cloud deployment models and give a detail note about them. (T1: pgs 192-196) **(ND2016)** • Public • Private • Hybrid	C403.3	BTL 1
2	Explain in detail about different type of service models in cloud computing? (T1: pgs 200-205) **(ND2016)** • IaaS • PaaS • SaaS	C403.3	BTL 2
3	Explain in detail about various levels of virtualization? (T1: pgs 130-133)	C403.3	BTL 2
4	Explain in detail about binary translation with Full Virtualization? (T1: pgs 141-143)	C403.3	BTL 2
5	Explain in detail about Para-virtualization with compiler support? (T1: pgs 143-144)	C403.3	BTL 2
6	Explain in detail about CPU Virtualization? (T1: pgs 147-148)	C403.3	BTL 2
7	Explain in detail about Virtual Cluster and Resource management? (T1: pgs 155-169)	C403.3	BTL 2
8	Explain in detail about virtualization for Data-Center Automation? (T1: pgs 169-177)	C403.3	BTL 2
9	Discuss how virtualization is implemented in different layers (Apr/May 2017) (T1: pgs 130-133)	C403.3	BTL 2
10	What do you mean by data center automation using virtualization (Apr/May 2017) (TB – Pg no: 169 -178)	C403.3	BTL 1
11	Describe service and deployment models of cloud computing environment with illustrations. How do they fit in NIST cloud architecture (Nov/Dec 2017) (TB – Pg no:192)	C403.3	BTL 2
12	What is virtualization? Describe para and full virtualization. Compare and	C403.3	BTL 2

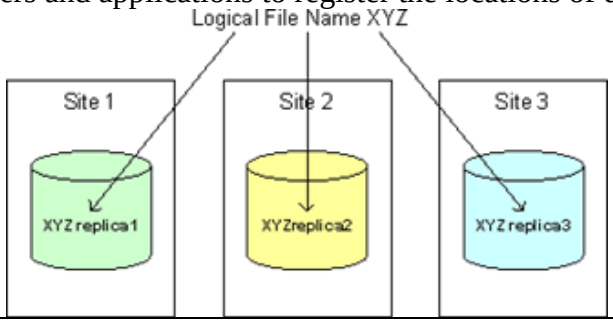
	contrast them (Nov/Dec 2017) (TB – Pg no: 141 – 144)		
13	With architecture, elaborate the various deployment models and reference models of cloud computing (Apr/May 2018) (TB – Pg no:192)	C403.3	BTL 2
14	"Virtualization is the wave of the future". Justify. Explicate the process of CPU, memory and I/O device virtualization in data center. (Apr/May 2018) (TB – Pg no:140)	C403.3	BTL 2

UNIT – IV - PROGRAMMING MODEL

Open source grid middleware packages – Globus Toolkit (GT4) Architecture , Configuration – Usage of Globus – Main components and Programming model - Introduction to Hadoop Framework - Mapreduce, Input splitting, map and reduce functions, specifying input and output parameters, configuring and running a job – Design of Hadoop file system, HDFS concepts, command line and java interface, dataflow of File read & File write.

PART A

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	**What are the functionalities of grid middleware?** Grid middleware is a specific software product, which enables the sharing of heterogeneous resources, and Virtual Organizations. It is installed and integrated into the existing infrastructure of the involved company or companies, and provides a special layer placed among the heterogeneous infrastructure and the specific user applications. Middleware glues the allocated resources with specific user applications. Major grid middlewares are Globus Toolkit, gLite, UNICORE, BONIC, CGSP, Condor-G and Sun Grid Engine etc.	C403.4	BTL 1
2	**Define Utility Computing?** Utility computing is referred to as the provision of grid computing and applications as service either as an open grid utility or as a hosting solution for one organization or a VO(Virtual Organization). Major players in the utility computing market are Sun Microsystems, IBM, and HP.	C403.4	BTL 1
3	**Write short notes on GT4?** The Globus Toolkit was initially motivated by a desire to remove obstacles that prevent seamless collaboration, and thus sharing of resources and services, in scientific and engineering applications. The toolkit addresses common problems and issues related to grid resource discovery, management, communication, security, fault detection, and portability.	C403.4	BTL 1
4	**List the functional modules in Globus GT4 library? / Services offered in GT4. (ND2016)** • **GRAM** (Global Resource Allocation Manager) – Grid resource access and management. • **Nexus** – used for unicast and multicast communication • **GSI** (Grid Security Infrastructure) – Used for Authentication and security. • **MDS** (Monitory and Discovery Service) – Distributed acces to structure and state information. • **HBM** (Heart Beat Monitor) – monitor heart beat of system components. • **GASS** (Global Access of Secondary Storage) – Grid access of data in remote secondary storage. • **GridFTP** (Grid File Transfer) – used for inter-node fast file	C403.4	BTL 1

	transfer.		
5	**Define GridFTP?** • GridFTP is a high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks. • The GridFTP protocol is based on FTP, the highly-popular Internet file transfer protocol. • GridFTP adds additional features such as parallel data transfer, third party data transfer and striped data transfer.	C403.4	BTL 1
6	**List the functional layers of GSI?** • Authorization • Authentications • Delegation • Message Protection • Message Format	C403.4	BTL 1
7	**Explain the different types of GT4 Data management?** • Globus Toolkit 4 Data Management tools within the toolkit fall into either of two categories data replication and data movement • Data Replication consists of Replica Location Service (RLS) • Data Movement consists of GridFTP and Reliable File Transfer (RFT)	C403.4	BTL 1
8	**Explain data replication in GT4?** Replica Location Service provides the capability to track and maintain multiple locations of data across the grid. It is a distributed registry system that allows users and applications to register the locations of data. 	C403.4	BTL 1
9	**Write short notes RFT?** Reliable File Transfer (RFT) - A Web Services Resource Framework that schedules file transfers based on a set of criteria of when specific resources and bandwidth is available	C403.4	BTL 1
10	**List the security issues of Globus Toolkit?** • Has to cross administrative domains. • Need agreed mechanisms and standards. • Focus on Internet security mechanisms, modified to handle the special needs of Grid computing. • Distributed resources must be protected from unauthorized access	C403.4	BTL 1
11	**Write short notes on distributed file system?** When a dataset outgrows the storage capacity of a single physical machine, it becomes necessary to partition it across a number of separate machines.	C403.4	BTL 1

	File systems that manage the storage across a network of machines are called *distributed file systems*		
12	**Write short notes on HDFS?** HDFS is a file system that is designed for use for MapReduce jobs that read input in large chunks of input, process it, and write potentially large chunks of output. HDFS does not handle random access particularly well. For reliability, file data is simply mirrored to multiple storage nodes. This is referred to as *replication* in the Hadoop community.	C403.4	BTL 1
13	**What are the advantages of using Hadoop? (ND 2016)** • Scalable • Flexible • Fast • Resilient to Failure • Independent	C403.4	BTL 1
14	**List the areas where HDFS is not fit for use?** • ***Low-latency data access*** - Applications that require low-latency access to data, in the tens of milliseconds range, will not work well with HDFS. • ***Lots of small files*** - the limit to the number of files in a filesystem is governed by the amount of memory on the namenode • ***Multiple writers, arbitrary file modifications*** - no support for multiple writers, or for modifications at arbitrary offsets in the file.	C403.4	BTL 1
15	**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.	C403.4	BTL 1
16	**What is the role of namenode in HDFS?** The namenode (the master) 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.	C403.4	BTL 1
17	**What is the role of datanode in HDFS?** Datanodes (workers) are the workhorses 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.	C403.4	BTL 1
18	**Write down the instructions for setting up Hadoop in pseudo-distributed mode?** • **fs.default.name**, set to *hdfs://localhost/*, which is used to set a default filesystem for Hadoop. Filesystems are specified by a URI, and here we have used an hdfs URI to configure Hadoop to use HDFS by default. The HDFS daemons will use this property to determine the host and port for the HDFS namenode. • **dfs.replication**, to 1 so that HDFS doesn't replicate filesystem	C403.4	BTL 1

	blocks by the default factor of three. When running with a single datanode, HDFS can't replicate blocks to three datanodes, so it would perpetually warn about blocks being under-replicated. This setting solves that problem.		
19	**Explain how data can be read from Hadoop URL using java?** Files can be read from a Hadoop filesystem by using a java.net.URL object to open a stream to read the data from. The general idiom is: *InputStream in = null;* *try {* *in = new URL("hdfs://host/path").openStream();* *// process in* *} finally {* *IOUtils.closeStream(in);* *}* Java recognize Hadoop's hdfs URL by calling the **setURLStreamHandlerFactory** method on URL with an instance of **FsUrlStreamHandlerFactory**.	C403.4	BTL 1
20	**Explain how data can be written in Hadoop file system?** The FileSystem class has a number of methods for creating a file. The simplest is the method that takes a **Path** object for the file to be created and returns an output stream to write to: *public FSDataOutputStream create(Path f) throws IOException* As an alternative to creating a new file, you can append to an existing file using the **append()** method : *public FSDataOutputStream append(Path f) throws IOException*	C403.4	BTL 1
21	**Explain how an application can be notified after a data being written to datanode?** *package org.apache.hadoop.util;* *public interface Progressable {* *public void progress();* *}*	C403.4	BTL 1
22	**How to create a File System directory using java?** FileSystem provides a method to create a directory: *public boolean mkdirs(Path f) throws IOException* This method creates all of the necessary parent directories if they don't already exist. It returns true if the directory (and allparent directories) was (were) successfully created	C403.4	BTL 1
23	**Explain how file or directory location can be retrieved in File System?** The ***FileStatus*** class encapsulates filesystem metadata for files and directories, including file length, block size, replication, modification time, ownership, and permission information. The method **getFileStatus()** on FileSystem provides a way of getting a **FileStatus** object for a single file or directory. If no file or directory exists, a **FileNotFoundException** is thrown. To find the existence of a file or directory, the **exists()** method on FileSystem is used: *public boolean exists(Path f) throws IOException*	C403.4	BTL 1

24	**Write the syntax for deleting a file or directory in 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).	C403.4	BTL 1
25	**Explain how a data can be made persistence and visible to all readers?** HDFS provides a method for forcing all buffers to be synchronized to the datanodes via the **sync()** method on **FSDataOutputStream**. After a successful return from sync(), HDFS guarantees that the data written up to that point in the file is persisted and visible to all new readers. *Path p = new Path("p");* *FSDataOutputStream out = fs.create(p);* *out.write("content".getBytes("UTF-8"));* *out.flush();* ***out.sync();*** *assertThat(fs.getFileStatus(p).getLen(), is(((long) "content".length())));*	C403.4	BTL 1
26	**Write short notes on MapReduce?** MapReduce model was introduced by Google as a method of solving a class of petascale problems with large clusters of inexpensive machines. The model is based on two distinct steps for an application: • *Map*: An initial ingestion and transformation step, in which individual input records can be processed in parallel. • *Reduce*: An aggregation or summarization step, in which all associated records must be processed together by a single entity.	C403.4	BTL 1
27	**Write short notes on Hadoop?** Hadoop is the Apache Software Foundation top-level project that holds the various Hadoop subprojects that graduated from the Apache Incubator. The Hadoop project provides and supports the development of open source software that supplies a framework for the development of highly scalable distributed computing applications. The Hadoop framework handles the processing details, leaving developers free to focus on application logic.	C403.4	BTL 1
28	**Explain Input Splitting in Mapreduce?** For the framework to be able to distribute pieces of the job to multiple machines, it needs to fragment the input into individual pieces, which can in turn be provided as input to the individual distributed tasks. Each fragment of input is called an *input split*.	C403.4	BTL 1
29	**Write short notes on IdentityMapper?** It is used in jobs that only need to reduce the input, and not transform the raw input. *public class IdentityMapper<K, V>extends MapReduceBase implements Mapper<K, V, K, V> {* ***public void map(K key, V val,*** ***OutputCollector<K, V> output, Reporter reporter)throws IOException {*** ***output.collect(key, val);***	C403.4	BTL 1

	}} The line output.collect(key, val), which passes a key/value pair back to the framework for further processing.		
30	**Write short notes on IdentityReducer?** The Hadoop framework calls the reduce function one time for each unique key. The framework provides the key and the set of values that share that key. IdentityReducer produces one output record for every value. *public class IdentityReducer<K, V>extends MapReduceBase implements Reducer<K, V, K, V> {* ***public void reduce(K key, Iterator<V> values, OutputCollector<K, V> output, Reporter reporter) throws IOException {*** ***while (values.hasNext()) {*** ***output.collect(key, values.next());*** ***}}*** The line output.collect() writes all keys and values directly to output	C403.4	BTL 1
31	**List the available input formats in Hadoop framework?** • **KeyValueTextInputFormat:** Key/value pairs, one per line. • **TextInputFormant:** The key is the line number, and the value is the line. • **NLineInputFormat:** Similar to KeyValueTextInputFormat, but the splits are based on *N* lines of input rather than *Y* bytes of input. • **MultiFileInputFormat:** An abstract class that lets the user implement an input format that aggregates multiple files into one split. • **SequenceFIleInputFormat:** The input file is a Hadoop sequence file, containing serialized key/value pairs.	C403.4	BTL 1
32	**Explain how to configuration the output of a MapReduce job?** *FileOutputFormat.setOutputPath(conf, MapReduceIntroConfig.getOutputDirectory());* *conf.setOutputKeyClass(Text.class);* *conf.setOutputValueClass(Text.class);* The conf.setOutputKeyClass(Text.class) and conf.setOutputValueClass(Text.class) settings inform the framework of the types of the key/value pairs to expect for the reduce phase.	C403.4	BTL 1
33	**What are the information required to configure the reduce phase?** • The number of reduce tasks; if zero, no reduce phase is run • The class supplying the reduce method • The input key and value types for the reduce task; by default, the same as the reduce output • The output key and value types for the reduce task • The output file type for the reduce task output	C403.4	BTL 1
34	**How to run a job in MapReduce?** *logger .info("Launching the job.");* ***final RunningJob job = JobClient.runJob(conf);*** *logger.info("The job has completed.");*	C403.4	BTL 1

	The method **runJob()** submits the configuration information to the framework and waits for the framework to finish running the job. The response is provided in the job object		
35	**Write the significant of GRAM (Apr/May 2017)** The Globus Toolkit includes a set of service components collectively referred to as the Globus Resource Allocation Manager (GRAM). GRAM simplifies the use of remote systems by providing a single standard interface for requesting and using remote system resources for the execution of "jobs". The most common use (and the best supported use) of GRAM is remote job submission and control. This is typically used to support distributed computing applications. For most Grid-based projects, we recommend using GRAM as a project-wide standard for remote job submission and resource management. GRAM is designed to provide a single common protocol and API for requesting and using remote system resources, by providing a uniform, flexible interface to, local job scheduling systems.	C403.4	BTL 1
36	**Name the different modules in hadoop framework (Apr/May 2017)** **The Apache Hadoop Module:** **Hadoop Common:** this includes the common utilities that support the other Hadoop modules **HDFS:** the Hadoop Distributed File System provides unrestricted, high-speed access to the application data. **Hadoop YARN:** this technology accomplishes scheduling of job and efficient management of the cluster resource. **MapReduce:** highly efficient methodology for parallel processing of huge volumes of data.	C403.4	BTL 1
37	**"HDFS is fault tolerant. Is it true? Justify your answer. (Nov/Dec 2017)** HDFS is highly fault tolerant. It handles faults by the process of replica creation. The replica of users data is created on different machines in the HDFS cluster. So whenever if any machine in the cluster goes down, then data can be accessed from other machine in which same copy of data was created. HDFS also maintains the replication factor by creating replica of data on other available machines in the cluster if suddenly one machine fails.	C403.4	BTL 1
38	**What is the purpose of heart beat in hadoop (Nov/Dec 2017)** In **Hadoop**, Namenode and Datanode are two physically separated machines, therefore **Heartbeat** is the signal that is sent by the datanode to the namenode after the regular interval to time to indicate its presence, i.e. to indicate that it is alive. • In case Namenode does not receive the heartbeat from a Datanode in a certain amount of time(within 10 mins), Namenode then considers that datanode as a dead machine. • Datanode along with heartbeat also sends the block report to Namenode, block report typically contains the list of all the blocks on a datanode.	C403.4	BTL 1

39	**How does divide and conquer strategy relates to map reduce paradigm? (Apr/May 2018)** In MapReduce, you divide the work up serially, execute work packets in parallel, and tag the results to indicate which results go with which other results. The merging is then serial for all the results with the same tag, but can be executed in parallel for results that have different tags. In more previous systems, the merge step became a bottleneck for all but the most truly trivial tasks. With MapReduce it *can* still be if the nature of the tasks requires that all merging be done serially. If, however, the task allows some degree of parallel merging of results, then MapReduce gives a simple way to take advantage of that possibility. Most other systems do one of two things: either execute all the merging serially just because it *might* be necessary for some tasks, or else statically define the parallel merging for a particular task. MapReduce gives you enough data at the merging step to automatically schedule as much in parallel as possible, while still ensuring (assuming you haven't made mistakes in the mapping step) that coherency is maintained.	C403.4	BTL 1
40	**Brief out the main components of Globus toolkit (Apr/May 2018)** • Common runtime components • Security • Data management • Information services • Execution management	C403.4	BTL 1

PART B

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	Explain in detail about Globus Toolkit GT4? (T1: Pgs 446-450) **(ND2016)**	C403.4	BTL 1
2	Give a detailed note on Hadoop Framework. (Ref. Notes) **(ND2016)**	C403.4	BTL 1
3	Explain in detail about parts of Hadoop MapReduce job? (R1: Pgs 27-31)	C403.4	BTL 1
4	Explain in detail about map and reduce functions? (R1: Pgs 31-35)	C403.4	BTL 1
5	How to configure and run a job in Hadoop MapReduce? (R1: Pgs 36-55)	C403.4	BTL 1
6	Explain in detail about command line interface and java interface in HDFS? (R2: Pgs 45-46,51-62)	C403.4	BTL 1
7	Explain the anatomy of File Read and File Write? (R2: Pgs 62-69)	C403.4	BTL 1
8	Discuss Map reduce with suitable diagrams (Apr/May 2017) (R1: Pgs 31-35)	C403.4	BTL 1
9	Elaborate HDFS concepts with suitable illustrations (Apr/May 2017)	C403.4	BTL 1
10	Illustrate data flow in HDFS with file read/write operations with suitable diagrams (Nov/Dec 2017)	C403.4	BTL 1
11	What is GT4? Describe in detail the components of GT4 with a suitable diagram (Nov/Dec 2017) (T1: Pgs 446-450)	C403.4	BTL 1
12	List the characteristics of globus toolkit. With neat sketch describe the	C403.4	BTL 1

	architecture of globus GT4 and the services offered (Apr/May 2018) (T1: Pgs 446-450)		
13	With an illustration, Emphasize the significance of map reduce paradigm in Hadoop framework. List out the assumptions and goal sets in HDFS architecture for processing the data based on divide and conquer strategy (Apr/May 2018) (R1: Pgs 31-35)	C403.4	BTL 1

UNIT – V – SECURITY

Trust models for Grid security environment – Authentication and Authorization methods – Grid security infrastructure – Cloud Infrastructure security: network, host and application level – aspects of data security, provider data and its security, Identity and access management architecture, IAM practices in the cloud, SaaS, PaaS, IaaS availability in the cloud, Key privacy issues in the cloud.

PART A

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	**Discuss on the application and use of identity and access management. (ND2016)** **Identity management**, also known as **identity and access management** (**IAM**) is, in computer security, the security and business discipline that "enables the right individuals to access the right resources at the right times and for the right reasons". It addresses the need to ensure appropriate access to resources across increasingly heterogeneous technology environments and to meet increasingly rigorous compliance requirements.	C403.5	BTL 6
2	**Define Transport Layer Security (TLS). (ND2016)** Transport Layer Security (TLS) is a protocol that provides privacy and data integrity between two communicating applications. It's the most widely deployed security protocol used today, and is used for Web browsers and other applications that require data to be securely exchanged over a network, such as file transfers, VPN connections, instant messaging and voice over IP.	C403.5	BTL 1
3	**Define the goals of security** • Confidentiality: Data is only available to those who are authorized • Integrity: Data is not changed except by controlled processes • Availability: Data is available when required.	C403.5	BTL 1
4	**Define data integrity** Data integrity requires that no unauthorized users can change or modify the data concerned. For example, you want to broadcast a message to the public, which is definitely not confidential to anyone. You have to ensure the data integrity of your message from modification by unauthorized people. In this instance, you may have to stamp or add your signature to certify the message.	C403.5	BTL 1
5	**Mention the additional concerns that required in terms of availability** The term “availability” addresses the degree to which a system, sub-system or equipment is operable and in a usable state. Additional concerns deal more with people and their actions: Authentication: Ensuring that users are who they say they are; Authorization: Making a decision about who may access data or a service;	C403.5	BTL 1

	Assurance: Being confident that the security system functions correctly Non-repudiation: Ensuring that a user cannot deny an action; Auditability: Tracking what a user did to data or a service.		
6	**Define ACL** Access Control Lists (ACL) associated with files or directories. ACLs are files listing individuals authorized to login to an account (e.g. the UNIX .rhosts file), configuration files naming authorized users of a node and sometimes files read over the network.	C403.5	BTL 1
7	**Define delegation** Delegation is a means by which a user or process authorized to perform an operation can grant the authority to perform that operation to another process. Delegation can be used to implement distributed authorization	C403.5	BTL 1
8	**List out the use of Assurance mechanism** Assurance mechanisms allow the requester of a service to decide whether a candidate service provider meets the requesters' requirements for security, trustworthiness, reliability or other characteristics. Assurance mechanisms can be implemented through certificates	C403.5	BTL 1
9	**Define Nonrepudiation and Audiability** Nonrepudiation means that it can be verified that the sender and the recipient were, in fact, the parties who claimed to send or receive the message, respectively. Auditability is about keeping track of what is happening on a system. The idea is that if there is an intrusion, then the system operator can find out exactly what has been done and in whose name.	C403.5	BTL 1
10	**Define trust, reliability, privacy** Trust: People can justifiably rely on computer-based systems to perform critical functions securely, and on systems to process, store and communicate sensitive information securely; • Reliability: The system does what you want, when you want it to; • Privacy: Within certain limits, no one should know who you are or what you do.	C403.5	BTL 1
11	**List out the common goals that achieved using Cryptography** • Message confidentiality: Only an authorized recipient is able to extract the contents of a message from its encrypted form; • Message integrity: The recipient should be able to determine if the message has been altered during transmission; • Sender authentication: The recipient can identify the sender, and verify that the purported sender did send the message; • Sender non-repudiation: The sender cannot deny sending the message.	C403.5	BTL 1
12	**Define the use of Symmetric cryptosystems** Using symmetric (conventional) cryptosystems, data is transformed (encrypted) using an encrypted key and scrambled in such a way that it can only be unscrambled (decrypted) by a symmetric transformation using the same encryption key.	C403.5	BTL 1

13	**Define Data Encryption Standard and its two components** Data Encryption Standard (DES) DES consists of two components – an algorithm and a key. The DES algorithm involves a number of iterations of a simple transformation which uses both transposition and substitution techniques applied alternately. DES is a so-called private-key cipher; here data is encrypted and decrypted with the same key. Both sender and receiver must keep the key a secret from others.	C403.5	BTL 1
14	**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.	C403.5	BTL 1
15	**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.	C403.5	BTL 1
16	**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.	C403.5	BTL 1
17	**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.	C403.5	BTL 1
18	**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.	C403.5	BTL 1
19	**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.	C403.5	BTL 1
20	**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.	C403.5	BTL 1

21	**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.	C403.5	BTL 1
22	**What is the primary information of GSI** GSI authentication, a certificate includes four primary pieces of information: • A subject name, which identifies the person or object that the certificate represents; • The public key belonging to the subject; • The identity of a CA that has signed the certificate to certify that the public key and the identity both belong to the subject; • The digital signature of the named CA.	C403.5	BTL 1
23	**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.	C403.5	BTL 1
24	**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	C403.5	BTL 1
25	**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.	C403.5	BTL 1
26	**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.	C403.5	BTL 1
27	**Define Data remanence** Data remanence is the residual representation of data that has been in some	C403.5	BTL 1

	way nominally erased or removed.		
28	**What are the IAM processes operational activities.** • Provisioning • Credential and attribute management • Entitlement management • Compliance management • Identity federation management	C403.5	BTL 1
29	**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.	C403.5	BTL 1
30	**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	C403.5	BTL 1
31	**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.	C403.5	BTL 1
32	**List the IAM process business category** • User management • Authentication management • Authorization management • Access management • Data management and provisioning • Monitoring and auditing	C403.5	BTL 1
33	**What are the key components of IAM automation process?** • User Management, New Users • User Management, User Modifications • Authentication Management • Authorization Management	C403.5	BTL 1
34	**List out the key policy issues** Data security involves encrypting the data as well as ensuring that appropriate policies are enforced for data sharing. In addition, resource allocation and memory management algorithm s have to be secure. Finally, data mining techniques may be applicable for malware detection in the	C403.5	BTL 1

	clouds – an approach which is usually adopted in intrusion detection systems (IDSs)		
35	**List out the six specific areas of the cloud computing environment** There are six specific areas of the cloud computing environment where equipment and software require substantial security attention These six areas are: (1) security of data at rest, (2) security of data in transit, (3) authentication of users/applications/ processes, (4) robust separation n between data belonging to different customers, (5) cloud legal and regulatory issues, and (6) incident response	C403.5	BTL 1
36	**Mention the issues in security of cloud computing** • The types of attackers and their capability of attacking the cloud. • The security risks associated with the cloud, and where relevant considerations of attacks and countermeasures. • Emerging cloud security risks	C403.5	BTL 1
37	**Define Network Level Security**. All data on the network need to be secured. Strong network traffic encryption techniques such as Secure Socket Layer (SSL) and the Transport Layer Security (TLS) can be used to prevent leakage of sensitive information. Several key security elements such as data security, data integrity, authentication and authorization, data confidentiality, web application security, virtualization vulnerability, availability, backup, and data breaches should be carefully considered to keep the cloud up and running continuously.	C403.5	BTL 1
38	**Define Application level security** Studies indicate that most websites are secured at the network level while there may be security loopholes at the application level which may allow information access to unauthorized users. Software and hardware resources can be used to provide security to applications.	C403.5	BTL 1
39	**Define Data Security** Majority of cloud service providers store customers' data on large data centres. Although cloud service providers say that data stored is secure and safe in the cloud, customers' data may be damaged during transition operations from or to the cloud storage provider.	C403.5	BTL 1
40	**List out the various advantages in Cloud computing architecture** Cloud computing architectures to its users numerous advantages that can be briefly summarized to: • Reduced cost since services are provided on demand with pay-as-you-use • billing system • Highly abstracted resources • Instant scalability and exibility • Instantaneous provisioning • Shared resources, such as hardware, database, etc.	C403.5	BTL 1

	• Programmatic management through API of Web services • Increased mobility - information is accessed from any location		
41	**Mention the foundational infrastructure requirements for cloud computing security** The foundational infrastructure for a cloud must be inherently secure whether it is a private or public cloud or whether the service is SAAS, PAAS or IAAS. It will require • Inherent component-level security • Stronger interface security • Resource lifecycle management	C403.5	BTL 1
42	**Mention the importance of transport level security. (Nov/Dec 2016)** Transport level security is based on Secure Sockets Layer (SSL) or Transport Layer Security (TLS) that runs beneath HTTP. SSL and TLS provide security features including authentication, data protection, and cryptographic token support for secure HTTP connections. To run with HTTPS, the service endpoint address must be in the form https://. The integrity and confidentiality of transport data, including SOAP messages and HTTP basic authentication, is confirmed when you use SSL and TLS. Web services applications can also use Federal Information Processing Standard (FIPS) approved ciphers for more secure TLS connections.	C403.5	BTL 1
43	**Discuss on application and use of identity and access management. (Nov/Dec 2016)** Identity management, also known as identity and access management (IAM) is, in computer security, the security and business discipline that "enables the right individuals to access the right resources at the right times and for the right reasons". It addresses the need to ensure appropriate access to resources across increasingly heterogeneous technology environments and to meet increasingly rigorous compliance requirements.	C403.5	BTL 1
44	**What are the various challenges in building trust environment? (Apr/May 2017)** • 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.	C403.5	BTL 1
45	**Write a brief note on security requirements of a grid. (Apr/May 2017)** ■ To protect application and data from the owner/administrator of the system ■ To protect local programs and data on the system on which another remote user's process may also be getting executed ■ Data, Code and resources accepted after proper *authentication* ■ *Integrity* of data and code is required to be verified.	C403.5	BTL 1
46	**List any four host security threads in public IaaS (ND2017)**	C403.5	BTL 1

	• Man in the middle attack • flooding attack • Data leakages		
47	**Identify the trust model based on site's trust worthiness (ND2017)** • A Generalized Trust Model • Reputation-Based Trust Model • A Fuzzy-Trust Model	C403.5	BTL 1

PART-B

S. No.	Question	Course Outcome	Blooms Taxonomy Level
1	Explain in detail about Trust models for Grid security environment. (TB – Pg no: 461 – 463) (ND2016)	C403.5	BTL 1
2	Briefly write a note on Authentication and Authorization methods Refer Notes	C403.5	BTL 1
3	Draw the neat architecture of Grid security infrastructure (TB – Pg no: 466 – 470) (ND2016)	C403.5	BTL 6
4	Explain the different level of Cloud Infrastructure security: network, host and application level Refer Notes	C403.5	BTL 1
5	Briefly discuss on Identity and access management architecture with neat architecture SaaS, PaaS, IaaS availability in the cloud, Refer Notes	C403.5	BTL 6
6	Illustrate the Key privacy issues in the cloud computing environment. Refer Notes	C403.5	BTL 2
7	Explain trust model for grid security environment (Nov/Dec 2016) (TB – Pg no: 461 – 463)	C403.5	BTL 1
8	Write in detail about cloud security infrastructures (Nov/Dec 2016)	C403.5	BTL 2
9	Write a detailed note on identity and access management architecture (Apr/May 2017) Refer Notes	C403.5	BTL 2
10	Explain grid security infrastructure (Apr/May 2017) (TB – Pg no: 466 – 470)	C403.5	BTL 1
11	What is the purpose of GSI? Describe the functionality of various layers in GSI. (ND2017) (TB – Pg no: 466 – 470)	C403.5	BTL 1
12	What is the purpose of IAM? Describe its functional architecture with an illustration. (ND2017) Refer Notes	C403.5	BTL 1