



**JEPPIAAR INSTITUTE OF TECHNOLOGY**

**Self Belief | Self Discipline | Self Respect**



## **QUESTION BANK**

**REGULATION :2013**

**YEAR : IV**

**SEMESTER : 07**

**BATCH : 2016-2020**

**DEPARTMENT**

**OF**

**INFORMATION TECHNOLOGY**



## **JEPPIAAR INSTITUTE OF TECHNOLOGY**

**“Self-Belief | Self Discipline | Self Respect”**



### **INSTITUTION VISION**

Jeppiaar Institute of Technology aspires to provide technical education in futuristic technologies with the perspective of innovative, industrial and social application for the betterment of humanity.

### **INSTITUTIONMISSION**

- To produce competent and disciplinedhigh quality professionals with the practical skills necessary to excel as innovative professionals and entrepreneurs for the benefit of the society.
- To improve the quality of education through excellence in teaching and learning, research, leadership and by promoting the principles of scientific analysis, and creative thinking.
- To provide excellent infrastructure, serene and stimulating environment that is most conducive to learning.
- To strive for productive partnership between the Industry and the Institute for research and development in the emerging fields and creating opportunities for employability.
- To serve the global community by instilling ethics, values and life skills among the students needed to enrich their lives.



## **JEPPIAAR INSTITUTE OF TECHNOLOGY**

**“Self-Belief | Self Discipline | Self Respect”**



### **DEPARTMENT VISION**

To facilitate the evolution of problem solving skills along with knowledge application in the field of Information Technology, understanding industrial and global requirements for the benefit of the society.

### **DEPARTMENT MISSION**

- To produce creative and productive computing graduates in software development being aware of global requirements and maximize employability.
- To enhance evolution of professional skills and development of leadership traits among the students to grow into successful entrepreneurs.
- To offer students an advantageous infrastructure to apply their research thoughts and develop their technical expertise .
- To escalate the moral code and honesty in the professional activities.

### **Program Educational Objectives (PEOs)**

**PEO1:** To provide students with a fundamental knowledge in Science, mathematics and computing skills for creative and innovative application.

**PEO2:** To enable students competent and employable by providing excellent Infrastructure to learn and contribute for the welfare of the society.

**PEO3:** To channelize the potentials of the students by offering state of the art amenities to undergo research and higher education.

**PEO4:** To evolve computing engineers with multi-disciplinary understanding and maximize Job Opportunities.

**PEO5:** To facilitate students obtain profound understanding nature and social requirements and grow as professionals with values and integrity.

### **Program Specific Outcomes (PSOs)**

**PSO 1:** To create the ability to analyze and enhance coding skills by participating in various competitions.

**PSO 2:** Students are able to provide solutions for Social Problems by creating Mobile Application Development using Android Studio and Chatbot.

**PSO 3:** Students are able to deal with real time problems using Machine Learning Tools and Big data Analytics.

## BLOOM'S TAXONOMY

### **Definition:**

**Bloom's taxonomy** is a classification system used to define and distinguish different levels of human cognition like thinking, learning, and understanding.

### **Objectives:**

- To classify educational learning objectives into levels of complexity and specificity.  
The classification covers the learning objectives in cognitive, affective and sensory domains.
- To structure curriculum learning objectives, assessments and activities.

### **Levels in Bloom's Taxonomy:**

- **BTL 1 – Remember** - The learner recalls, restate and remember the learned information.
- **BTL 2 – Understand** - The learner embraces the meaning of the information by interpreting and translating what has been learned.
- **BTL3–Apply**-The learner makes use of the information in a context similar to the one in which it was learned.
- **BTL 4 – Analyze** - The learner breaks the learned information into its parts to understand the information better.
- **BTL 5 – Evaluate** - The learner makes decisions based on in-depth reflection, criticism and assessment.
- **BTL6–Create**-The learner creates new ideas and information using what has been previously learned.

## TABLE OF CONTENTS

<b>IT6701 INFORMATION MANAGEMENT</b>		
<b>Unit No.</b>	<b>Topic</b>	<b>Page No.</b>
	<b>Syllabus</b>	<b>1.1</b>
<b>I</b>	<b>Database Modelling ,management and Development</b>	<b>1.2</b>
<b>II</b>	<b>Data Security and Privacy</b>	<b>1.11</b>
<b>III</b>	<b>Information Governance</b>	<b>1.20</b>
<b>IV</b>	<b>Information Architecture</b>	<b>1.27</b>
<b>V</b>	<b>Information Lifecycle Management</b>	<b>1.37</b>
<b>CS6701 CRYPTOGRAPHY AND NETWORK SECURITY</b>		
	<b>Syllabus</b>	<b>2.1</b>
<b>I</b>	<b>Introduction &amp; Number Theory</b>	<b>2.3</b>
<b>II</b>	<b>Block Ciphers and Public key Cryptography</b>	<b>2.17</b>
<b>III</b>	<b>Hash Functions and Digital Signatures</b>	<b>2.27</b>
<b>IV</b>	<b>Security practice and System Security</b>	<b>2.38</b>
<b>V</b>	<b>E mail ,IP &amp; Web security</b>	<b>2.53</b>
<b>IT6702 DATAWAREHOUSING AND MINING</b>		
	<b>Syllabus</b>	<b>3.1</b>
<b>I</b>	<b>Dataware housing</b>	<b>3.2</b>
<b>II</b>	<b>Business Modelling</b>	<b>3.8</b>
<b>III</b>	<b>Data Mining</b>	<b>3.16</b>
<b>IV</b>	<b>Association Rule Mining and Classification</b>	<b>3.25</b>
<b>V</b>	<b>Clustering and Trends in Datamining</b>	<b>3.34</b>
<b>CS6703 GRID AND CLOUD COMPUTING</b>		
	<b>Syllabus</b>	<b>4.1</b>
<b>I</b>	<b>Introduction</b>	<b>4.3</b>

<b>II</b>	<b>Grid Services</b>	<b>4.12</b>
<b>III</b>	<b>Virtualization</b>	<b>4.19</b>
<b>IV</b>	<b>Programming Models</b>	<b>4.27</b>
<b>V</b>	<b>Security</b>	<b>4.35</b>
<b>IT6004 SOFTWARE TESTING</b>		
	<b>Syllabus</b>	<b>5.1</b>
<b>I</b>	<b>Introduction</b>	<b>5.2</b>
<b>II</b>	<b>Test Case Design</b>	<b>5.10</b>
<b>III</b>	<b>Levels of Testing</b>	<b>5.16</b>
<b>IV</b>	<b>Test Management</b>	<b>5.24</b>
<b>V</b>	<b>Test Automation</b>	<b>5.30</b>

**IT6701 INFORMATION MANAGEMENT****UNIT I DATABASE MODELLING, MANAGEMENT AND DEVELOPMENT 9**

Database design and modelling - Business Rules and Relationship; Java database Connectivity (JDBC), Database connection Manager, Stored Procedures. Trends in Big Data systems including NoSQL - Hadoop HDFS, MapReduce, Hive, and enhancements.

**UNIT II DATA SECURITY AND PRIVACY 9**

Program Security, Malicious code and controls against threats; OS level protection; Security – Firewalls, Network Security Intrusion detection systems. Data Privacy principles. Data Privacy Laws and compliance.

**UNIT III INFORMATION GOVERNANCE 9**

Master Data Management (MDM) – Overview, Need for MDM, Privacy, regulatory requirements and compliance. Data Governance – Synchronization and data quality management.

**UNIT IV INFORMATION ARCHITECTURE 9**

Principles of Information architecture and framework, Organizing information, Navigation systems and Labelling systems, Conceptual design, Granularity of Content.

**UNIT V INFORMATION LIFECYCLE MANAGEMENT 9**

Data retention policies; Confidential and Sensitive data handling, lifecycle management costs. Archive data using Hadoop; Testing and delivering big data applications for performance and functionality; Challenges with data administration;

**TOTAL: 45 PERIODS****TEXT BOOKS:**

1. Alex Berson, Larry Duboy MASTER DATA MANAGEMENT AND DATA GOVERNANCE, 2/E, Tata McGraw Hill, 2011
2. Security in Computing, 4/E, Charles P. Pfleeger, Shari Lawrence Pfleeger, Prentice Hall; 2006
3. Information Architecture for the World Wide Web; Peter Morville, Louis Rosenfeld ; O'Reilly Media; 1998

**REFERENCES:**

1. Jeffrey A. Hoffer, HeikkiTopi, V Ramesh - MODERN DATABASE MANAGEMENT, 10 Edition, PEARSON, 2012
2. <http://nosql-database.org/> Next Gen databases that are distributed, open source and scalable.
3. <http://ibm.com/big-data> - Four dimensions of big data and other ebooks on Big Data Analytics
4. Inside Cyber Warfare: Mapping the Cyber Underworld- Jeffrey Carr, O'Reilly Media; Second Edition 2011.

### **UNIT-1 DATABASE MODELLING, MANAGEMENT AND DEVELOPMENT**

Database design and modelling - Business Rules and Relationship; Java database Connectivity (JDBC), Database connection Manager, Stored Procedures. Trends in Big Data systems including NoSQL - Hadoop HDFS, MapReduce, Hive, and enhancements.

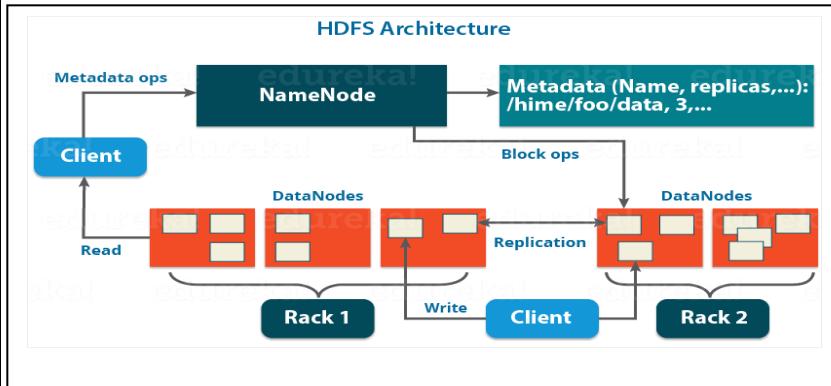
#### **PART \* A**

Q.N o	Questions
1	<p><b>Define data modeling.</b> BTL1</p> <p><i>Data modeling</i> is the analysis of data objects and their relationships to other data objects. It is often the first step in database design. Data modeling is a process used to define and analyze data requirements needed to support the business processes within the scope of corresponding information systems in organizations.</p>
2	<p><b>List the types of data models.</b> BTL1</p> <p>A <b>data model</b> (or <b>datamodel</b>) is an abstract model that organizes elements of data and standardizes how they relate to one another and to properties of the real world entities. For instance, a data model may specify that the data element representing a car be composed of a number of other elements which, in turn, represent the colour and size of the car and define its owner.</p> <ul style="list-style-type: none"> <li>■ Flat model</li> <li>■ Hierarchical model</li> <li>■ Network model</li> <li>■ Relational model</li> <li>■ Concept-oriented model</li> <li>■ Star schema</li> </ul>
3	<p><b>Define schemas.</b> BTL1</p> <p>The term "<i>schema</i>" refers to the organization of data as a blueprint of how the database is constructed (divided into database tables in the case of relational databases). The formal <i>definition</i> of a database <i>schema</i> is a set of formulas (sentences) called integrity constraints imposed on a database.</p>
4	<p><b>Where physical model is used?</b> BTL1</p> <p>A <b>physical data model</b> is used by <b>database</b> administrators to estimate the size of the <b>database</b> systems and to perform capacity planning. The <b>physical data model</b> constraints such as size, configuration and security can vary based on the underlying <b>database</b> system.</p>
5	<p><b>Name any two sources of Business rules.</b> BTL1</p> <ul style="list-style-type: none"> <li>■ Top management (policy makers) and managers</li> <li>■ Written documentation</li> </ul> <p style="text-align: center;">Procedures</p>

	<p style="text-align: center;">Standards Operations manuals ■ Direct interviews with end users</p>														
<b>Define Business Rules with an example.BTL1</b>	<ul style="list-style-type: none"> <li>■ Excellent tool to document various aspects of business domain.</li> <li>■ Descriptions of policies, procedures, or principles within a specific organization</li> <li>■ Use for describing characteristics of data</li> </ul> <p><b>Example:</b></p> <p>Department offers Course →</p> <p>Course generates Class →</p> <p>Professor teaches Class →</p>														
<b>Summarize the functionalities of JDBC.BTL2</b>	<ul style="list-style-type: none"> <li>■ allows multiple implementations to exist and is used by the same application</li> <li>■ API provides a mechanism for dynamically loading the correct Java packages and registering them with the JDBC Driver Manager</li> <li>■ JDBC connections support creating and executing statements</li> <li>■ JDBC connections are often managed via a connection pool rather than obtained directly from the driver.</li> </ul>														
<b>Describe the connection object .BTL2</b>	A wrapper around the actual database connection. The wrapper handles its relationship with the pool internally and hides the details of the pool from the application.														
<b>Differentiate between JDBC and ODBC. BTL2</b>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; padding: 5px;"><b>ODBC</b></th><th style="text-align: center; padding: 5px;"><b>JDBC</b></th></tr> </thead> <tbody> <tr> <td style="padding: 5px;">ODBC Stands for Open Database Connectivity.</td><td style="padding: 5px;">JDBC Stands for java database connectivity.</td></tr> <tr> <td style="padding: 5px;">Introduced by Microsoft in 1992.</td><td style="padding: 5px;">Introduced by SUN Micro Systems in 1997.</td></tr> <tr> <td style="padding: 5px;">We can use ODBC for any language like C,C++,Java etc.</td><td style="padding: 5px;">We can use JDBC only for Java languages.</td></tr> <tr> <td style="padding: 5px;">We can choose ODBC only windows platform.</td><td style="padding: 5px;">We can Use JDBC in any platform.</td></tr> <tr> <td style="padding: 5px;">Mostly ODBC Driver developed in native languages like C,C++.</td><td style="padding: 5px;">JDBC Stands for java database connectivity.</td></tr> <tr> <td style="padding: 5px;">For Java applications it is not recommended to use ODBC</td><td style="padding: 5px;">For Java application it is highly recommended to use JDBC</td></tr> </tbody> </table>	<b>ODBC</b>	<b>JDBC</b>	ODBC Stands for Open Database Connectivity.	JDBC Stands for java database connectivity.	Introduced by Microsoft in 1992.	Introduced by SUN Micro Systems in 1997.	We can use ODBC for any language like C,C++,Java etc.	We can use JDBC only for Java languages.	We can choose ODBC only windows platform.	We can Use JDBC in any platform.	Mostly ODBC Driver developed in native languages like C,C++.	JDBC Stands for java database connectivity.	For Java applications it is not recommended to use ODBC	For Java application it is highly recommended to use JDBC
<b>ODBC</b>	<b>JDBC</b>														
ODBC Stands for Open Database Connectivity.	JDBC Stands for java database connectivity.														
Introduced by Microsoft in 1992.	Introduced by SUN Micro Systems in 1997.														
We can use ODBC for any language like C,C++,Java etc.	We can use JDBC only for Java languages.														
We can choose ODBC only windows platform.	We can Use JDBC in any platform.														
Mostly ODBC Driver developed in native languages like C,C++.	JDBC Stands for java database connectivity.														
For Java applications it is not recommended to use ODBC	For Java application it is highly recommended to use JDBC														

	because performance will be down due to internal conversion and applications will become platform Dependent.	because there we no performance & platform dependent problem.	
	ODBC is procedural.	JDBC is object oriented.	
	It provides connection between front-end application(other than java) and back-end.	JDBC is used to provide connection between JAVA and database.	
<b>Discuss OLEDB.BTL2</b>			
10	Object Linking and Embedding Database (OLE DB) is a group of APIs used to facilitate and abstract access to application data of different file formats, including spreadsheets, structured query language (SQL)-based database management systems (DBMS), indexed-sequential files, and personal databases.  OLE DB is based on the Component Object Model (COM) and is part of the Microsoft Data Access Components (MDAC) software package, which is used to read and write data.		
11	<b>Illustrate stored procedure with an example.</b> BTL3  A <b>stored procedure</b> is a set of Structured Query Language (SQL) statements with an assigned name, which are <b>stored</b> in a relational database management system as a group, so it can be reused and shared by multiple programs.  CREATE PROCEDURE dbo.uspGetAddress @City nvarchar(30) AS SELECT * FROM Person.Address WHERE City = @City GO		
12	<b>Demonstrate ACID properties.</b> BTL3 <b>Atomicity</b> All changes to data are performed as if they are a single operation. <b>Consistency</b> Data is in a consistent state when a transaction starts and when it ends. <b>Isolation</b> The intermediate state of a transaction is invisible to other transactions. <b>Durability</b> After a transaction successfully completes, changes to data persist and are not undone, even in the event of a system failure.		
13	<b>Discover Map parameters.</b> BTL3 <b>Map Function</b> – It takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (Key-Value pair).  <b>Example</b> – (Map function in Word Count)		

	<b>Input</b>	Set of data	Bus, Car, bus, car, train, car, bus, car, train, bus, TRAIN,BUS, buS, caR, CAR, car, BUS, TRAIN	
	<b>Output</b>	Convert into another set of data  (Key,Value)	(Bus,1), (Car,1), (bus,1), (car,1), (train,1),  (car,1), (bus,1), (car,1), (train,1), (bus,1),  (TRAIN,1),(BUS,1), (buS,1), (caR,1), (CAR,1),  (car,1), (BUS,1), (TRAIN,1)	
14	<b>Analyse Hadoop as a Service.</b> BTL4	Hadoop as a service (HaaS), also known as Hadoop in the cloud, is a big data analytics framework that stores and analyzes data in the cloud using Hadoop		
15	<b>Point out any 2 features of Hadoop Cluster.</b> BTL4	<ul style="list-style-type: none"> <li>■ Highly scalable</li> <li>■ Commodity hardware based</li> </ul>		
16	<b>Compare Map Stage and Reduce stage.</b> BTL4	<p>Map Stage: The map stage uses a map() function that you must implement. When the map stage runs, it repeatedly calls the reader to get one input record at a time and applies the map() function to the record. Parameters: key ,value and context object</p> <p>Reduce Stage: The reduce stage uses a reduce() function that you must implement. When this stage executes, the reduce() function is called for each unique key in the shuffled intermediate data set. The reduce function takes a key and the list of values associated with that key and emits a new value based on the input Parameters: key,iterable collection of values, context object</p>		
17	<b>Explain the features of Hive.</b> BTL5	<ul style="list-style-type: none"> <li>■ It stores schema in a database and processed data into HDFS.</li> <li>■ It is designed for OLAP.</li> <li>■ It provides SQL type language for querying called HiveQL or HQL.</li> <li>■ It is familiar, fast, scalable, and extensible.</li> </ul>		
18	<b>Assess Hadoop Map Reduce.</b> BTL5	HadoopMapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. A MapReduce job usually splits the input data-set into independent chunks which are processed by the <i>map tasks</i> in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the <i>reduce tasks</i> .		
19	<b>Design HDFS architecture.</b> BTL6			



10<sup>6</sup>

#### **Prepare a query in HiveQL.BTL6**

20  
**SELECT [ALL | DISTINCT]select\_expr,select\_expr,...  
 FROM table\_reference  
 [WHERE where\_condition]  
 [GROUP BY col\_list]  
 [HAVING having\_condition]  
 [CLUSTER BY col\_list][DISTRIBUTE BY col\_list][SORT BY col\_list]]  
 [LIMIT number];**

#### **Illustrate Entity.BTL1**

An object in the real world that is distinguishable from other objects such as the musical instrument or concert.

#### **Describe attributes.BTL1**

20 A property or description of an entity. A concert entity could have attributes describing the concert's name, performers, location, time etc.

#### **Write about domain.BTL1**

21 A set of possible values for an attribute

#### **Describe Relationship.BTL1**

22 An Association among two or more entities

#### **Define Entity Set.BTL1**

23 A collection of similar entities such as all of the musical instruments

#### **Write relationship set.BTL1**

24 A collection of similar relationships

#### **Discuss many to many relationship.BTL1**

25 A key constraint which indicates that many of one entity can be associated with many of another entity...eg: concert and artist(an artist can perform in many concert)

#### **Define one to many relationship.BTL1**

26 A key constraint which indicates that one entity can be associated with many of other entities. Eg: artist play with many instruments

#### **Write about participation constraint.BTL1**

27 Determines whether relationships must involve certain entities.

#### **State weak entity set.BTL1**

28 An entity that cannot be identified uniquely without considering some primary key attributes of another

	identifying owner entity.
29	<b>Write the about term foreign key.</b> BTL1 <b>Referral to primary key</b> of another table. Foreign key columns can only contain values that exist in primary key column that they refer to.
30	<b>Define primary key.</b> BTL1 One or more columns within table that together form a <b>unique combination of values</b> by which each record can be pointed out separately.
31	<b>Illustrate normalisation and its types.</b> BTL1 A process of refining the relational schema <b>Types:</b> First normal form Second normal form Third normal form Boyce – Codd normal form
32	<b>List characteristic of big data.</b> BTL2 Volume Velocity Veracity
33	<b>Describe is hadoop ecosystem.</b> BTL2 Collection of tools to work along with key components Components: HDFS and Mapreduce
34	<b>Write three daemons in HDFS.</b> BTL2 Namenode Datanode Secondary namenode
35	<b>Outline name node.</b> BTL2 Maintains, manages and administers data blocks saved on slave machines and managed by datanodes.
36	<b>Define rack awareness.</b> BTL2 Important to prevent HDFS from placing all the copies of block in same rack which results in loss of data.
37	<b>Define NOSql Database environemt.</b> BTL1 It is a no relational and large distributed database system that enables fast, ad-hoc organization and analysis of extremely high volume, dissimilar data.
38	<b>Define hive.</b> BTL1 ■ Hive is an ETL and Data warehousing tool developed on top of Hadoop Distributed File System (HDFS) ■ Provides the SQL like interface for accessing data stored in hadoop.
39	<b>Illustrate Hadoop.</b> BTL1 <b>Hadoop</b> is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.
<b>PART * B</b>	
1	<b>Explain about the JDBC in detail.(13 M)</b> BTL1 - Register the driver class(3M) - Creating Connection(3M) - Creating statements(3M) - Executing Queries(3M) - Close Connection(1M)
2	<b>Explain the following SELECT statement with syntax. (13 M )</b> BTL1

	<p>Examples in HiveQL.</p> <p>i) Computing with Columns (2M)  <b>hive&gt; SELECT * FROM employee WHERE salary&gt;30000;</b></p> <p>ii) WHERE Clauses (3M)</p> <p>[WHERE where_condition]</p> <p>iii) GROUP BY Clauses(4M)</p> <p>[GROUP BY col_list]</p> <p>iv) HAVING Clauses (4M)</p> <p>[HAVING having_condition]</p> <p>[CLUSTER BY col_list   [DISTRIBUTE BY col_list] [SORT BY col_list]] [LIMIT number]</p>
3	<p><b>List the various data models in database design. (13 M)BTL1</b></p> <ul style="list-style-type: none"> <li>■ Database(1M)</li> <li>■ Database modelling(1M)</li> <li>■ ER Model <ul style="list-style-type: none"> <li>▪ Entity set</li> <li>▪ Attributes</li> </ul> </li> <li>■ Types of Attributes(2 M) <ul style="list-style-type: none"> <li>• Composite versus Simple Attribute</li> <li>• Single valued versus multivalued attribute</li> <li>• Stored versus derived attributes</li> <li>• Null values</li> <li>• Complex attributes</li> </ul> </li> <li>■ Relationship between the entities</li> <li>■ Cardinality ratio(1M)</li> <li>■ Participation constraint(1M) <ul style="list-style-type: none"> <li>■ Total/mandatory participation</li> <li>■ Partial/optional participation</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>■ ER to Relational Data Model(2M)           <ul style="list-style-type: none"> <li>■ The Mapping</li> <li>■ Type</li> <li>■ Diagram</li> <li>■ DDL</li> <li>■ DML</li> <li>■ Commands</li> </ul> </li>   <li>Normalization(3M)           <ul style="list-style-type: none"> <li>■ First Normal form</li> <li>■ Second Normal form</li> <li>■ Third Normal form</li> <li>■ Boyce-codd Normal form</li> </ul> </li> </ul>
4	<p><b>Explain Data organization in HDFS. (8M) BTL2</b></p> <ul style="list-style-type: none"> <li>-HDFS Architecture(2M)</li> <li>-Rack Awareness in HDFS(2M)</li> <li>-Preparing HDFS Writes(2M)</li> <li>-Reading Data from HDFS(2M)</li> </ul>
5	<p><b>Briefly explain limitations and Restrictions in Stored procedures.(13M) BTL2</b></p> <ul style="list-style-type: none"> <li>■ Limited Coding Functionality(2M)</li> <li>■ Portability (2M)</li> <li>■ Testing (2M)</li> <li>■ Location of Business Rules (2M)</li> <li>■ Utilization of Set-based Processing (1M)</li> <li>■ Cost (1M)</li> <li>■ Alternatives to Stored Procedures(1M)</li> <li>■ In-line or Parameterized Queries(1M)</li> <li>■ Object Relational Mapping (ORM)(1M)</li> </ul>
6	<p><b>Discuss MapR converged data platform.(13 M)BTL2</b></p> <ul style="list-style-type: none"> <li>Map Reduce(2M)           <ul style="list-style-type: none"> <li>-Reading the Data into MapReduce Program(2M)</li> <li>-MapReduce Flow(2M)</li> <li>-Mapping of physical and Logical division of data(2M)</li> <li>-Diagram of MapReduce Flow(2M)</li> <li>-Sample Working of the word count program(3M)</li> </ul> </li> </ul>

7	<p><b>Explain about Hadoop Ecosystem or architecture. (13 M)BTL2</b></p> <ul style="list-style-type: none"> <li>- open-source software framework (3M)</li> <li>- massive storage for any kind of data,</li> <li>- enormous processing power</li> <li>-the ability to handle virtually limitless concurrent tasks or jobs.</li> </ul> <p>-High level Hadoop Architecture(2M)</p> <ul style="list-style-type: none"> <li>- High level Hadoop 1.xArchitecture(2M)</li> </ul> <p>High level Hadoop 1.xArchitecture -Characteristics of(2M)</p> <p>Hadoop -Components of Hadoop Echo system -Diagram (2M)</p> <pre> graph TD     subgraph HADOOP_ECOYSTEM [HADOOP ECOSYSTEM]         direction TB         DS[HDFS Distributed file system] --- DP[Map Reduce Cluster Management]         DS --- YARN[YARN Cluster &amp; Resource Management]         DP --- DA[Hive Pig Mahout Avro Scoop]         DA --- DP         YARN --- DA         DS --- DS         DS --- DS     end     subgraph DataManagement [Data Management]         Oozie[Oozie Workflow Monitoring]         Chukwa[Chukwa Monitoring]         Flume[Flume Monitoring]         ZooKeeper[Zoo Keeper Management]     end     subgraph DataAccess [Data Access]         Hive[Hive SQL]         Pig[Pig Data Flow]         Mahout[Mahout Machine Learning]         Avro[Avro RPC, Serialization]         Scoop[Scoop RDBMS Connector]     end     subgraph DataProcessing [Data Processing]         MapReduce[Map Reduce Cluster Management]         YARN[YARN Cluster &amp; Resource Management]     end     subgraph DataStorage [Data Storage]         HDFS[HDFS Distributed file system]         HBase[HBase Column DB Storage]     end </pre> <p>Hadoop ecosystem(2M)</p>
8	<p><b>Analyze various databases used in NoSQL.(13M)BTL4</b></p> <ul style="list-style-type: none"> <li>-NoSQL means Not Only SQL, (1M)</li> <li>- No structure</li> <li>- Massive Database</li> </ul> <p><b>Key-Value databases(3M)</b></p> <p><b>Document databases(3M)</b></p> <p><b>Column family stores(3M)</b></p> <p><b>Graph Database(3M)</b></p>
9	<p><b>Explain about Hive Unit testing framework.(13M)BTL5</b></p> <ul style="list-style-type: none"> <li>- an ETL - Data warehousing tool (6M)</li> <li>-built on top of Hadoop.             <ul style="list-style-type: none"> <li>• Analysis of huge datasets</li> <li>• Ad-hoc queries</li> <li>• Data encapsulation</li> </ul> </li> </ul> <p>Components: (7M)</p> <p><b>Hive Clients</b></p> <p><b>Hive Services</b></p> <p><b>Processing framework and Resource Management</b></p> <p><b>Distributed Storage</b></p>

**PART \* C**

<b>1</b>	<b>Develop an example to show how to transform enhanced ER diagram into Relations. (15 M) BTL6</b> -ER Diagram(5M) -entity(2M) -Attribute(2M) -Diagram(3M) -ER Diagram to Relational Table(3M)
<b>2</b>	<b>Develop a program to establish Java Database connectivity.(15 M) BTL6</b> - JDBC (3M) - Steps:(9M) Diagram(3M)
<b>3</b>	<b>Difference between SQL and NOSQL.(15 M) BTL4</b> SQL(2M) Advantages(3M) Disadvantages(3M) NOSQL(2M) Advantages(3M) Disadvantages(2M)

JIT-2106

**SUBJECT Name: Information Management****Subject Handler:M. Dinesh Kumar****UNIT II - DATA SECURITY AND PRIVACY**

Program Security, Malicious code and controls against threats; OS level protection; Security – Firewalls, Network Security Intrusion detection systems. Data Privacy principles. Data Privacy Laws and compliance.

<b>Part A</b>	
<b>S. No</b>	<b>Question</b>
<b>1</b>	<b>Illustrate Session Hijacking.( NOV/DEC2016)BTL1</b> TCP session hijacking is a security attack on a user session over a protected network. The most common method of session hijacking is called IP spoofing, when an attacker uses source-routed IP packets to insert commands into an active communication between two nodes on a network and disguising itself as one of the authenticated users. This type of attack is possible because authentication typically is only done at the start of a TCP session.
<b>2</b>	<b>Give an example for Pseudonymity.(NOV/DEC 2016) BTL1</b> Facebook requires individuals to use their real names. On some forums and other venues, the user's pseudonym and IP address may be displayed along with location information, which can make their identity relatively easy to detect.
<b>3</b>	<b>Write in brief about secure program.BTL1</b> Security program implies some degree of trust that the program enforces expected confidentiality, integrity, and availability.
<b>4</b>	<b>State bug, error and fault.BTL1</b> ERROR: An error is a mistake, misconception, or misunderstanding on the part of a software developer. Error normally arises in software; it leads to change the functionality of the program. BUG: A bug is the result of a coding error. FAULT: An incorrect step, process or data definition in a computer program which causes the program to perform in an unintended or unanticipated manner.
<b>5</b>	<b>List the type of program security.BTL2</b> <ul style="list-style-type: none"> <li>• validation error (incomplete or inconsistent)</li> <li>• domain error</li> <li>• serialization and aliasing</li> <li>• inadequate identification and authentication</li> <li>• boundary condition violation</li> <li>• other exploitable logic errors</li> </ul>
<b>6</b>	<b>Categorize the program flaws. BTL1</b> <ul style="list-style-type: none"> <li>• validation error (incomplete or inconsistent): permission checks</li> <li>• domain error: controlled access to data</li> <li>• serialization and aliasing: program flow order</li> <li>• inadequate identification and authentication: basis for authorization</li> <li>• boundary condition violation: failure on first or last case</li> <li>• other exploitable logic errors</li> </ul>
<b>7</b>	<b>Write about malicious code attack.BTL2</b>

	Malicious code brings unanticipated or undesired effects in programs or program parts, by an agent intent to damage
<b>8</b>	<b>Define Transient virus.</b> BTL1 A Transient virus has a life that depends on the life of its hosts; the virus runs when its attached program executes and terminates when its attached program ends.
<b>9</b>	<b>State zero dayexploit.</b> BTL3 A zero-day vulnerability, also known as a computer zero day, is a flaw in software, hardware or firmware that is unknown to the party or parties responsible for patching or otherwise fixing the flaw. Zero day may refer to the vulnerability itself, or it may refer to a zero-day exploit, an attack that uses the zero-day vulnerability to attack vulnerable systems.
<b>10</b>	<b>List different controls against threats.</b> BTL2 <ul style="list-style-type: none"> <li>• Developmental</li> <li>• Operating system</li> <li>• Administrative</li> </ul>
<b>11</b>	<b>Define Trojan horse and Logic bomb, Trapdoor or backdoor, Worm and Rabbit.</b> BTL1 A <b>Trojan horse</b> is malicious code that, in addition to its primary effect, has a second, non obvious malicious effect A <b>logic bomb</b> is a class of malicious code that "detonates" or goes off when a specified condition occurs A <b>trapdoor or backdoor</b> is a feature in a program by which someone can access the program other than by the obvious, direct call, perhaps with special privileges A <b>worm</b> is a program that spreads copies of itself through a network. A <b>rabbit</b> as a virus or worm that self-replicates without bound, with the intention of exhausting some computing resource.
<b>12</b>	<b>Explain Appended Viruses and Documented Viruses.</b> BTL2 A program virus attaches itself to a program; then, whenever the program is run, the virus is activated. This kind of attachment is usually easy to program. Currently, the most popular virus type is what we call the document virus, which is implemented within a formatted document, such as a written document, a database, a slide presentation, a picture, or a spreadsheet
<b>13</b>	<b>Illustrate by Virus Signatures.</b> BTL3 A virus cannot be completely invisible. Code must be stored somewhere, and the code must be in memory to execute. Moreover, the virus executes in a particular way, using certain methods to spread. Each of these characteristics yields a telltale pattern, called a <b>signature</b> , that can be found by a program that knows to look for it. The virus's signature is important for creating a program, called a <b>virus scanner</b> , that can automatically detect and, in some cases, remove viruses.
<b>14</b>	<b>List the Operating System functions.</b> BTL1 <ul style="list-style-type: none"> <li>• Memory management</li> <li>• Task or process management</li> <li>• Storage management</li> <li>• Device or input/output management</li> <li>• Kernel or scheduling</li> </ul>
<b>15</b>	<b>Summarize the ways of separation in an operating system can occur.</b> BTL3 <ul style="list-style-type: none"> <li>• Physical separation;</li> </ul>

	<ul style="list-style-type: none"> <li>• Temporal separation;</li> <li>• Logical separation;</li> <li>• Cryptographic separation</li> </ul>
16	<b>Define Base/Bounds register.</b> BTL1 <p>Base Register is the register, acts as a address holder of the base storage location from where the data was stored continuously. Assembler uses the Base register value to find the data that is required.</p>
17	<b>Mention few kinds of objects for which protection are desirable.</b> BTL3 <ul style="list-style-type: none"> <li>• Memory</li> <li>• Sharable I/O devices, such as disks</li> <li>• Serially reusable I/O devices, such as printers and tape drives</li> <li>• Sharable programs and subprocedures</li> <li>• Networks</li> <li>• Sharable data</li> </ul>
18	<b>Discuss by Kerberos and mention the two systems required by it.</b> BTL1 <p>Kerberos is a protocol for authenticating service requests between trusted hosts across an untrusted network, such as the internet. Kerberos is built in to all major operating systems, including Microsoft Windows, Apple OS X, FreeBSD and Linux.</p>
19	<b>Define Firewalls. (APR /MAY 2017)</b> BTL1 <p>A firewall is a special form of reference monitor. By carefully positioning a firewall within a network, we can ensure that all network accesses that we want to control must pass through it</p>
20	<b>Write the types of firewalls.</b> BTL1 <ul style="list-style-type: none"> <li>• Packet filtering gateways or screening routers</li> <li>• Stateful inspection firewalls</li> <li>• Application proxies</li> <li>• Guards</li> <li>• Personal firewalls</li> </ul>
21	<b>Define Intrusion detection system.</b> BTL1 <p>Intrusion detection systems complement these preventive controls as the next line of defense. An intrusion detection system (IDS) is a device, typically another separate computer, that monitors activity to identify malicious or suspicious events</p>
22	<b>List the different types of IDS.</b> BTL2 <ul style="list-style-type: none"> <li>• Signature based</li> <li>• Heuristic</li> </ul>
23	<b>Define Privacy.</b> BTL1 <p>Privacy is a human right, although people can legitimately disagree over when or to what extent privacy is deserved; this disagreement may have cultural, historical, or personal roots.</p>
24	<b>List all the eight dimensions of privacy.</b> BTL2 <ul style="list-style-type: none"> <li>• Information collection</li> <li>• Information usage</li> <li>• Information retention</li> </ul>

	<ul style="list-style-type: none"> <li>• Information disclosure</li> <li>• Information security</li> <li>• Access control</li> <li>• Monitoring</li> <li>• Policy changes</li> </ul>
25	<p><b>Define Pseudonymity.BTL1</b></p> <p>Sometimes, full anonymity is not wanted. A person may want to order flower bulbs but not be placed on a dozen mailing lists for gardening supplies. But the person does want to be able to place similar orders again, asking for the same color tulips as before. This situation calls for pseudonyms, unique identifiers that can be used to link records in a server's database but that cannot be used to trace back to a real identity.</p>
26	<p><b>Write about Civil law.BTL1</b></p> <p>Civil law is the part of a country's set of laws which is concerned with the private affairs of citizens, for example marriage and property ownership, rather than with crime.</p>
27	<p><b>Mention the functions of IDS.BTL1</b></p> <ul style="list-style-type: none"> <li>• Monitoring users and system activity</li> <li>• Auditing system configuration for vulnerabilities and misconfigurations</li> <li>• Assessing the integrity of critical system and data files</li> <li>• Recognizing known attack patterns in system activity</li> <li>• Identifying abnormal activity through statistical analysis</li> </ul>
	<b>Part B</b>
1	<p><b>Illustrate in detail how to use controls during software development against the program threats. (NOV/DEC 2016) (13 M)</b> BTL2</p> <p><b>Answer: Page (299 -319)- Charles P. Pfleeger</b></p> <ul style="list-style-type: none"> <li>➢ <b>DEVELOPMENTAL CONTROLS(3M)</b></li> <li>➢ software development to ferret out and fix problems</li> <li>➢ <b>The Nature of Software Development(4M)</b> <ul style="list-style-type: none"> <li>• capturing the requirements</li> <li>• proposing a solution to the problem</li> <li>• the design as a blueprint</li> <li>• meets the requirements</li> <li>• to make sure that the end products are consistent</li> </ul> </li> <li>➢ <b>Modularity-</b> dividing a task into subtasks(3M)</li> <li>➢ <b>Encapsulation</b> - isolated from the effects of other components(1M)</li> <li>➢ <b>Information Hiding</b> - component hides its precise implementation(2M)</li> </ul>
2	<p><b>Examine the high protection mechanisms in general purpose operating systems. (APR / MAY 2017) ) (13M) BTL2</b></p> <p><b>Answer: Page (319 -330)- Charles P. Pfleeger</b></p> <p><b>Protected objects</b></p> <ul style="list-style-type: none"> <li>➢ Rise of Multiprogramming(4M) <ul style="list-style-type: none"> <li>• Memory</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>• Sharable i/o devices, such as disks</li> <li>• Serially reusable i/o devices, such as printers and tape drives</li> <li>• Sharable programs and subprocedures</li> <li>• Networks</li> <li>• Sharable data</li> </ul> <p>➤ <b>Security in operating system(4M)</b></p> <ul style="list-style-type: none"> <li>• user's objects separate from other</li> <li>• Physical separation</li> <li>• Temporal separation</li> <li>• Logical Separation</li> </ul> <p>➤ <b>Ways of Sharing and Separation(5M)</b></p> <ul style="list-style-type: none"> <li>• Do not protect.</li> <li>• Isolate.</li> <li>• Share all or share nothing.</li> <li>• Share via access limitation.</li> <li>• Share by capabilities.</li> <li>• Limit use of an object.</li> </ul>
3	<p><b>Write in brief about the malicious code and its types.(NOV/DEC 2016) (APR / MAY 2017) (13M) BTL2</b></p> <p><b>Answer:</b> Page (299 -319)- Charles P. Pfleeger</p> <ul style="list-style-type: none"> <li>➤ behaves in unexpected ways(2M)</li> <li>➤ could do anything to any programs(3M)</li> <li>➤ Trapdoor - undocumented entry point to a module(1M)</li> <li>➤ Kinds of Malicious code (7M) <ul style="list-style-type: none"> <li>• Virus -Attaches itself to program then propagates</li> <li>• Trojan horse- Contains unexpected, additional functionality</li> <li>• Logic bomb-Triggers action when condition occurs</li> <li>• Time bomb-Triggers action when specified time occurs</li> <li>• Trapdoor-Allows unauthorized access to functionality</li> <li>• Worm-Propagates copies of itself through a network</li> <li>• Rabbit-Replicates itself without limit to exhaust resource</li> </ul> </li> </ul>
4	<p><b>Discuss in detail about how to secure the programs. (13M)BTL6</b></p> <p><b>Answer:</b> Page (88-145)- Charles P. Pfleeger</p> <ul style="list-style-type: none"> <li>➤ Secure program(4M)</li> <li>➤ Types of Flaws(4M) <ul style="list-style-type: none"> <li>○ validation error</li> <li>○ domain error</li> <li>○ serialization and aliasing</li> <li>○ inadequate identification and authentication</li> <li>○ boundary condition violation:</li> <li>○ other exploitable logic errors</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>➤ Non Malicious Code(2M)</li> <li>➤ Buffer Overflows (3M)</li> </ul>
5	<p><b>Elaborate how the viruses get appended to the programs in detail. (13M)BTL2</b></p> <p><b>Answer:</b> Page (299 -319)- Charles P. Pfleeger</p> <ul style="list-style-type: none"> <li>➤ it must be activated by being executed(2M)</li> <li>➤ <b>Appended Viruses</b> - attaches itself to a program(2M)</li> <li>➤ <b>Viruses That Surround a Program</b> - that runs the original program(2M)</li> <li>➤ <b>Integrated Viruses and Replacements</b> - virus replaces some of its target(2M)</li> <li>➤ <b>Home of Virus</b>(5M) <ul style="list-style-type: none"> <li>• Hard to detect.</li> <li>• Not easily destroyed or deactivated.</li> <li>• Spreads infection widely.</li> <li>• Can reinfect its home program or other programs.</li> <li>• Easy to create.</li> <li>• Machine independent and operating system independent</li> </ul> </li> </ul>
6	<p><b>Discuss any two examples of malicious Code in detail.(13M)BTL6</b></p> <p><b>Answer:</b> Page (299 -319)- Charles P. Pfleeger</p> <ul style="list-style-type: none"> <li>➤ Anonymous code written to affect users(2M)</li> <li>➤ <b>Trapdoors</b> – Undocumented Entry point of module(1M)</li> <li>➤ <b>Causes of Trapdoors</b>- remove trapdoors during program development(1M)</li> <li>➤ <b>Salami Attack</b> - way odd bits of meat(1M)</li> <li>➤ <b>Rootkits and the Sony XCP</b>- goes to great lengths not to be discovered(1M)</li> <li>➤ <b>Privilege Escalation</b> - run in the context of the invoking user(1M)</li> <li>➤ <b>Interface Illusions</b> – spoofing attack(3M)</li> <li>➤ <b>Keystroke Logging</b>- retains a surreptitious copy of all keys pressed(1M)</li> <li>➤ <b>Man-in-the-Middle Attacks</b> - malicious program interjects itself between two other programs(1M)</li> <li>➤ <b>Timing Attacks</b> – Works faster than human(1M)</li> </ul>
7	<p><b>Describe about the Operating systems security methods and their role in computer security. (13M)BTL6</b></p> <p><b>Answer:</b> Page (319-320)- Charles P. Pfleeger</p> <ul style="list-style-type: none"> <li>➤ <b>Memory protection</b> <ul style="list-style-type: none"> <li>- control memory access rights(3M)</li> </ul> </li> <li>➤ <b>Segmentation</b> <ul style="list-style-type: none"> <li>- dividing a computer's memory(3M)</li> </ul> </li> <li>➤ <b>Paged virtual memory</b> <ul style="list-style-type: none"> <li>- divided into equal-sized blocks(3M)</li> </ul> </li> <li>➤ <b>Simulated segmentation</b> <ul style="list-style-type: none"> <li>- interpret the machine code instructions(4M)</li> </ul> </li> </ul>
8	<p><b>Write in detail about the Firewalls.(13M)BTL6.</b></p>

	<ul style="list-style-type: none"> <li>➤ filters all traffic (3M)</li> <li>➤ Design of Firewall(3M) <ul style="list-style-type: none"> <li>• Always invoked</li> <li>• Tamperproof</li> <li>• Small - simple enough for rigorous analysis</li> </ul> </li> <li>➤ Types of Firewall <ul style="list-style-type: none"> <li>• Packet filtering gateways (2M)</li> <li>• Stateful inspection firewalls (2M)</li> <li>• Application proxies (1M)</li> <li>• Guards (1M)</li> <li>• Personal firewalls(1M)</li> </ul> </li> </ul>
<b>Part * C</b>	
1	<p><b>Analyze in detail about the Intrusion detection Systems.(15 M)BTL6 (APR / MAY 2017)</b></p> <p><b>Answer: Page (380 -386)- Charles P. Pfleeger</b></p> <ul style="list-style-type: none"> <li>➤ A device(2M)</li> <li>➤ prevents malicious event(2M)</li> <li>➤ <b>Variety of functions(6M)</b> <ul style="list-style-type: none"> <li>• Monitoring users</li> <li>• system activity</li> <li>• Auditing system configuration for vulnerabilities - misconfigurations</li> <li>• Assessing the integrity of critical system</li> <li>• Recognizing known attack patterns in system activity</li> <li>• Identifying abnormal activity</li> </ul> </li> <li>➤ <b>Types (3M)</b> <ul style="list-style-type: none"> <li>• Signature based</li> <li>• Heuristic based</li> </ul> </li> </ul>
2	<p><b>Demonstrate about the laws and compliance for the information to be protected legally.(15 M)BTL6</b></p> <p><b>Answer: Page (641 -645)- Charles P. Pfleeger</b></p> <p><b>U.S. Privacy Laws(1M)</b></p> <p>-the 1974 Privacy Act (5 USC 552a),</p> <p><b>Federal Trade Commission (FTC) determined that in order to obey the Privacy Act(3M)</b></p> <ul style="list-style-type: none"> <li>• Notice</li> <li>• Choice</li> <li>• Access</li> <li>• Security</li> <li>• Enforcement</li> </ul> <p><b>e-Government Act of 2002(3M)</b></p> <p>- Control on government data collection through web sites</p> <p><b>Controls on Commercial Web Sites(2M)</b></p>

	<ul style="list-style-type: none"> <li>-deceptive trade or unfair business practices</li> </ul> <p><b>No Deceptive Practices(2M)</b></p> <ul style="list-style-type: none"> <li>- entities with whom the information</li> </ul> <p><b>Non-U.S. Privacy Principles(2M)</b></p> <ul style="list-style-type: none"> <li>-Special protection for sensitive data</li> </ul> <p>-Data Transfer</p> <ul style="list-style-type: none"> <li>-Independent Oversight</li> </ul> <p><b>Anonymity(2M)</b></p> <ul style="list-style-type: none"> <li>-want to do some things anonymously</li> </ul>

JIT-2106

### UNIT III INFORMATION GOVERNANCE

Master Data Management (MDM) – Overview, Need for MDM, Privacy, regulatory requirements and compliance.  
Data Governance – Synchronization and data quality management.

#### PART \*A

Q.N o	Questions
1	<b>What is Data Scrubbing?(NOV/DEC 2016)</b> BTL 1 Data scrubbing is the process of amending or removing data in a database that is incorrect, incomplete, improperly formatted, or duplicated.
2	<b>List the five areas to be considered for defining Compliance policies.(NOV/DEC 2016)</b> BTL 1 <ul style="list-style-type: none"> <li>• Retention</li> <li>• Immutability</li> <li>• Privacy</li> <li>• Auditing</li> <li>• Expiration</li> </ul>
3	<b>Illustrate the need of MDM. (APR /MAY 2017)</b> BTL 1 MDM claims that some entities (master entities) are more important than others because they are widely distributed across the enterprise as well as reside and are maintained in multiple systems and application silos.
4	<b>Define Customer Data Integration (CDI).</b> BTL 1 <b>Customer Data Integration (CDI)</b> is a Master Data Management framework focused on the customer data domain. It is a comprehensive set of technology components, services, and business processes that create, maintain, and make available an accurate, timely, integrated, and complete view of a customer across lines of business, channels, and business partners.
5	<b>Summarize the CDI predecessors include.</b> BTL 1 Customer Information File (CIF); Extract, Transform, and Load (ETL) technologies; Enterprise Data Warehouse (EDW); Operational Data Store (ODS); data quality (DQ) technologies; Enterprise Information Integration (EII); and Customer Relationship Management (CRM) systems.
6	<b>Define Customer Information File (CIF).</b> BTL 1 <b>Customer Information File (CIF)</b> is typically a legacy environment that represents some basic static information about the customers. CIF systems have a number of constraints including limited flexibility and extensibility; they are not well suited to capturing and maintaining real-time customer data, customer privacy preferences, customer behavior traits, and customer relationships. CIF systems are often used to feed the company's Customer Relationship Management systems.
7	<b>Briefly explain Extract, Transform and Load Technologies (ETL).</b> BTL 1 Extract, Transform, and Load (ETL) tools are designed to extract data from multiple data sources, perform complex transformations from source formats to the target formats, and efficiently load the transformed and formatted data into a target database such as CDI Data Hub.
8	<b>Define Operational Data Store (ODS).</b> BTL 1 <b>Operational Data Store (ODS)</b> is a data technology that allows transaction level detail data records to be stored in a non summarized form suitable for analysis and reporting.
9	<b>What is meant by Enterprise Information Integration (EII)?</b> BTL 1 <b>Enterprise Information Integration (EII)</b> tools are designed to aggregate distributed data in memory or nonpersistent storage, thus potentially delivering a "just-in-time" customer data view
10	<b>Define Customer Relationship Management systems (CRM).</b> BTL 1

	<b>Customer Relationship Management (CRM)</b> is a set of technologies and business processes designed to understand customers, improve customer experience, and optimize customer-facing business processes across marketing, sales, and servicing channels.
<b>11</b>	<b>Describe Data Quality Technologies (DQT).</b> BTL 1 <b>Data Quality Technologies (DQ)</b> , strictly speaking, are not customer data platforms, but they play an important role in making these platforms useful and useable whether they are built as data warehouses, operational data stores, or customer information files
<b>12</b>	<b>List the key benefits of MDM and CDI.</b> BTL 2 <ul style="list-style-type: none"> <li>• allows organizations to avoid costly penalties and bad publicity</li> <li>• helps to create new opportunities</li> <li>• A CDI Data Hub solution designed and deployed by a financial institution will most likely represent the authoritative source of customer personal and potentially financial data</li> </ul>
<b>13</b>	<b>Define Privacy.</b> BTL 1 Privacy is a human right, although people can legitimately disagree over when or to what extent privacy is deserved; this disagreement may have cultural, historical, or personal roots
<b>14</b>	<b>Describe Governance.</b> BTL 1 It is defined as “a process focused on managing data quality, consistency, usability, security, and availability of information.”
<b>15</b>	<b>Define Data Governance. (APR /MAY 2017)</b> BTL 1 Data Governance initiatives cross functional, organizational, and system boundaries and are enormously challenging in creating and aligning an effective and efficient program within the current organizational culture, programs, and maturity level
<b>16</b>	<b>Illustrate Enterprise Data Governance.</b> BTL 2 Enterprise Data governance (DG) is the overall management of the availability, usability, integrity and security of data used in an enterprise. A sound data governance program includes a governing body or council, a defined set of procedures and a plan to execute those procedures.
<b>17</b>	<b>List the two industry segments that have to deal with a customer’s privacy.</b> BTL 1 <ul style="list-style-type: none"> <li>• customer behaviour traits,</li> <li>• customer relationships</li> </ul>
<b>18</b>	<b>Mention few Regulatory Compliance Requirements.</b> BTL 1 <ul style="list-style-type: none"> <li>• The CEO/CFO has reviewed the report.</li> <li>• The report does not contain any untrue or misleading statement of a material fact or omit to state a material fact</li> </ul>
<b>19</b>	<b>Illustrate Data Quality.</b> BTL 1 Data should be relevant to their purposes, accurate, complete, and up-to-date.
<b>20</b>	<b>Describe MDM benchmark development.</b> BTL 1 The creation and maintenance of the data quality Benchmark Master—for example, a benchmark or high-quality authoritative source for customer, product, and location data
<b>21</b>	<b>Define MDM benchmark proliferation.</b> BTL 1 Proliferation of the benchmark data to other systems, which can occur through the interaction of the enterprise systems with the MDM Data Hub via messages, Web Service calls, API calls, or batch processing
<b>22</b>	<b>Summarize the two critical components of the quality improvement process.</b> BTL 1 <b>MDM benchmark development</b> MDM benchmark proliferation

	<b>Describe Master Data.</b> BTL 1
23	<b>Master data</b> are those entities, relationships, and attributes that are critical for an enterprise and foundational to key business processes and application systems.
24	<b>Define Master Data Management (MDM).</b> BTL 1 Master Data Management is very broad and may cover customer data, product data, supplier data, employee data, reference data, and other key types of data that should be used to consistently manage the entire enterprise in an integrated fashion
25	<b>What is Data synchronization?</b> BTL 1 Data synchronization is one of the most difficult and frequently underestimated areas of Master Data Management.
26	<b>Describe business rule.</b> BTL 1 The concept of business rules has been used in information systems for some time. There are many software products that help organizations manage their business rules (for example, JRules from ILOG, an IBM company).
27	<b>Define data quality management.</b> BTL 1 MDM Data Hub is an architectural concept and technology that is at the very core of most MDM implementations. At a high level, the MDM Data Hub's function is to resolve the most important master entities that are distributed across the enterprise, and to understand their intra- and inter-relationships.

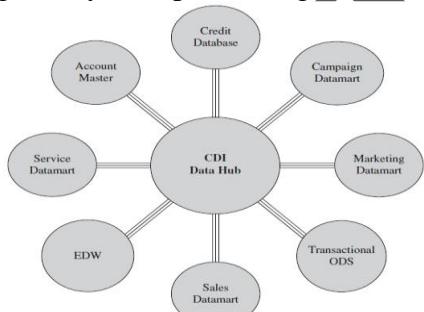
**PART \* B**

1 Explain in detail Regulatory Compliance Requirements and their impact on MDM IT Infrastructure. (13 M) (NOV/DEC 2016) BTL 2

**Answer:** Page( 271- 280)- Alex Berson

-the complex process of identifying, assessing, measuring, monitoring. Mitigate the full range of risks they face. (3 M)

-notion of customer risk Customer data protection, confidentiality, integrity risks –acts as subject to new regulatory, compliance legislation, industry-wide rules(3M)



(2M)

**Integrated Risk Management: Benefits and Challenges**(5 M)

- ability to provide accurate, verifiable, and consistent information
- ability to satisfy compliance requirements using clean, reliable, secure consistent data.
- ability to mitigate transaction risk associated with the data issues
- Flexibility in implementing, managing new organizational structures
- ability to define, implement, measure enterprise-wide
- ability to avoid delays related to data issues when delivering new products, services

2 Explain briefly about data quality improvement cycle and its activities with a neat diagram. (NOV/DEC

**2016) (13M)BTL 2**

**Answer: Page( 405 )- Alex Berson**

-develop sound – details materials on how to componentize Data Governance.(2M)

- Organizational Structure and Awareness.(4M)
- Stewardship
- Policy
- Value Generation
- Data Risk Management and Compliance
- Information Security and privacy
- Data Architecture
- Data Quality Management
- Classification and Metadata

**Levels:(7M)**

- **Level 1: Initial** Ad hoc operations that rely on individuals' knowledge, decision making.
- **Level 2: Managed** Projects - managed but lack cross-project, cross-organizational consistency, repeatability.
- **Level 3: Defined** Consistency in standards across projects – achieves organizational units.
- **Level 4: Quantitatively Managed** - sets quantitative quality goals leveraging statistical/quantitative techniques in the organization.
- **Level 5: Optimizing** Quantitative process improvement objectives - firmly established, continuously revised to manage process improvement.

**3 Explain about Enterprise architecture framework of MDM and its key challenges. (13M)(APR / MAY - 2017) BTL 2**

**Answer: Page( 5 -53 )- Alex Berson**

-Tremendous changes in the enterprise(2M)

-represents management challenges

-Consumption based approach

Stakeholders(8M)

1. Senior management
2. Sponsors
3. Line of business leaders
4. LOB professionals
5. Technology Managers
6. Architects
7. Finance
8. External partners
9. Legal Department
10. Security

• Need Careful analysis

• Candidate legacy- analysed

• MDM projects - shows technical challenges, risks to IT managers

Technical Challenges(3M)

1. Implementation cost - time to market

	<p>2. Data governance 3. Data quality, synchronization, integration and federation 4. Data visibility security regulatory.</p>
<b>4</b>	<p><b>Explain Data governance framework and its functions.(13 M) (APR / MAY -2017) BTL 2</b>  <b>Answer: Page( 399 – 423 )- Alex Berson</b>  a process on managing data quality, consistency, usability, security, availability of information(3 M)</p> <p><b>Mike2.0 Framework-</b> methodology concentrates on Information Development(3 M)</p> <p><b>The Data Governance Institute Framework-</b> recognizes a number of Data Governance. (3 M)</p> <p><b>The IBM Data Governance Council Framework and Maturity Model – deals with</b> componentize Data Governance (4M)</p>
<b>5</b>	<p><b>Discuss in detail the overview, need and the evolution of MDM.(13 M)BTL 6</b>  <b>Answer: Page( 5 -53 )- Alex Berson</b>  -historically striven to create, maintain authoritative, timely information sources.(2M)  -Includes the Sarbanes-Oxley Act and the Basel II Capital Accord. (2M)  -Evolved into a state with a wide variety of customer information stores and applications that manage customer data(2M)</p> <ul style="list-style-type: none"> <li>-Customer Information File (CIF);(1M)</li> <li>- Extract, Transform, and Load (ETL) technologies(1M)</li> <li>- Enterprise Data Warehouse (EDW)(1M)</li> <li>-Operational Data Store (ODS)(1M)</li> <li>-Data quality (DQ) technologies(1M)</li> <li>-Enterprise Information Integration (EII)(1M)</li> <li>-Customer Relationship Management (CRM) systems(1M)</li> </ul>
<b>6</b>	<p><b>Explain the Challenges Classification Dimensions and the key benefits of Master Data Management. (13M)BTL2</b>  <b>Answer: Page( 55 – 74)- Alex Berson</b></p> <p><b>Challenges(3M)</b></p> <ul style="list-style-type: none"> <li>• Significant, often require approaches, capabilities that exceed those available for mastering customer information</li> <li>• ability to match, link data records in order to find, integrate similar records</li> <li>• large part depends on the domain of data matching</li> </ul> <p><b>Classification Dimensions(5M)</b></p> <ul style="list-style-type: none"> <li>-include the Design and Deployment dimension, the Use Pattern dimension, and the Information Scope or Data Domain dimension</li> <li>-regardless of the industry or master data domain</li> <li>-cover most of the major differences between various MDM variants</li> </ul>

- **Design and Deployment**
- **Use Pattern**
- **Analytical MDM**
- **Operational MDM**
- **Collaborative MDM**
- **Information Scope or Data Domain**

#### **Key benefits of MDM (5M)**

- natural need to establish a single, authoritative, accurate, timely, secured master data system
- compliance - allows organizations to avoid costly penalties, bad publicity
- a single authoritative system of record positions the enterprise with gradual sunset in a number of legacy systems
- offers a number of critical capabilities
- Accurate, complete customer data
- contains information about prospects allows enterprises to increase their prospect-to-customer conversion ratio

**7 Explain in detail the Key Technical Implications of Data Security and Privacy Regulations on MDM Architecture.(13M)BTL2**

**Answer: Page( 195 -270 )- Alex Benson**

#### **Risk Taxonomy(2M)**

- the probability - a threat agent to exploit a defined vulnerability
- highly visible with the emergence of a broad set of government regulations, standards
  - **Transaction risk (2M)**
  - **Reputational risk (2M)**
  - **Strategic risk (2M)**
  - **Compliance (legal) risk (2M)**

-defined as policies, procedures, practices involved in identification, analysis, assessment, control, avoidance, minimization, or elimination of unacceptable risks

#### **Risk Analysis(3M)**

- brings together all the elements of risk management
- Quantitative risk analysis - uses the following variables to calculate the risk exposure

SLE - the single loss expectancy (expressed as the monetary value of the loss).

ARO - the annualized rate of occurrence.

ALE - the annualized loss expectancy

### **PART \* C**

**1 Explain in detail about Data synchronization. (15 M)BTL2**

**Answer:Page 367 – 395- Alex Berson**

### **DATA SYNCHRONIZATION(3M)**

- most difficult - frequently underestimated areas of Master Data Management.
- a typical implementation of an MDM synchronization solution in the context of the following use case.

Use Case: Delivering Customer Information to the Touch Point in Real Time(12 M)

#### *Real-Time/Near-Real-Time Synchronization Components*

- Legacy System Data Entry Validation Component
- Legacy System Message Creation and Canonical Message Format Component
- Legacy System Message-Processing Components
- Message Validation and Translations
- Transaction Manager and Transaction Logging Service
- Match Suspect Extractor
- Identity Store
- Change Capture
- Purge, Archival, and Audit Support
- Enterprise Record Locator
- Enterprise Attribute Locator
- Race Condition Controller

**2 Explain in detail the Data quality Management and its processes in MDM. (15M)BTL2**

**Answer: Page( 405 )- Alex Berson**

### **Data quality Management (5M)**

- an architectural concept - technology at the very core of most MDM implementations
- resolves the most important master entities distributed across the enterprise
- grows into powerful, multifaceted service-oriented toolkits
- match, link, merge, cleanse, validate, standardize, steward, transforms data
- accomplish by leveraging advanced algorithms and master data modeling patterns

### **Data Quality Processes(8M)**

#### **MDM benchmark development**

- creation, maintenance of the data quality Benchmark Master
- for example, a benchmark or high-quality authoritative source for customer, product, and location data.
- The MDM benchmark - includes the relationships between master entities.
- **MDM benchmark proliferation (2M)**
- the benchmark data to other systems
- interaction of the enterprise systems with the MDM Data Hub via messages, Web Service calls, API

**Subject Code:IT 6701**

**SUBJECT Name: Information Management**

**YEAR: / Sem : IV / 7**

**Subject Handler:M. Dinesh Kumar**

## **UNIT IV INFORMATION ARCHITECTURE**

Principles of Information architecture and framework, Organizing information, Navigation systems and Labelling systems, Conceptual design, Granularity of Content.

### PART \*A

Q.N o	Questions
1	<b>List out the components of an information architecture system. (April/May 2017) BTL1</b> <ul style="list-style-type: none"> <li>• organization systems</li> <li>• Navigation systems</li> <li>• Labeling systems</li> <li>• Searching systems</li> </ul>
2	<b>Mention the use of labeling system in information architecture system. (April/May 2017) BTL1</b> <p>The labeling system is used for representing thoughts and concepts on a website. The goal of labeling is to convey the meaningful information efficiently to the users without consuming much space. Labels are often used for representing organization and navigation systems.</p>
3	<b>What is information architecture? (Nov/Dec 2016) BTL1</b> <p>A foundation discipline describing the theory, principles, guidelines, standards, conventions and factors for managing information as a resource.</p> <p>The combination of organization, labeling and navigation schemes within an information system. The structural design of an information space to facilitate task completion and intuitive access to content.</p>
4	<b>Write the differences between library and web site. (Nov/Dec 2016) BTL1</b> <p>A Web site is a related collection of World Wide Web (WWW) files that includes a beginning file called a home Page:. A company or an individual tells you how to get to their Web site by giving you the address of their home Page:. From the home Page; you can get to all the other Page:s on their site.</p> <p>A library is a collection of sources of information and similar resources, made accessible to a defined community for reference or borrowing. It provides physical or digital access to material, and may be a physical building or room, or a virtual space, or both.</p>
5	<b>Write the different phases of information architecture development. BTL2</b> <ul style="list-style-type: none"> <li>• Research or analysis</li> <li>• Strategic planning</li> <li>• Conceptual design</li> <li>• Implementation</li> <li>• Administration</li> </ul>
6	<b>Define search system. BTL3</b> <p>The search system is another important component of information architecture that allows an user to search for specific contents over a website. The search engine are the basic foundation of a search system. They are basically software applications running on web servers that perform search based on user queries.</p>
7	<b>List the sources of labelling systems. BTL2</b> <p>There are two sources in labelling system. They are</p> <ol style="list-style-type: none"> <li>i) own site</li> <li>ii) competitive site</li> </ol>
8	<b>What are the types of navigation system? BTL1</b> <ul style="list-style-type: none"> <li>• Global navigation system</li> </ul>

	<ul style="list-style-type: none"> <li>• Local navigation system</li> <li>• Contextual navigation system</li> <li>• Supplemental navigation system</li> </ul>
9	<b>What are the different types of label? BTL1</b> <ul style="list-style-type: none"> <li>• Label as contextual links</li> <li>• Label as heading</li> <li>• Labels within navigation systems</li> <li>• Iconic labels</li> <li>• Labels as index terms</li> </ul>
10	<b>List the main organization structure.BTL2</b> <ul style="list-style-type: none"> <li>• Hierarchical or top down structure</li> <li>• Database or bottom up structure</li> <li>• Hypertext structure</li> </ul>
11	<b>Write any four responsibilities of information architecture.BTL1</b> <ul style="list-style-type: none"> <li>• Collect information through various sources such as emails, focus groups</li> <li>• Organize huge amounts of information on large websites and intranets so that people can accurately find what they are looking for.</li> <li>• Understand user goals and needs.</li> <li>• Understand business and organization's needs.</li> </ul>
12	<b>What are the three dimensions of information ecology? BTL1</b> <p>The three dimensions of information ecology are</p> <ul style="list-style-type: none"> <li>• Content (includes content objective, volume of contents, documents, data types, governance and ownership)</li> <li>• Context (includes business goals, funding, policies, technology, constraints and resources)</li> <li>• Users (includes audience, task, needs, experience and information seeking behavior)</li> </ul>
13	<b>Define organization system. BTL1</b> <p>The organization system is responsible for classifying the collected information in a correct manner for users. The information architect organizes the information such that people can search them easily, and they can find the right answer to their questions.</p>
14	<b>Define granularity of contents. BTL1</b> <p>The granularity of contents is related to the organization contents at different levels. Various levels of granularities in information architecture include journals, articles, paragraphs and sentences. Granularity deals with articulating the contents hierarchically at different levels according to certain criteria.</p>
15	<b>What are the classification organization systems? BTL1</b> <p>Two classification</p> <ul style="list-style-type: none"> <li>• Organization schemes</li> <li>• Organization structures</li> </ul>
16	<b>Define organization schemes. BTL1</b> <p>Organization scheme are related to organizing the information in a correct manner by categorizing the contents and making relationship between each pieces.</p>
17	<b>List the classifications of exact or objective organization schemes.BTL2</b>

	<ul style="list-style-type: none"> <li>• Alphabetical scheme</li> <li>• Chronological scheme</li> <li>• Geographical scheme</li> </ul>
18	<p><b>Summarize the classification of ambiguous or subjective organization schemes.</b> BTL2</p> <p>Topic scheme Task scheme Audience scheme Metaphor Scheme</p>
19	<p><b>Define organization structure.</b> BTL1</p> <p>Organization structure plays an important role in designing websites. It helps architects to define relationships between pieces and content. A successful organization structure allows users to predict the information they want on a particular site.</p>
20	<p><b>What are advanced approaches related to navigation system?</b> BTL1</p> <ul style="list-style-type: none"> <li>• Personalization and customizations</li> <li>• Visualization and social navigation</li> </ul>
21	<p><b>What is Architectural PageMockups?</b> BTL1</p> <p>Architectural Page: mockups are useful tools during conceptual design for complementing the blueprint view of the site. Mockups are quick and dirty textual documents that show the content and links of major Pages on the web site. They enable you to clearly (yet inexpensively) communicate the implications of the architecture at the Page level.</p>
22	<p><b>List challenges of organizing information.</b> BTL2</p> <ul style="list-style-type: none"> <li>• Ambiguity</li> <li>• Heterogeneity</li> </ul>
23	<p><b>Define Heterogeneity.</b> BTL1</p> <p>Heterogeneity refers to an object or collection of objects composed of unrelated or unlike parts. You might refer to grandma's homemade broth with its assortment of vegetables, meats, and other mysterious leftovers as heterogeneous.</p>
24	<p><b>List the sections which explore frequently used data.</b> BTL2</p> <p>The following sections explore three frequently used exact organization schemes.</p> <ul style="list-style-type: none"> <li>• Alphabetical</li> <li>• Chronological</li> <li>• Geographical</li> </ul>
25	<p><b>Define Metaphor.</b> BTL1</p> <p>Metaphors are commonly used to help users understand the new by relating it to the familiar. You need not look further than your desktop computer with its folders, files, and trash can or recycle bin for an example.</p>
<b>PART * B</b>	
1	<p><b>Discuss about how to Create New Labeling Systems. (13M)(NOV/DEC 2016)</b> BTL6</p> <p><b>Answer:</b> Page: 4.22 – B.S. Charulatha</p> <p><b>General Guidelines(4M)</b></p> <ul style="list-style-type: none"> <li>• Narrow scope whenever possible</li> <li>• Develop consistent labeling systems, not labels</li> </ul> <p><b>Sources of Labeling Systems(5M)</b></p>

	<ul style="list-style-type: none"> <li>• your site - create a table of the existing labels</li> <li>• comparable and competitive sites - find the labeling pattern</li> <li>• controlled vocabularies and thesauri - seek out focused vocabularies to help specific audience - like ERIC</li> <li>• content analysis - focus on things like titles, summaries, and abstracts</li> <li>• content authors - make their own suggestions</li> <li>• user advocates and subject matter experts - work with librarians and the like who can speak on behalf of the user - those who know what the users want</li> <li>• users - learn how the site's users will use the information - card sorting exercises where users are asked to cluster labels of existing content into their own categories and then label the categories or where they are given existing categories and asked to sort content into those categories</li> </ul>
	<p><b>Fine Tuning the Labels(4M)</b></p> <ul style="list-style-type: none"> <li>• Sort the list of terms alphabetically and remove duplicates</li> <li>• Review for consistency of usage, punctuation, letter case, etc.</li> <li>• Look for obvious gaps in the system - is the future considered?</li> <li>• Remember that you'll need to continually improve and work on your labeling system as users and content continue to change</li> </ul>
2	<p><b>Describe the challenges of Organizing Information. (13M)( (NOV/DEC 2016) BTL5</b></p> <p><b>Answer: Page: 4.11– B.S. Charulatha</b></p> <p><b>Ambiguity(5M)</b></p> <p>-built upon the foundation of language,</p> <p>-language - ambiguous</p> <ul style="list-style-type: none"> <li>• A throw, fling, or toss.</li> <li>• Used for waterproofing.</li> <li>• Rise and fall of the bow.</li> <li>• Salesman's persuasive line of talk.</li> <li>• Determined by the frequency of vibration.</li> </ul> <p><b>Heterogeneity(5M)</b></p> <p>-An object or collection of objects - composed of unrelated or unlike parts.</p> <p>-Assortment of vegetables, meats, other mysterious leftovers - heterogeneous.</p> <p>-End of the scale.</p> <p>For example, Ritz crackers - homogeneous.</p> <p><b>Differences in Perspectives(3M)</b></p> <ul style="list-style-type: none"> <li>- By striving to understand the intended audiences through user.</li> <li>-Provides multiple navigation pathways.</li> <li>- Organizing information for public consumption.</li> </ul>
3	<p><b>Explain in detail about Information Architecture Components.(13M)BTL2</b></p> <p><b>Answer: Page: 4.4– B.S. Charulatha</b></p> <ul style="list-style-type: none"> <li>• <b>4 components of Information Components (5M)</b></li> </ul>

**1. Organization systems**

- categorizing information, e.g., by subject or chronology.

**2. Labeling systems**

- Representation of information, e.g., scientific terminology ("Acer") or lay terminology ("maple").

**3. Navigation systems**

- browse or move through information, e.g., clicking through a hierarchy.

**4. Searching systems**

- search for information, e.g., executing a search query against an index.

- **EIGHT PRINCIPLES OF INFORMATION ARCHITECTURE** (8M)

**1. The principle of objects** – Treat content as a living, breathing thing, with a lifecycle, behaviors and attributes.

**2. The principle of choices** – Create Page:s that offer meaningful choices to users.

**3. The principle of disclosure** – Show only enough information to help people understand what kinds of information they'll find as they dig deeper.

**4. The principle of exemplars** – Describe the contents of categories with examples.

**5. The principle of front doors** – Assume users will visit some Page: than home Page:..

**6. The principle of multiple classifications** – Offer users several different classification schemes

**7. The principle of focused navigation** – Don't mix apples and oranges in your navigation scheme.

**8. The principle of growth** – Assume tomorrow content will have some today's content.

**4 Briefly discuss about Organization schemes.(APR/MAY 2017)(13M)BTL6**

**Answer: Page: 4.9– B.S. Charulatha**

- Used for navigation (4M)

The three major types:

1. Exact
2. Ambiguous
3. Hybrid

- **Exact Organization Schemes** (4M)

1. Divide information into well-defined and mutually exclusive sections
2. Example -organization of the phone book's white Page:s.
3. Easy to design and maintain.
  - -Three frequently used schemes
    1. Alphabetical
    2. Chronological
    3. Geographical

- **Ambiguous Organization Schemes** (3M)

1. Divide information into categories that defy exact definition
2. Mired in the ambiguity of language
3. More important and useful than exact organization schemes
4. Depends upon the quality of the scheme

- **Types**

1. Task
2. Topic
3. Metaphor
4. Audience

	<p><b>Hybrids Organization schemes(2M)</b></p> <ul style="list-style-type: none"> <li>- blending elements of multiple schemes</li> <li>- to skim through each menu item to find the option</li> </ul>
<b>5</b>	<p><b>Explain the Organization structure for designing the website. (13M)BTL 2</b></p> <p><b>Answer: Page: 4.9– B.S. Charulatha</b></p> <ul style="list-style-type: none"> <li>• Very important role in the design of web sites (4M)</li> <li>• Interact with organization structures every day</li> <li>• Experience physical structure frame by frame from beginning to end</li> <li>• Items are placed according to physical proximity</li> <li>• Possesses unique strengths and weaknesses</li> </ul> <p><b>Example:</b> Street structure in NYC vs. Paris</p> <p>The main organizational structures are:</p> <ul style="list-style-type: none"> <li>• Hierarchy</li> <li>• Hypertext</li> <li>• Database</li> </ul> <p><b>Hierarchical structure</b> (4M)</p> <p>A well-designed hierarchy forms good foundation for many web sites.</p> <ul style="list-style-type: none"> <li>• Mutually exclusive subdivisions and parent- child relationships</li> <li>• Users are familiar with hierarchies.</li> <li>• Hierarchy to develop a mental model of the site's structure and their locationwithin the site.</li> <li>• Takes a top-down approach.</li> </ul> <p><b>Designing hierarchies</b></p> <ul style="list-style-type: none"> <li>• Categories should be mutually exclusive.</li> <li>• Consider the balance between breadth and depth in the hierarchy. <ul style="list-style-type: none"> <li>• Breadth: Be sensitive to the cognitive limits of your user</li> <li>• Depth: forced to click through more than four or five levels</li> </ul> </li> </ul> <p><b>Hypertext structure(3M)</b></p> <ul style="list-style-type: none"> <li>• Content chunks are connected via links in a loose web of relationships.</li> </ul> <p><b>Features:</b></p> <ul style="list-style-type: none"> <li>• Great flexibility</li> <li>• Substantial potential for complexity and confusion</li> <li>• Rarely useful as the primary structure</li> <li>• Often used to complement structures based on the other two models.</li> </ul> <p>Example: Microsoft Help</p> <p><b>Database structure(2M)</b></p> <p>A database is a collection of records, each of which has a number of associated fields.</p> <p><b>Features:</b></p> <ul style="list-style-type: none"> <li>• Allows field-specific searching</li> <li>• Permits repackaging of information into different formats for different audiences</li> <li>• Records must follow rigid rules</li> <li>• Best for listings, catalogues, directories, and other subsites with structured, homogeneous data.</li> </ul>

	<p>Example: The Oracle of Bacon</p> <p><b>6 Define labelling systems. Discuss about varieties of labels in Information Architecture. (13M) (APR/MAY 2017) BTL1</b></p> <p><b>Answer: Page: 4.22– B.S. Charulatha</b></p> <ul style="list-style-type: none"> <li>-to represent the larger pieces of information present in our web site. (2M)</li> <li>-to communicate information efficiently.</li> <li>-should speak the same language as the users .</li> <li>-educate the user about new concepts.</li> </ul> <p><b>Varieties of Labels</b></p> <p><b>Contextual Links</b></p> <ul style="list-style-type: none"> <li>-describe the hypertext links within the body of the document and occur within the context of the surrounding text (3M)</li> <li>-easy to create, but hard to make work well - not developed systematically but in an ad hoc manner</li> <li>-they rely on context</li> <li>-ensure they are representational by asking the user where he/she expects to be taken</li> </ul> <p><b>Headings</b></p> <ul style="list-style-type: none"> <li>-describe the chunk of information that follows</li> <li>-establish a hierarchy and visual consistency with the text through numbering, font size, color, style, etc.</li> <li>-must be obvious and convey sequence (2M)</li> </ul> <p><b>Within Navigation Systems</b></p> <ul style="list-style-type: none"> <li>-require consistent application because they typically occur throughout the site</li> <li>-use to build a sense of familiarity</li> <li>-use scope notes to help users understand the label more clearly (2M)</li> </ul> <p><b>Index Terms</b></p> <ul style="list-style-type: none"> <li>-often called keywords, descriptive metadata, taxonomies, controlled vocabularies and thesauri</li> <li>-can be used to describe any type of content</li> <li>-support precise searching</li> <li>frequently visible to users - IUPUI Site Index - but can be hidden in the metadata tag of the HTML document (2M)</li> </ul> <p><b>Iconic Labels</b></p> <ul style="list-style-type: none"> <li>-most often used as navigation labels</li> <li>-more limited language than text - use with caution</li> </ul> <p><b>7 Explain the types of Navigation Systems. (13M)( APR/MAY 2017) BTL2</b></p> <p><b>Answer: Page: 4.32– B.S. Charulatha</b></p> <p>A good navigation system: (3M)</p> <ul style="list-style-type: none"> <li>• Helps exploration/orientation</li> <li>• Leads users to what they seek</li> <li>• Informs about the available products/services/tasks</li> </ul> <p style="padding-left: 20px;">Doing this requires anticipating the users' needs</p> <p style="padding-left: 20px;">✓ <a href="#">Amazon</a></p> <p><b>Built-in navigational features(2M)</b></p> <p>Most browsers offer built-in navigational features:</p> <ul style="list-style-type: none"> <li>• URL : direct access to any Page:</li> <li>• Back/forward : bi-directional backtracking</li> </ul>
	JIT-2106/IT/S.S.Vasantha Raja/IVYr/SEM07/IT6701/InformationManagement/UNIT1-5/QB+KEYS/Ver1.0

- History : random access to Page:s visited
- Bookmarks : save the location of Page:s visited
- Color coding of links : helps users understand where they have been and retrace their steps through a site
- Mouseover effects : may indicate site structure

### Purposes of navigation systems(2M)

Navigational systems can aid users by:

- Providing context: Users must have a good idea of where the Page: fits into the overall site.
  - Company logo
  - Page: title
  - Subsite/task indicators
  - Properly named links out of the site

Example: [DePaul CTI](#)

- Providing flexibility: Multiple means of navigation are important. At the very least, provide a link back to the main Page: for a site/subsite.

### Types of navigation systems(2M)

- Hierarchical: follows the information hierarchy closely.
- Global: Used for quick access of site.
- Local: Used in conjunction with a global system.
- Embedded links: Never used alone.

### Types of navigation elements(2M)

- Integrated: Integrated within the Page: and thus context-related.
  - Navigation bars (graphic or text)
  - Pull-down menus
  - Frames
- Remote: Complement other navigation systems
  - Table of contents
  - Index
  - Site map

### Building a navigation system(2M)

- Use the information hierarchy as the primary navigation system.
- Major categories are global navigation system.
- Local navigation – depended on global system choices
- Site's size and goal determines required navigation systems.
- Test your navigation on users.

## PART \* C

<b>1</b>	<b>How to connect and manage granular content within your site. (15M)(APR/MAY 2017) BTL 1</b>
	<b>Answer: Page: 4.20– B.S. Charulatha</b>

-Should not be independent (4M)

	<ul style="list-style-type: none"> <li>-Navigation can drown this content</li> <li>-Revision of options</li> <li>-Supplemental Navigation System(4M)</li> <li>-Site maps</li> <li>-Indexes</li> <li>-levels of granularity(4M)</li> <li>-Challenges in Indexing(3M)</li> <li>-index creation</li> <li>-Configuration</li> <li>-Search</li> </ul>
2	<p><b>Explain the following conceptual design concepts (i)Metaphor Exploration (ii)Scenario (iii)Architectural Page:Mockups (iv)Design Sketches.(15M) BTL 2</b></p> <p><b>Answer:</b> Page: 4.25– B.S. Charulatha</p> <p><b>(i)Metaphor Exploration (3M)</b></p> <ul style="list-style-type: none"> <li>– can be a powerful tool for communicating complex ideas</li> <li>– creative relationships or by mapping the familiar onto the new</li> <li>– mapped the familiar and respected metaphor of the physical highway infrastructure</li> </ul> <p>Three types of metaphor</p> <ol style="list-style-type: none"> <li>1. <b>Organizational metaphors</b> - organization to convey quickunderstanding of a new system's organization</li> <li>2. <b>Functional metaphors</b> - make a connection between the tasks</li> <li>3. <b>Visual metaphors</b>- leverage familiar graphic elements such as images, icons, and colors</li> </ol> <p><b>(ii)Scenario(3M)</b></p> <ul style="list-style-type: none"> <li>– scenarios that show how people with different needs and behaviors</li> <li>– think about the primary intended audiences</li> <li>– select three or four major user types</li> <li>– these scenarios to be easy and fun to write</li> </ul> <p><b>High-Level Architecture Blueprints(3M)</b></p> <ul style="list-style-type: none"> <li>– most useful for exploring primary organization schemes and approaches.</li> <li>– map out the organization and labeling of major areas</li> <li>– great for stimulatingdiscussions focused on the organization and management</li> </ul> <p><b>(iii)Architectural Page:Mockups(3M)</b></p> <ul style="list-style-type: none"> <li>– useful for presenting a bird's-eye view of the web site</li> <li>– helping people to envision the contents</li> <li>– used to show multiple ways</li> <li>– inexpensive</li> <li>– used in conjunction with scenarios</li> <li>– somebasic usability tests</li> </ul> <p><b>(iv)Design Sketches.(3M)</b></p> <ul style="list-style-type: none"> <li>– sketches on paper of major Page:s in the web site</li> <li>– To develop a sense of the desired graphic identity or look and feel</li> </ul>

- |  |   |
|--|---|
|  | <ul style="list-style-type: none"><li>– a great way to pool the collective knowledge of dynamic content</li><li>– <b>Web-Based Prototypes-</b> highly skilled graphic designer creates beautiful Web-based prototypes</li></ul> |
|--|---|

JIT-2106

**Subject Code:IT 6701**  
**SUBJECT Name: Information Management**

**YEAR: / Sem : IV / 7**  
**Subject Handler:M. Dinesh Kumar**

**UNIT V INFORMATION LIFECYCLE MANAGEMENT**

Data retention policies; Confidential and Sensitive data handling, lifecycle management costs. Archive data using Hadoop; Testing and delivering big data applications for performance and functionality; Challenges with data administration;

<b>PART *A</b>	
<b>Q.N o</b>	<b>Questions</b>
<b>1</b>	<p><b>What is data retention policy? BTL 1</b></p> <p>A data retention policy, or records retention policy, is an organization's established protocol for retaining information for operational or regulatory compliance needs.</p> <p>When writing a data retention policy, you need to determine how to:</p> <ul style="list-style-type: none"> <li>• Organize information so it can be searched and accessed at a later date</li> <li>• Dispose of information that is no longer needed</li> </ul>
<b>2</b>	<p><b>List the Issues involved in data retention.BTL 2</b></p> <ul style="list-style-type: none"> <li>• Knowing how long the data must be retained.</li> <li>• Another difficult issue involved in data retention is that of data storage and who is responsible for managing the storage.</li> </ul> <p>Finally, when researcher leave the institution, the institution and researcher should come to agreement over whether the researcher may take the original data or an identical copy of the data</p>
<b>3</b>	<p><b>Illustrate Restricted data and confidential data. BTL 2</b></p> <p><b>Restricted Data</b> - super sensitive information. Restricted data is "notice-triggering", meaning, we need to notify people if there has been unauthorized access or disclosure of this information. Leaks of this type of information can lead to identity theft, news coverage/publicity, and reputational damage and costs to the university.</p> <p>Examples: Social Security Number (SSN), driver's license/state ID numbers, financial account numbers, credit card numbers, personal medical and medical insurance information, and passwords.</p> <p><b>Confidential Data</b> - Moderately sensitive information. Not notice-triggering. This information needs to be protected from unauthorized access.</p> <p>Examples: home address and phone, birth date.</p>
<b>4</b>	<p><b>What is the difference between sensitive information and confidential information? BTL 1</b></p> <ul style="list-style-type: none"> <li>• Sensitive Data - institutional data that is not legally protected, but should not be made public and should only be disclosed under limited circumstances. Users must be granted specific authorization to access since the data's unauthorized disclosure, alteration, or destruction may cause perceivable damage to the institution.</li> </ul> <p>Confidential/Regulated Data - institutional data for which there is a legal obligation not to disclose. These data elements require the highest levels of restriction due to the risk or harm that will result from disclosure or inappropriate use</p>
<b>5</b>	<p><b>What is Information life cycle management (ILM)? BTL 1</b></p> <p>ILM is a comprehensive approach to managing the flow of an information system's data and associated metadata from creation and initial storage to the time when it becomes obsolete and is deleted.</p>
<b>6</b>	<p><b>Write the phases involved in life cycle management.BTL 2</b></p> <ul style="list-style-type: none"> <li>• Creation and Receipt: deals with records from their point of origination.</li> </ul>

	<ul style="list-style-type: none"> <li>• Distribution: is the process of managing the information once it has been created or received. This includes both internal and external distribution</li> <li>• Use: takes place after information is distributed internally, and can generate business decisions, document further actions, or serve other purposes.</li> <li>• Maintenance: is the management of information. This can include processes such as filing, retrieval and transfers.</li> <li>• Disposition: is the practice of handling information that is less frequently accessed or has met its assigned retention periods.</li> </ul>
7	<b>Define data archiving. BTL 1</b> Data archiving is the process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that is still important to the organization and may be needed for future reference, as well as data that must be retained for regulatory compliance
8	<b>Draw data archiving process diagram. BTL 3</b> <pre> graph LR     Preproc([Preproc]) --&gt; Write([Write])     Write --&gt; Delete([Delete])          subgraph Left [ ]         direction TB         L1[Checks] --&gt; L2[Completed]         L2 --&gt; L3[Status Can be Archived]         L3 --- Note1[Selected documents checked to see whether they can be archived -&gt; If successful: Status Can be Archived.]     end          subgraph Middle [ ]         direction TB         M1[Documents with status Can be Archived] --&gt; M2[Archive files]         M2 --- Note2[Data that can be archived is written sequentially to the archive file. Actual archiving run.]     end          subgraph Right [ ]         direction TB         R1[Read data record from archive file] --&gt; R2[If successful: Delete archived documents]         R2 --- Note3[Archive files read -&gt; If successful: - Documents physically deleted from database - Archive index created automatically]     end </pre>
9	<b>How to Create an Archive? BTL 1</b> Usage: hadoop archive -archiveName name -p <parent><src>* <dest> -archiveName is the name of the archive you would like to create. An example would be foo.har. The name should have a *.har extension. The parent argument is to specify the relative path to which the files should be archived to. Example would be : -p /foo/bar a/b/c e/f/g Here /foo/bar is the parent path and a/b/c, e/f/g are relative paths to parent. Note that this is a Map/Reduce job that creates the archives. You would need a map reduce cluster to run this. If you just want to archive a single directory /foo/bar then you can just use hadoop archive -archiveNamezoo.har -p /foo/bar /outputdir
10	<b>What is life cycle costing? BTL 1</b> The process of identifying and documenting all the costs involved over the life of an asset is known as Life Cycle Costing (LCC). A life cycle cost analysis involves the analysis of the costs of a system or a component over its entire life span. Typical costs for a system may include: Acquisition costs, Operating costs, and Maintenance costs
11	<b>How to Estimate Life Cycle Costs? BTL 1</b> The life cycle cost of an asset can be expressed by the simple formula: Life Cycle Cost = initial (projected) capital costs + projected life-time operating costs + projected life-time maintenance costs +

	projected capital rehabilitation costs + projected disposal costs - projected residual value.															
12	<b>What are the traditional databases Approach for bigdata? BTL 1</b> In this approach, an enterprise will have a computer to store and process big data. Here data will be stored in an RDBMS like Oracle Database, MS SQL Server or DB2 and sophisticated software can be written to interact with the database, process the required data and present it to the users for analysis purpose.															
13	<b>What is Hadoop? BTL 1</b> Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.															
14	<b>List out the industry fields were big data applications are used.BTL 2</b> Banking and Securities, Communications, Media and Entertainment, Healthcare Providers, Education, Manufacturing and Natural Resources, Government, Insurance, Retail and Whole sale trade, Transportation, Energy and Utilities.															
15	<b>List out some of the big data apps in real world.BTL 2</b> Big Data Apps: Roambi, Esri ArcGIS, cloudera enterprises, zaloni bedrock, tamr															
16	<b>Give the big data characteristics.BTL 2</b> <ul style="list-style-type: none"> <li>• Data Volume: The big word in big data itself defines the volume.</li> <li>• Data velocity: Deals with the speed of the data coming from various sources</li> <li>• Data Variety: Is the measure of the richness of data representation-text, images, video, audio etc</li> </ul> Data value: It Measures the usefulness of data in making decisions															
17	<b>What are the challenges in big data?(NOV/DEC 2016) BTL 1</b> <ul style="list-style-type: none"> <li>• Meeting the need for speed: In today's hypercompetitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly</li> <li>• Understanding the data: It takes a lot of understanding to get data in the right shape so that you can use visualization as part of data analysis.</li> <li>• Addressing data quality: the value of data for decision-making purposes will be jeopardized if the data is not accurate or timely (data not in proper context).</li> </ul>															
18	<b>Difference between Big data Testing and Traditional database Testing.BTL 2</b> <table border="1"> <thead> <tr> <th>Properties</th> <th>Traditional database testing</th> <th>Big data testing</th> </tr> </thead> <tbody> <tr> <td><b>Data</b></td> <td>Tester work with structured data</td> <td>Tester works with both structured as well as unstructured data</td> </tr> <tr> <td><b>Infrastructure</b></td> <td>Testing approach is well defined and time-tested</td> <td>Testing approach requires focused R&amp;D efforts</td> </tr> <tr> <td><b>Validation Tools</b></td> <td>It does not require special test environment as the file size is limited</td> <td>It requires special test environment due to large data size and files (HDFS)</td> </tr> <tr> <td></td> <td>Tester uses either the Excel based macros or UI based automation tools</td> <td>No defined tools, the range is vast from programming tools like MapReduce to HIVEQL</td> </tr> </tbody> </table>	Properties	Traditional database testing	Big data testing	<b>Data</b>	Tester work with structured data	Tester works with both structured as well as unstructured data	<b>Infrastructure</b>	Testing approach is well defined and time-tested	Testing approach requires focused R&D efforts	<b>Validation Tools</b>	It does not require special test environment as the file size is limited	It requires special test environment due to large data size and files (HDFS)		Tester uses either the Excel based macros or UI based automation tools	No defined tools, the range is vast from programming tools like MapReduce to HIVEQL
Properties	Traditional database testing	Big data testing														
<b>Data</b>	Tester work with structured data	Tester works with both structured as well as unstructured data														
<b>Infrastructure</b>	Testing approach is well defined and time-tested	Testing approach requires focused R&D efforts														
<b>Validation Tools</b>	It does not require special test environment as the file size is limited	It requires special test environment due to large data size and files (HDFS)														
	Tester uses either the Excel based macros or UI based automation tools	No defined tools, the range is vast from programming tools like MapReduce to HIVEQL														
19	<b>What are the Parameters for Performance Testing? BTL 1</b>															

	<p>Various parameters to be verified for performance testing are</p> <ul style="list-style-type: none"> <li>• Data Storage: How data is stored in different nodes</li> <li>• Commit logs: How large the commit log is allowed to grow</li> <li>• Concurrency: How many threads can perform write and read operation</li> <li>• Caching: Tune the cache setting "row cache" and "key cache."</li> <li>○ Timeouts: Values for connection timeout, query timeout, etc</li> </ul>
20	<p><b>List the needs for big data test environment.</b> BTL 2</p> <p>Test Environment needs depend on the type of application you are testing. For Big data testing, test environment should encompass</p> <ul style="list-style-type: none"> <li>• It should have enough space for storage and process large amount of data</li> <li>• It should have cluster with distributed nodes and data</li> <li>• It should have minimum CPU and memory utilization to keep performance high</li> </ul>
21	<p><b>List the advantages of Client-side Validation.</b> (NOV/DEC 2016) BTL 3</p> <ul style="list-style-type: none"> <li>• Allow for more interactivity by immediately responding to users' actions</li> <li>• Execute quickly because they don't require a trip to the server</li> <li>• May improve the usability of Web sites for users whose browsers support scripts</li> <li>• Can be substituted with alternatives (for example, HTML) if users' browsers do not support scripts</li> </ul>
22	<p><b>What is ILM?</b> BTL 1</p> <p>Information life cycle management (ILM) is a comprehensive approach to managing the flow of an information system's data and associated <u>metadata</u> from creation and initial storage to the time when it becomes obsolete and is deleted. Unlike earlier approaches to data storage management, ILM involves all aspects of dealing with data, starting with user practices, rather than just automating storage procedures, as for example, hierarchical storage management (HSM) does. Also in contrast to older systems, ILM enables more complex criteria for storage management than data age and frequency of access.</p>
23	<p><b>What is Data Staging Validation?</b> BTL 1</p> <p>First step of big data testing, also referred as pre-Hadoop stage involves process validation.</p> <ul style="list-style-type: none"> <li>➤ Data from various source like RDBMS, weblogs, social media, etc. should be validated to make sure that correct data is pulled into system</li> <li>➤ Comparing source data with the data pushed into the Hadoop system to make sure they match</li> <li>➤ Verify the right data is extracted and loaded into the correct HDFS location.</li> </ul>
24	<p><b>What are the test environment needs required for testing bigdata ?</b> BTL 1</p> <p>Test Environment needs depend on the type of application you are testing. For Big data testing, test environment should encompass</p> <ul style="list-style-type: none"> <li>➤ It should have enough space for storage and process large amount of data</li> <li>➤ It should have cluster with distributed nodes and data</li> <li>➤ It should have minimum CPU and memory utilization to keep performance high</li> </ul>
25	<p><b>Why is life cycle costing important to a utility?</b> BTL 3</p> <p><b>Life-cycle cost analysis (LCCA)</b> is a tool to determine the most cost-effective option among different competing alternatives to purchase, own, operate, maintain and, finally, dispose of an object or process, when</p>

	each is equally appropriate to be implemented on technical grounds.
26	<b>Discuss Sensitive Information.</b> BTL 2 Sensitive information is data that must be protected from unauthorized access to safeguard the <u>privacy</u> or <u>security</u> of an individual or organization.
27	<b>Define confidential Information.</b> BTL 1 "Confidential Information" means any nonpublic information pertaining to company's business
28	Write about <b>data retention policy.</b> BTL 2 A data retention policy, or records retention policy, is an organization's established <u>protocol</u> for retaining information for operational or <u>regulatory compliance</u> needs.
29	<b>Define Data Administration.</b> BTL 1 Data administration is the process by which data is monitored, maintained and managed by a data administrator and/or an organization. Data administration allows an organization to control its data assets, as well as their processing and interactions with different applications and business processes.
30	<b>Define MapReduce.</b> BTL 1 MapReduce is the heart of Hadoop®. It is this programming paradigm that allows for massive scalability across hundreds or thousands of servers in a Hadoop cluster. The MapReduce concept is fairly simple to understand for those who are familiar with clustered scale-out data processing solutions.
<b>PART * B</b>	
1	<b>Explain the steps to test and deliver a big data applications.(13M)(NOV/DEC 2016) BTL 2</b> <b>Answer:</b> Page: 5.18-5.20–B.S. Charulatha
	<pre> graph LR     DS[Data Source&lt;br/&gt;(RDBMS, web logs, social media etc.)] --&gt; SH[Source Hadoop]     SH --&gt; ETL[ETL Process]     ETL --&gt; TDW[Target Data Warehouse]     TDW --&gt; BI[BI]          subgraph TE [Test Entry Points]         SH --&gt; TE         TDW --&gt; TE     end          TE --&gt; S1["- Check proper data is pulled into system&lt;br/&gt;- Compare source data with the data landed on hadoop&lt;br/&gt;- Check right data is extracted and loaded into the correct HDFS location"]     TE --&gt; S2["- Map reduce process works correctly&lt;br/&gt;- Data aggregation and segregation rules are implemented on the data&lt;br/&gt;- Key value pairs are generated&lt;br/&gt;- Validating the data after post map reduce process"]          subgraph Marks [ ]         S1 --&gt; M1["(5M)"]         S2 --&gt; M1     end   </pre> <p style="text-align: center;"><b>Testing Big Data</b></p> <p style="text-align: center;">© Guru99.com</p>
	<b>Step 1: Data Staging Validation(2M)</b>
	<ul style="list-style-type: none"> <li>• Data from various source</li> <li>• Comparing source data</li> <li>• Verify the right data</li> </ul>

**Step 2: "MapReduce" Validation(2M)**

- Map Reduce process works correctly
- Data aggregation or segregation rules are implemented on the data
- Key value pairs are generated
- Validating the data after Map Reduce process

**Step 3: Output Validation Phase(2M)**

- To check the transformation rules
- To check the data integrity
- To check that there is no data corruption

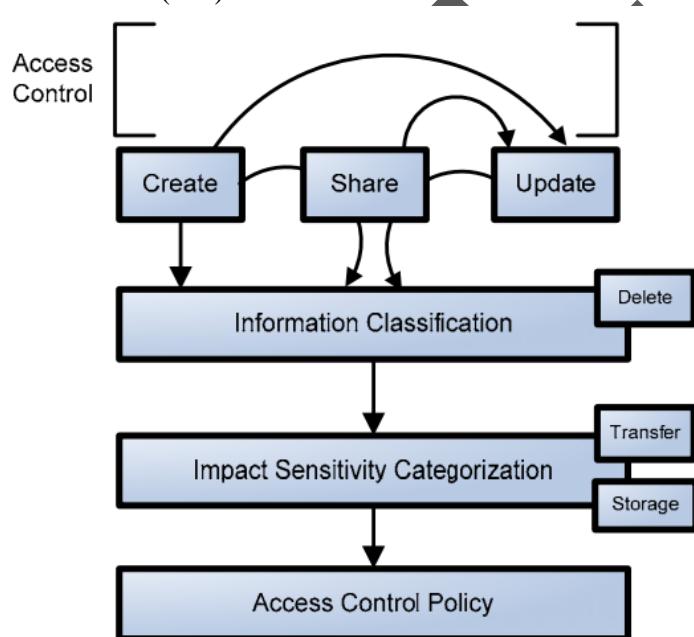
**Step 4: Performance Testing Approach(2M)**

- Data ingestion and Throughput
- Data Processing
- Sub-Component Performance

2 Explain Information life cycle management and its stages. (13)(NOV/DEC 2016) BTL 2

**Answer: Page: 5.10-5.15–B.S. Charulatha**

1. Creation (7M)
2. Retention (organisation, storage, security, etc)
3. Maintenance
4. Use (retrieval, access levels etc)
5. Disposal (timely, with appropriate and secure media destruction methods used)

**DIAGRAM(6M)**

**3** What are the big data solutions from traditional approach to Hadoop solution?(13M)BTL 3  
**Answer:** Page: 341-342 - Peter Morville

Properties	Traditional database testing	Big data testing
<b>Data</b>	Structured Data test	Structured and Unstructured Data test
	Well defined testing approach	R&D Based test approach
<b>Infrastructure</b>	It does not require special test environment as the file size is limited	It requires special test environment due to large data size and files (HDFS)
<b>Validation Tools</b>	Excel based macros or UI based automation tools	No defined tools
	Tools used with basic knowledge	More Skills Needed

**4** Explain about data retention policy and its importance. (13M)BTL 3

**Answer:** Page: 5.15-5.16-B.S. Charulatha

- protocol for retaining information (3M)
- for operational or regulatory compliance needs
  - Organize information
  - Dispose of information

**Regulatory compliance**(2M)

- the value of data over time
- only delete data that is not subject

**Proper data disposal** (4M)

- Old data retention policy
- Temporary Records
- Final Records
- Permanent Records
- Accounting and Corporate Tax Records
- Workplace Records
- Employment, Employee, and Payroll Records
- Bank Records
- Legal Records
- Historical Records

**The policy is also helpful to:**(4M)

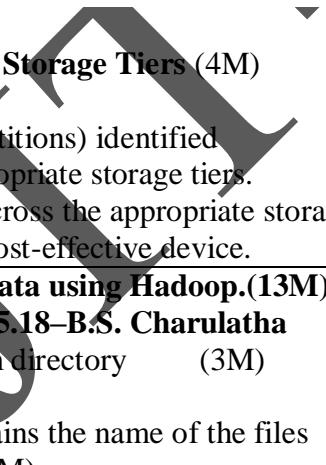
- provide a system for complying with document retention laws;
- ensure that valuable documents are available when needed;
- save money, space and time.
- protect against allegations of selective document destruction; and
- provide for the routine destruction of non-business, superfluous and outdated documents.

**5 Explain in detail about life cycle costing. (13M)BTL 3**

**Answer: Page: 5.1-5.5-B.S. Charulatha**

-Cost savings that can result from using multiple tiered storage. (4M)

-the costs savings can be quite significant and if the data is suitable for database compression,



Storage Tier	Single Tier using High Performance Disks	Multiple Storage Tiers	Multiple Tiers with Database Compression
High Performance (50gb)	\$3,600	\$3,600	\$3,600
Low Cost (500gb)	\$36,000	\$7,000	\$7,000
Online Archive (2Tb)	\$144,000	\$14,000	\$8,500
	<b>\$180,300</b>	<b>\$24,600</b>	<b>\$19,100</b>

(5M)



**Assigning Classes to Storage Tiers (4M)**

- The data classes (partitions) identified
- Assigned to the appropriate storage tiers.
- Distribute the data across the appropriate storage devices
- Stored on the most cost-effective device.

**6 Explain archiving data using Hadoop.(13M) BTL 3**

**Answer: Page: 5.15-5.18-B.S. Charulatha**

-Maps to a file system directory (3M)

-Contains metadata

-The \_index file contains the name of the files

**Create an Archive(3M)**

Usage: hadoop archive -archiveName name -p <parent><src>\* <dest>-archive

Example would be :- p /foo/bar a/b/c e/f/g

**Look Up Files in Archives(3M)**

har://scheme-hostname:port/archivepath/fileinarchive

	<p>har:///archivepath/fileinarchive</p> <p><b>The output should be:</b>(4M)      har:///user/zoo/foo.har/dir1      har:///user/zoo/foo.har/dir2</p>
<b>PART * C</b>	
1	<p><b>Discuss the challenges and issues in bigdata.</b>BTL 1</p> <p><b>Answer: Page:5.21- B.S Charulatha</b></p> <p><b>Challenges in Big Data</b> (8M)</p> <ul style="list-style-type: none"> <li>• <b>Automation-</b> requires someone with a technical expertise.</li> <li>• <b>Virtualization-</b> managing images in Big data is a hassle</li> <li>• <b>Large Dataset-</b>Automate the testing effort, testing in different platform</li> </ul> <p><b>Performance testing challenges</b> (7M)</p> <ul style="list-style-type: none"> <li>• <b>Diverse set of technologies:</b> Requires testing in isolation</li> <li>• <b>Unavailability of specific tools:</b> No single tool can perform the end-to-end testing.</li> <li>• <b>Test Scripting:</b> High degree of scripting is needed to design test scenarios and test cases</li> <li>• <b>Test environment:</b> Needs special test environment due to large data size</li> <li>• <b>Monitoring Solution:</b> Limited solutions exists that can monitor the entire environment</li> </ul>
X X X X X X X X	
2	<p><b>Explain some real time big data apps in detail (15M)</b> BTL 2</p> <p><b>Answer: Page: 5.21–5.24 B.S. Charulatha</b></p> <ul style="list-style-type: none"> <li>• <b>Procurement with Big data:</b> Demand can be forecasted properly as per different conditions available with Big Data. (2M)</li> <li>• <b>Big data in Product development:</b> What product to be developed to increase sales (1M)</li> <li>• <b>Big data in manufacturing sector:</b> Used to identify machinery and process variations that may be indicators of quality problems. (1M)</li> <li>• <b>Big data for product distribution:</b> Analysis could be done to ensure proper distribution in proper market. (2M)</li> <li>• <b>Big data in Marketing field:</b> helps in knowing better marketing strategy that could increase. (2M)</li> <li>• <b>Price Management using Big data:</b> Price management plays a key role and Big data helps business in knowing market trend for it. (2M)</li> <li>• <b>Merchandising:</b> Plays a major role in sales for retail market. (2M)</li> <li>• <b>Big data in Sales:</b> It helps in increasing sale for the business. It also helps in optimizing assignment of sales resources and accounts, product mix and other operations. (1M)</li> <li>• <b>Big data in Social Media:</b> Driving factor behind every marketing decision made by social media companies and it is driving personalization to the extreme. (1M)</li> <li>• <b>Big data in Ecosystem conservation:</b> Learn how ecosystem has been benefited by big data (1M)</li> </ul>

CS6701 CRYPTOGRAPHY AND NETWORK SECURITY

LTPC 3003

## **OBJECTIVES:**

The student should be made to:

- \*Understand OSI security architecture and classical encryption techniques.
  - \*Acquire fundamental knowledge on the concepts of finite fields and number theory.
  - \*Understand various block cipher and stream cipher models.
  - \*Describe the principles of public key cryptosystems, hash functions and digital

## **UNIT I      INTRODUCTION & NUMBER THEORY**

10

Services, Mechanisms and attacks-the OSI security architecture-Network security model- Classical Encryption techniques (Symmetric cipher model, substitution techniques, transposition techniques, steganography).FINITE FIELDS AND NUMBER THEORY: Groups, Rings, Fields-Modular arithmetic- Euclid's algorithm-Finite fields- Polynomial Arithmetic –Prime numbers-Fermat's and Euler's theorem-Testing for primality -The Chinese remainder theorem- Discrete logarithms.

UNIT II BLOCK CIPHERS & PUBLIC KEY CRYPTOGRAPHY

10

Data Encryption Standard-Block cipher principles-block cipher modes of operation-Advanced Encryption Standard (AES)-Triple DES-Blowfish-RC5 algorithm. Public key cryptography: Principles of public key cryptosystems-The RSA algorithm-Key management – Diffie Hellman Key exchange-Elliptic curve arithmetic-Elliptic curve cryptography.

UNIT III HASH FUNCTIONS AND DIGITAL SIGNATURES

8

Authentication requirement – Authentication function – MAC – Hash function – Security of hash function and MAC – MD5 – SHA – HMAC – CMAC – Digital signature and authentication protocols – DSS – El Gamal – Schnorr

UNIT IV SECURITY PRACTICE & SYSTEM SECURITY

8

Authentication applications – Kerberos – X.509 Authentication services – Internet Firewalls for Trusted System: Roles of Firewalls – Firewall related terminology- Types of Firewalls – Firewall designs – SET for E-Commerce Transactions. Intruder – Intrusion detection system – Virus and related threats – Countermeasures – Firewalls design principles – Trusted systems – Practical ofcryptography and security.

UNIT V E-MAIL / IP & WEB SECURITY

9

E-mail Security: Security Services for E-mail-attacks possible through E-mail – establishing keys privacy-authentication of the source-Message Integrity-Non-repudiation-Pretty Good Privacy-S/MIME. IPSecurity: Overview of IPSec – IP and IPv6-Authentication Header-Encapsulation Security Payload (ESP)-Internet Key Exchange (Phases of IKE, ISAKMP/IKE Encoding). Web Security: SSL/TLS Basic Protocol-computing the keys- client authentication-PKI as deployed by SSLAttacks fixed in v3- Exportability-Encoding-Secure Electronic Transaction (SET).

TOTAL: 45 PERIODS

**TEXT BOOKS:**

1. William Stallings, Cryptography and Network Security, 6th Edition, Pearson Education, March 2013. (UNIT I,II,III,IV).
2. Charlie Kaufman, Radia Perlman and Mike Speciner, "Network Security", Prentice Hall of India, 2002. (UNIT V).

**REFERENCES:**

1. Behrouz A. Ferouzan, "Cryptography & Network Security", Tata Mc Graw Hill, 2007.
2. Man Young Rhee, "Internet Security: Cryptographic Principles", "Algorithms and Protocols", Wiley Publications, 2003.
3. Charles Pfleeger, "Security in Computing", 4th Edition, Prentice Hall of India, 2006.
4. Ulysses Black, "Internet Security Protocols", Pearson Education Asia, 2000.
5. Charlie Kaufman and Radia Perlman, Mike Speciner, "Network Security, Second Edition, Private Communication in Public World", PHI 2002.
6. Bruce Schneier and Neils Ferguson, "Practical Cryptography", First Edition, Wiley Dreamtech India Pvt Ltd, 2003.
7. Douglas R Simson "Cryptography – Theory and practice", First Edition, CRC Press, 1995.
8. <http://nptel.ac.in/>.

Subject Code: CS6701

Year/Semester: IV/07

Subject Name: Cryptography &amp; Network Security

Subject Handler: Daya Mathew

**UNIT I-INTRODUCTION & NUMBER THEORY**

**Services, Mechanisms and attacks-the OSI security architecture-Network security model-Classical Encryption techniques (Symmetric cipher model, substitution techniques, transposition techniques, steganography). FINITE FIELDS AND NUMBER THEORY:** Groups, Rings, Fields-Modular arithmetic-Euclid's algorithm-Finite fields- Polynomial Arithmetic -Prime numbers-Fermat's and Euler's theorem-Testing for primality -The Chinese remainder theorem- Discrete logarithms.

**PART\* A**

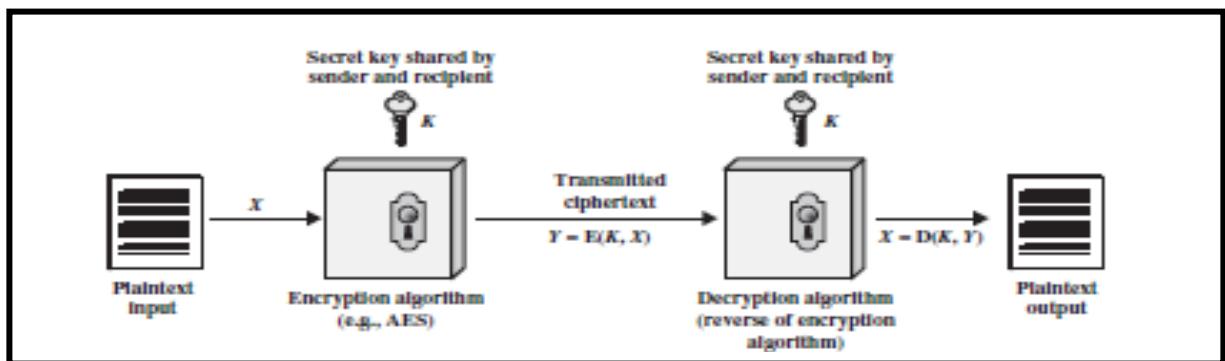
Q.N O	QUESTIONS
1.	<p><b>Specify the four categories of security threats.</b> BTL3</p> <ul style="list-style-type: none"> <li>✓ Interruption</li> <li>✓ Interception</li> <li>✓ Modification</li> <li>✓ Fabrication</li> </ul>
2.	<p><b>Define active and passive attack with example.</b> BTL1</p> <p><b>Passive attack:</b></p> <ul style="list-style-type: none"> <li>✓ Monitoring the message during transmission.</li> <li>✓ Difficult to detect</li> <li>✓ Does not affect system</li> </ul> <p><b>Eg: Interception</b></p> <p><b>Active attack:</b></p> <ul style="list-style-type: none"> <li>✓ It involves the modification of data stream or creation of false data stream</li> <li>✓ Easy to detect</li> <li>✓ Easily affects system</li> </ul> <p><b>E.g.: Fabrication, Modification, and Interruption</b></p>
3.	<p><b>Define integrity and non repudiation.</b> BTL1</p> <p><b>Integrity:</b> Service that ensures that only authorized person able to modify the message</p> <p><b>Non repudiation:</b> This service helps to prove that the person who denies the transaction is true or</p>

	false.				
4.	<p><b>Differentiate symmetric and asymmetric encryption.</b> BTL3</p> <table border="1"> <thead> <tr> <th>Symmetric</th><th>Asymmetric</th></tr> </thead> <tbody> <tr> <td>           It is a form of cryptosystem in which encryption and decryption performed using the same key.            Eg : DES,AES         </td><td>           It is a form of cryptosystem in which encryption and decryption performed using two keys. Eg : RDA, ECC         </td></tr> </tbody> </table>	Symmetric	Asymmetric	It is a form of cryptosystem in which encryption and decryption performed using the same key. Eg : DES,AES	It is a form of cryptosystem in which encryption and decryption performed using two keys. Eg : RDA, ECC
Symmetric	Asymmetric				
It is a form of cryptosystem in which encryption and decryption performed using the same key. Eg : DES,AES	It is a form of cryptosystem in which encryption and decryption performed using two keys. Eg : RDA, ECC				
5.	<p><b>Define cryptanalysis.</b> BTL1</p> <p>Cryptanalysis is a process of attempting to discover the key or plaintext or both.</p>				
6.	<p><b>Define security mechanism.</b> BTL1</p> <p>It is process that is designed to detect prevent, recover from a security attack.</p> <p><b>Example: Encryption algorithm, Digital signature, Authentication protocols</b></p>				
7.	<p><b>Define steganography.</b> BTL1</p> <p>Hide in plain sight .Hiding the message into some cover media. It conceals the existence of a message</p>				
8.	<p><b>Why network needs security?</b> BTL2</p> <p>When systems are connected through the network, attacks are possible during transmission time.</p>				
9.	<p><b>Define confidentiality and authentication.</b> BTL1</p> <p><b>Confidentiality:</b> It means how to maintain the secrecy of message. It ensures that the information in a computer system and transmitted information are accessible only for reading by authorized person.</p> <p><b>Authentication:</b> It helps to prove that the source entity only has involved the transaction.</p>				
10.	<p><b>Define cryptography.</b> BTL1</p> <p>It is a science of writing Secret code using mathematical techniques. The many schemes used for enciphering constitute the area of study known as cryptography.</p>				
11.	<p><b>Compare Substitution and Transposition techniques.</b> BTL2</p> <table border="1"> <thead> <tr> <th>SUBSTITUTION</th><th>TRANSPOSITION</th></tr> </thead> <tbody> <tr> <td>           A substitution techniques is one in which the letters of plaintext are replaced by other letter or by number or symbols   <b>Eg: Caeser cipher</b> </td><td>           It means, different kind of mapping is achieved by performing some sort of permutation on the plaintext letters.   <b>Eg: DES, AES</b> </td></tr> </tbody> </table>	SUBSTITUTION	TRANSPOSITION	A substitution techniques is one in which the letters of plaintext are replaced by other letter or by number or symbols  <b>Eg: Caeser cipher</b>	It means, different kind of mapping is achieved by performing some sort of permutation on the plaintext letters.  <b>Eg: DES, AES</b>
SUBSTITUTION	TRANSPOSITION				
A substitution techniques is one in which the letters of plaintext are replaced by other letter or by number or symbols  <b>Eg: Caeser cipher</b>	It means, different kind of mapping is achieved by performing some sort of permutation on the plaintext letters.  <b>Eg: DES, AES</b>				

12.	<p><b>Define Diffusion &amp; Confusion. BTL1</b></p> <p><b>Diffusion</b></p> <p>It means each plaintext digits affect the values of many ciphertext digits which is equivalent to each ciphertext digit is affected by many plaintext digits. It can be achieved by performing permutation on the data. It is the relationship between the plaintext and ciphertext</p> <p><b>Confusion:</b></p> <p>It can be achieved by substitution algorithm. It is the relationship between cipher text and key</p>
13	<p><b>Define Multiple Encryption. BTL2</b></p> <p>It is a technique in which the encryption is used multiple times. Eg: Double DES, Triple DES</p>
14	<p><b>Specify the design criteria of block cipher. BTL3</b></p> <p>Number of rounds</p> <p>Design of the function F</p> <p>Key scheduling</p>
15	<p><b>Define Reversible mapping. BTL1</b></p> <p>Each plain text is maps with the unique cipher text. This transformation is called reversible mapping</p>
16	<p><b>Specify the basic task for defining a security serviceBTL3</b></p> <p>A service that enhances the security of the data processing systems and the information transfer of an organization. The services are intended to counter security attack, and they make use of one or more security mechanism to provide the service</p>
17	<p><b>Define network security. BTL1</b></p> <p>This area covers the use of cryptographic algorithms in network protocols and network applications.</p>
18	<p><b>Define computer security. BTL1</b></p> <p>This term refers to the security of computers against intruders and malicious software.</p>
19	<p><b>What are hill cipher merits and demerits? BTL1</b></p> <p>Completely hides single letter and 2 letter frequency information.</p>
20.	<p><b>List-out the types of attack in ceaser cipher. BTL2</b></p> <p>Brute force attack.</p> <p>Just try all the 25 possible keys</p>
21	<p><b>Define integrity and nonrepudiation? BTL1</b></p> <p><b>Integrity:</b></p> <p>Service that ensures that only authorized person able to modify the message.</p> <p><b>Nonrepudiation:</b></p>

	This service helps to prove that the person who denies the transaction is true or false.
22	<b>Write short notes Congruence.</b> BTL3 Let $a, b, n$ be integers with $n \neq 0$ . We say that $a \equiv b \pmod{n}$ If $a - b$ is a multiple of $n$ .
23	<b>What is Key?</b> BTL1 A sequence of symbols that controls the operation of a cryptographic transformation. A key is normally a string of bits used by a cryptographic algorithm to transform plain text into cipher text or vice versa. The key should be the only part of the algorithm that it is necessary to keep secret.
24	<b>What is Plain text &amp; Ciphertext?</b> BTL1 <b>Plaintext:</b> An original message is known as the <b>plaintext(Readable format)</b> <b>Ciphertext:</b> coded message is called the <b>Cipher Text.(Unreadable format)</b>
25	<b>List the different Types of Ciphers.</b> BTL2 <ul style="list-style-type: none"> <li>✓ Shift Ciphers.</li> <li>✓ Affine Ciphers</li> <li>✓ Vigenere Cipher</li> <li>✓ Substitution Ciphers</li> <li>✓ Sherlock Holmes</li> <li>✓ Playfair and ADFGX Ciphers</li> <li>✓ Block ciphers</li> <li>✓ One-Time pads</li> </ul>

<b>PART *B</b>	
1	<p>i) Explain about symmetric cipher models (May/June 2012) (13M) BTL 4  <b>Answer:</b> Pageno.:57 to 59 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ Symmetric cipher model (2M)  <b>A symmetric key cipher</b> (also called a <b>secret-key cipher</b>, or a <b>one-key cipher</b>, or a <b>private-key cipher</b>, or a <b>shared-key cipher</b>) Shared _secret is one that uses the same (necessarily secret) <b>key</b> to encrypt messages as it does to decrypt messages.</li> <li>✓ Plain text (1M)  Original message or data</li> <li>✓ Encryption Algorithm (2M)  Various substitutions and transformations</li> <li>✓ Decryption Algorithm (2M)  Produces plaintext</li> <li>✓ Cipher Text (1M)  Scrambled Message</li> <li>✓ Principles of Security (2M)</li> <li>✓ Diagram (3M)</li> </ul>



2

**Explain the various substitution techniques. (AU Nov/Dec 2011) (13M) BTL4**

**Answer:**Pageno.:62 to 69 in William Stallings

- ✓ **Caesar cipher** : replacing each letter of the alphabet with the letter standing three places further down the alphabet (1M)
- ✓ **Formula :**(1M)

$$C = E(3, p) = (p + 3) \bmod 26$$

- ✓ **Example:**(1M)
- ✓ **Mono alphabetic Cipher**: permutation of the 26 alphabetic characters (1M))
- ✓ **Example:**(1M)
- ✓ **Hill Fair cipher**: polygraphic substitution *cipher* based on linear algebra (1M)
- ✓ **Formula:**(1M)

$$\begin{aligned} C &= E(K, P) = PK \bmod 26 \\ P &= D(K, C) = CK^{-1} \bmod 26 = PKK^{-1} = P \end{aligned}$$

- ✓ **Example:**(1M)
- ✓ **Play Fair cipher**: treats digrams in the plaintext as single units and translates these units into ciphertext
- ✓ **Example:**(1M)
- ✓ Poly alphabetic cipher : technique is to use different monoalphabetic substitutions as one proceeds through the plaintext message (2M)
- ✓ One time pad (2M): Each new message requires a new key of the same length as the new message

3

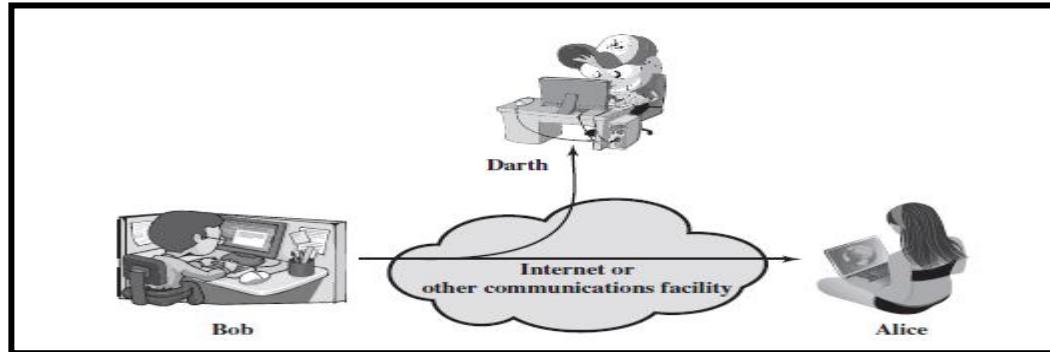
**Describe about transposition techniques.(10M) BTL2**

**Answer:**Pageno.:77 to 82 in William Stallings

	<ul style="list-style-type: none"> <li>✓ <b>Rail fence technique:</b> the plaintext is written down as a sequence of diagonals and then read off as a sequence of rows (1M)</li> <li>✓ <b>Example:</b> (2M)</li> <li>✓ <b>Coloumnar technique:</b> It is a transposition cipher that follows a simple rule for mixing up the characters in the plaintext to form the ciphertext. (1M)</li> <li>✓ <b>Example:</b> (2M)</li> <li>✓ <b>Rotor machines:</b> principle of multiple stages of encryption was a class of systems known as rotor machines. (2M)</li> <li>✓ <b>Stegnography:</b> conceal the existence of the message (2M)</li> </ul>
4	<p><b>Explain Security Services.(10M)BTL4</b>  <b>Answer:Pageno.:101 to 105 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ <b>Authentication:</b> assuring that a communication is authentic (2M)</li> <li>✓ <b>Access Control :</b> the ability to limit and control the access to host systems and applications via communications links. (2M)</li> <li>✓ <b>Data Confidentiality :</b> protection of transmitted data from passive attacks (2M)</li> <li>✓ <b>Data Integrity:</b>protecting information from being modified by unauthorized parties (2M)</li> <li>✓ <b>NonRepudation:</b> prevents either sender or receiver from denying a transmitted message. (2M)</li> </ul>
5	<p><b>Explain various Security Mechanisms.(8M) BTL4</b>  <b>Answer:Pageno.:105 to 107 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ <b>Pervasive Security Mechanisms:</b> Mechanisms that are not specific to any particular OSI security service or protocol layer. (2M) <ul style="list-style-type: none"> <li>(i) <b>Trusted Functionality :</b> perceived to be correct with respect to some criteria (1/2 M)</li> <li>(ii) <b>Security Label:</b> marking bound to a resource(1/2 M)</li> <li>(iii) <b>Event Detection:</b> Detection of security-relevant events. (1/2 M)</li> <li>(iv) <b>Security Audit Trail:</b> facilitate a security audit(1/2 M)</li> </ul> </li> <li>✓ <b>Specific Security Mechanism :</b> incorporated into the appropriate protocol (2M) <ul style="list-style-type: none"> <li>(i) <b>Encipherment :</b> mathematical algorithms to transform data into a form that is not readily intelligible (1/2 M)</li> <li>(v) <b>Digital Signature:</b> allows a recipient of the data unit to prove the source and integrity of the data unit and protect against forgery (1/2 M)</li> <li>(vi) <b>Access Control :</b> enforce access rights to resources. (1/2 M)</li> <li>(vii) <b>Data Integrity :</b> assure the integrity of a data unit or stream of data units. (1/2 M)</li> </ul> </li> </ul>
6	<p><b>Explain various Security Attacks.(13M) BTL4</b>  <b>Answer:Pageno.:107 to 109 in William Stallings</b></p>

- ✓ **Passive Attack:** nature of eavesdropping on, or monitoring of, transmissions (1M)

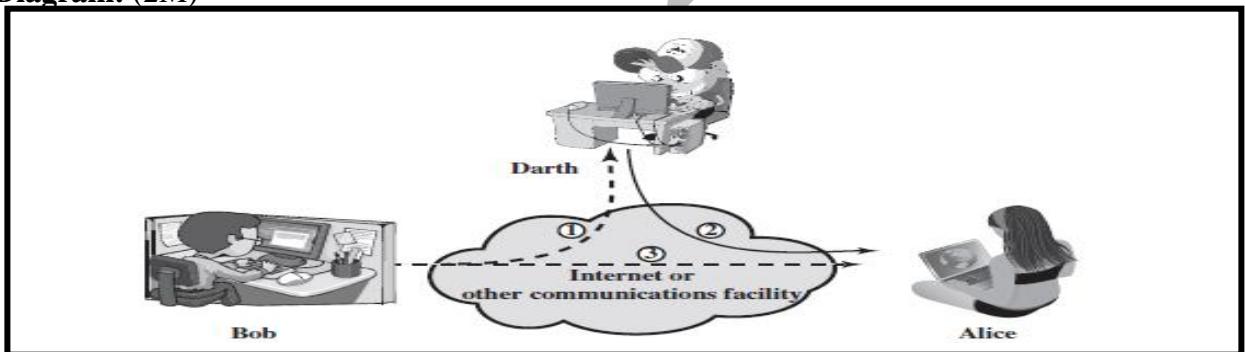
- Release of Message Contents(1/2 M)
- Traffic Analysis(1/2 M)
- Diagram(2M)



- ✓ **Active Attack :** involve some modification of the data stream or the creation of a false stream

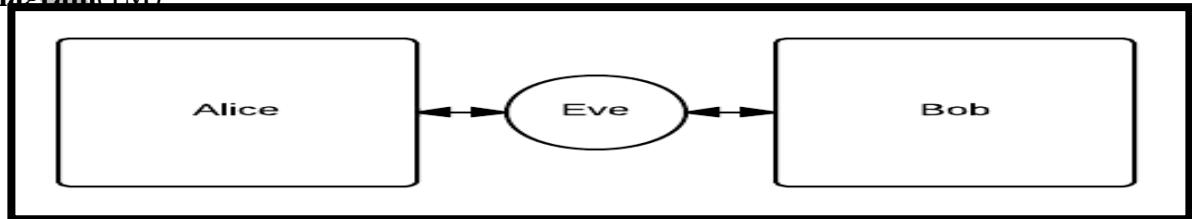
- Masquerade:** when one entity pretends to be a different entity (1/2M)
- Replay:** involves the passive capture of a data unit and its subsequent retransmission to produce an unauthorized effect (1/2M)
- Modification of message:** some portion of a legitimate message is altered, or that messages are delayed or reordered, to produce an unauthorized effect (1/2M)
- Denial of service:** prevents or inhibits the normal use or management of communications facilities (1/2M)

**Diagram:** (2M)



- ✓ **Man in Middle Attack :** attack where the attacker secretly relays and possibly alters the communication between two parties who believe they are directly communicating with each other. (2M)

**Diagram(1M)**



8	<p><b>Explain about finite fields.(13M) BTL4</b>  <b>Answer:</b>Pageno.:144 to 147 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ Groups, Rings, Fields (2M)            A <b>group</b> <math>G</math>, sometimes denoted by <math>\{G, \sim\}</math>, is a set of elements with a binary operation denoted by <math>\sim</math> that associates to each ordered pair <math>(a, b)</math> of elements in <math>G</math> an element <math>(a \sim b)</math> in <math>G</math>            Properties(1/2M)</li> <li>✓ Rings:<b>A ring</b> <math>R</math>, sometimes denoted by <math>\{R, +, *\}</math>, is a set of elements with two binary operations, called <i>addition</i> and <i>multiplication</i>, such that for all <math>a, b, c</math> in <math>R</math> the following axioms are obeyed.(1/2M)</li> <li>✓ Fields(1/2M)            A <b>field</b> <math>F</math>, sometimes denoted by <math>\{F, +, *\}</math>, is a set of elements with two binary operations, called <i>addition</i> and <i>multiplication</i>, such that for all <math>a, b, c</math> in <math>F</math> the following axioms are obeyed.(1/2M)</li> <li>✓ <b>Modular Arithmetic :</b>(2M)            If <math>a</math> is an integer and <math>n</math> is a positive integer, we define <math>a \bmod n</math> to be the remainder when <math>a</math> is divided by <math>n</math>. The integer <math>n</math> is called the <b>modulus</b>.</li> </ul> <p><b>Formula(1M)</b></p> <div style="border: 1px solid black; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"> <math display="block">a = qn + r \quad 0 \leq r &lt; n; q = [a/n]</math> <math display="block">a = [a/n] \times n + (a \bmod n)</math> <math display="block">11 \bmod 7 = 4; \quad -11 \bmod 7 = 3</math> </div> <p>✓ <b>Euclidean algorithm :</b>(2M)            One of the basic techniques of number theory is the Euclidean algorithm, which is a simple procedure for determining the greatest common divisor of two positive integers.</p> <p><b>Formula(1M)</b></p> <div style="border: 1px solid black; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"> <math display="block">\gcd(a, b) = \gcd(b, a \bmod b)</math> <math display="block">\gcd(55, 22) = \gcd(22, 55 \bmod 22) = \gcd(22, 11) = 11</math> </div> <p>✓ <b>Finite fields of form GF(P) :</b>(2M)            For a given prime, <math>p</math>, we define the finite field of order <math>p</math>, <math>\text{GF}(p)</math>, as the set <math>Z</math> of integers <math>\{0, 1, 2, \dots, p-1\}</math> together with the arithmetic operations modulo <math>p</math>.</p> <p>✓ <b>Polynomial arithmetic :</b>(1M)            Polynomial arithmetic in which the arithmetic on the coefficients is performed modulo <math>p</math>; that is, the coefficients are in <math>\text{GF}(p)</math>.</p>
9	<p><b>Explain about Euclidean algorithm. (13M) BTL4</b>  <b>Answer:</b>Pageno.:129 to 132 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ <b>Euclidean algorithm :</b>(2M)            One of the basic techniques of number theory is the Euclidean algorithm, which</li> </ul>

- is a simple procedure for determining the greatest common divisor of two positive integers.
- ✓ **Formula:** (2M)

$$\begin{array}{ll}
 a = q_1 b + r_1 & 0 < r_1 < b \\
 b = q_2 r_1 + r_2 & 0 < r_2 < r_1 \\
 r_1 = q_3 r_2 + r_3 & 0 < r_3 < r_2 \\
 \vdots & \vdots \\
 \vdots & \vdots \\
 \vdots & \vdots \\
 r_{n-2} = q_n r_{n-1} + r_n & 0 < r_n < r_{n-1} \\
 r_{n-1} = q_{n+1} r_n + 0 & \\
 d = \gcd(a, b) = r_n &
 \end{array}$$

- ✓ **Greatest common divisor**(2M)

$$\gcd(a, b) = \gcd(b, a \bmod b)$$

$$\gcd(55, 22) = \gcd(22, 55 \bmod 22) = \gcd(22, 11) = 11$$

- ✓ **Finding GCD**(2M)

$$a = q_1 b + r_1 \quad 0 \leq r_1 < b$$

- ✓ **Example**(5M)

10

**Explain Fermat Theorem. (10M) BTL4**

**Answer:**Pageno.:272 to 275 in William Stallings

- ✓ **Fermat's Theorem:** (2M)

Fermat's theorem states the following: If  $p$  is prime and  $a$  is a positive integer not divisible by  $p$ , then

- ✓ **Formulas :**(2M)

$$a^{p-1} = 1 \pmod{p}$$

- ✓ **Proof :**(3M)

$$\begin{aligned}
 a &= 7, p = 19 \\
 7^2 &= 49 = 11 \pmod{19} \\
 7^4 &= 121 = 7 \pmod{19} \\
 7^8 &= 49 = 11 \pmod{19} \\
 7^{16} &= 121 = 7 \pmod{19} \\
 a^{p-1} &= 7^{18} = 7^{16} \times 7^2 = 7 \times 11 = 1 \pmod{19}
 \end{aligned}$$

- ✓ **Example :**(3M)

11	<p><b>Explain about Euler's theorem. (13M) BTL4</b></p> <p><b>Answer:</b>Pageno.:273 to 276 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ <b>Euler Theorem :</b>(2M)</li> </ul> <p>Euler's theorem states that for every <math>a</math> and <math>n</math> that are relatively prime:</p> <ul style="list-style-type: none"> <li>✓ <b>Formula :</b>(3M)</li> </ul> $a^{\phi(n)} = 1 \pmod{n}$ <ul style="list-style-type: none"> <li>✓ <b>Proof :</b>true if <math>n</math> is prime, because in that case, <math>f(n) = (n - 1)</math> and Fermat's theorem holds. However, it also holds for any integer <math>n</math>. Recall that <math>f(n)</math> is the number of positive integers less than <math>n</math> that are relatively prime to <math>n</math>. (4M)</li> </ul> <p>Consider the set of such integers, labeled as</p> <ul style="list-style-type: none"> <li>✓ <b>Euler's Totient Function :</b>(4M)</li> </ul> <p>Before presenting Euler's theorem, we need to introduce an important quantity in number theory, referred to as <b>Euler's totient function</b>, written <math>f(n)</math>, and defined as the number of positive integers less than <math>n</math> and relatively prime to <math>n</math>. By convention,</p> $\phi(1) = 1.$
12	<p><b>I )Explain Chinese Remainder Theorem. (10M) BTL4</b></p> <p><b>Answer:</b>Pageno.:278 to 281 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ <b>Chinese Remainder Theorem :</b>(3M)</li> </ul> <p>The <i>Chinese remainder theorem</i> is a <i>theorem</i> which gives a unique solution to simultaneous linear congruences with coprime moduli</p> <ul style="list-style-type: none"> <li>✓ <b>Statements:</b>(2M)</li> </ul> $(A + B) \pmod{M} \leftrightarrow ((a_1 + b_1) \pmod{m_1}, \dots, (a_k + b_k) \pmod{m_k})$ $(A - B) \pmod{M} \leftrightarrow ((a_1 - b_1) \pmod{m_1}, \dots, (a_k - b_k) \pmod{m_k})$ $(A \times B) \pmod{M} \leftrightarrow ((a_1 \times b_1) \pmod{m_1}, \dots, (a_k \times b_k) \pmod{m_k})$ <ul style="list-style-type: none"> <li>✓ <b>Example :</b>(5M)</li> </ul>
13	<p><b>Explain Modular and Discrete Algorithm.(10M) BTL4</b></p> <p><b>Answer:</b>Pageno.:281 to 284 in William Stallings</p>

	<ul style="list-style-type: none"> <li>✓ <b>Modular algorithm :</b>(2M) If <math>a</math> is an integer and <math>n</math> is a positive integer, we define <math>a \bmod n</math> to be the remainder when <math>a</math> is divided by <math>n</math>. The integer <math>n</math> is called the <b>modulus</b>.</li> <li>✓ <b>Modular arithmetic formula :</b>(1M)</li> </ul> <div style="border: 1px solid black; padding: 10px; margin-top: 10px;"> <math display="block">a = qn + r \quad 0 \leq r &lt; n; q = [a / n]</math> <math display="block">a = [a / n] \times n + (a \bmod n)</math> <math display="block">11 \bmod 7 = 4; \quad -11 \bmod 7 = 3</math> </div> <ul style="list-style-type: none"> <li>✓ discrete algorithm(2M)</li> <li>✓ Fundamental to public key algorithm(2M)</li> <li>✓ Formula(3M)</li> </ul>
--	---

### PART\* C

1	<p>a) <b>Find <math>3^{21}</math> and <math>11</math> using Fermat's Theorem(7M) BTL5</b>  b) <b>Find <math>11^7</math> and <math>13</math> using fermat's theorem(8M) BTL5</b></p> <p><b>Answer:</b>Pageno.:272 to 274 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ <b>Formula:</b>(4M)  <math display="block">a^{p-1} \equiv 1 \pmod{p}</math></li> <li>✓ <b>Steps:</b>(3M)</li> <li>✓ <b>Formula:</b>(4M)  <math display="block">a^{p-1} \equiv 1 \pmod{p}</math></li> <li>✓ <b>Steps:</b>(4M)</li> </ul>
2.	<p><b>Encrypt the following using play fair cipher using the keyword MONARCHY "SWARAJ IS MY BIRTH RIGHT". Use X for blank spaces.(15M) BTL 6</b></p> <p><b>Answer:</b>Pageno.:65 to 67 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ <b>Rules:</b>(6M) <ul style="list-style-type: none"> <li>(i) If both the letters are in the same column, take the letter below each one (going back to the top if at the bottom)</li> <li>(ii) First, a plaintext message is split into pairs of two letters (digraphs)</li> <li>(iii) If both the letters are in the same column, take the letter below each one</li> <li>(iv) If both letters are in the same row, take the letter to the right of each one</li> <li>(v) If neither of the preceding two rules are true, form a rectangle with the two letters and take the letters on the horizontal opposite corner of the rectangle</li> </ul> </li> <li>✓ <b>Solution:</b>(9M)</li> </ul>

3	<p><b>Apply Caesar cipher and k=5 decrypt the given Cipher text “YMJTYMJWXNIJTKXNQJSHJ”. (15M)BTL5</b></p> <p><b>Answer:</b>Pageno.:62 to 63 in William Stallings</p> <p>✓ <b>Definition:</b>(3M) It is a type of substitution cipher where each letter in the original message (which in cryptography is called the plaintext) is replaced with a letter corresponding to a certain number of letters shifted up or down in the alphabet.</p> <p>✓ <b>Formula:</b>(2M)</p> <div style="border: 1px solid black; padding: 5px; display: inline-block;"> <math display="block">C = E(3, p) = (p + 3) \text{ mod } 26</math> </div> <p>✓ <b>Solution</b>(10M)</p>
4	<p><b>Encrypt the message “PAY” using hill cipher with the following key matrix and show the decryption to formulate original plain text. (15M)BTL6</b></p> $K = \begin{vmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{vmatrix}$ <p><b>Answer:</b>Pageno.:67 to 69 in William Stallings</p> <p>✓ <b>Encryption:</b>(5M) turn our keyword into a key matrix plaintext into digraphs into a column vector perform matrix multiplication modulo the length of the alphabet (i.e. 26) on each vector converted back into letters to produce the ciphertext.</p> <p>✓ <b>Decryption:</b>(3M) must find the inverse matrix</p> <p>✓ <b>Example:</b>(7M)</p>
5	<p><b>Assess the following cipher Text using brute force attack. (15M) BTL6</b></p> <p>CMTMROOEORW (Hint: Algorithm-Rail fence)</p> <p><b>Answer:</b>Pageno.:69 to 71 in William Stallings</p> <p><b>Defintion:</b>(3M) The simplest such cipher is the <b>rail fence</b> technique, in which the plaintext is written down as a sequence of diagonals and then read off as a sequence of rows</p> <p><b>Encryption:</b>(6M)</p> <p><b>Decryption:</b>(6M)</p>
6	<p><b>Explain Security Attacks.(15M) BTL4</b></p> <p><b>Answer:</b>Pageno.:39 to 41 in William Stallings</p>

	<ul style="list-style-type: none"> <li>➤ Security Attacks (3M)</li> <li>➤ Passive attacks (3M)</li> <li>➤ Active attacks (3M)</li> <li>➤ Network security model (3M)</li> <li>➤ Diagrams (3M)</li> </ul>
7	<p><b>Explain Network Security Services. (15M) BTL4</b></p> <p><b>Answer:</b>Pageno.:43 to 45 in William Stallings</p> <ul style="list-style-type: none"> <li>➤ Authentication (3M)</li> <li>➤ Access control (3M)</li> <li>➤ Data confidentiality (3M)</li> <li>➤ Data integrity (2M)</li> <li>➤ Non repudiation (2M)</li> <li>➤ Availability services (2M)</li> </ul>
8	<p><b>Explain Fermats and Euler Theorem.(15M) BTL4</b></p> <p><b>Answer:</b>Pageno.:272 to 275 in William Stallings</p> <ul style="list-style-type: none"> <li>➤ Explanation (5M)</li> <li>➤ Theorem (10M)</li> </ul>
9	<p><b>Illustrate how to solve <math>x^2 \equiv 1 \pmod{35}</math> using Chinese remainder theorem.(15M)BTL6</b></p> <p><b>Answer:</b>Pageno.:278 to 281 in William Stallings</p> <div style="border: 1px solid black; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"> <math display="block">A = \left( \sum_{i=1}^k a_i c_i \right) \pmod{M}</math> <math display="block">C_i = M_i \times (M_i^{-1} \pmod{m_i}) \text{ for } 1 \leq i \leq k</math> </div> <p style="text-align: right;">(2M)</p> <p style="text-align: right;">(3M)</p> <p><b>Encryption :</b>(5M)  <b>Decryption:</b>(5M)</p>
10	<p><b>Estimate <math>11^{13} \pmod{53}</math> using modular exponentiation.(15M) BTL5</b></p>

	<p><b>Answer:</b>Pageno.:112 to 113 in William Stallings</p> <ul style="list-style-type: none"> <li>➤ <math>Z_n = \{0, 1, \dots, (n - 1)\}</math> (2M)</li> <li>➤ <math>a+b \text{ mod } n = [a \text{ mod } n + b \text{ mod } n] \text{ mod } n</math> (3M)</li> </ul> <p><b>Encryption:</b>(5M) <b>Decryption:</b>(5M)</p>
11	<p><b>State the CRT and find X for the given set of congruent equations using CRT.</b></p> <p><math>X \equiv 2 \pmod{3}</math>  <math>X \equiv 3 \pmod{5}</math>  <math>X \equiv 2 \pmod{7}</math> (15M) BTL5</p> <p><b>Answer:</b>Page no.:278 to 281 in William Stallings</p> <div style="border: 1px solid black; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"> <math display="block">A = \left( \sum_{i=1}^k a_i c_i \right) \pmod{M}</math> <math display="block">C_i = M_i \times (M_i^{-1} \pmod{m_i}) \text{ for } 1 \leq i \leq k</math> </div> <p style="text-align: right;">(2M)</p> <p style="text-align: right;">(3M)</p> <p><b>Encryption :</b>(5M) <b>Decryption:</b>(5M)</p>
12	<p><b>Solve the following system of congruence's:</b></p> <p><math>X \equiv 12 \pmod{25}</math>  <math>X \equiv 9 \pmod{26}</math>  <math>X \equiv 23 \pmod{27}</math> (15M) BTL5</p> <p><b>Answer:</b>Pageno.:278 to 281 in William Stallings</p> <div style="border: 1px solid black; padding: 10px; width: fit-content; margin-left: auto; margin-right: auto;"> <math display="block">A = \left( \sum_{i=1}^k a_i c_i \right) \pmod{M}</math> <math display="block">C_i = M_i \times (M_i^{-1} \pmod{m_i}) \text{ for } 1 \leq i \leq k</math> </div> <p style="text-align: right;">(2M)</p> <p style="text-align: right;">(3M)</p> <p><b>Encryption :</b>(5M) <b>Decryption:</b>(5M)</p>

## UNIT 2- BLOCK CIPHERS & PUBLIC KEY CRYPTOGRAPHY

**Data Encryption Standard**-Block cipher principles-block cipher modes of operation-**Advanced Encryption Standard (AES)**-Triple DES-Blowfish-RC5 algorithm. Public key cryptography: Principles of public key cryptosystems-The RSA algorithm-Key management -Diffie Hellman Key exchange-Elliptic curve arithmetic-Elliptic curve cryptography.

### PART\* A

1	<b>Compare stream cipher with block cipher with example.</b> BTL1 <b>Stream cipher:</b> Processes the input stream continuously and producing one element at a time. <b>Example: caeser cipher.</b> <b>Block cipher:</b> Processes the input one block of elements at a time producing an output block for each input block. <b>Example: DES</b>
2	<b>Differentiate unconditionally secured and computationally secured .</b> BTL1 An Encryption algorithm is unconditionally secured means; the condition is if the cipher text generated by the encryption scheme doesn't contain enough information to determine corresponding plaintext. Encryption is computationally secured means, <ul style="list-style-type: none"> <li>✓ The cost of breaking the cipher exceeds the value of enough information.</li> <li>✓ Time required to break the cipher exceed the useful lifetime of information.</li> </ul>
3	<b>What are the design parameters of Feistel cipher network?</b> BTL1 <ul style="list-style-type: none"> <li>✓ Block size</li> <li>✓ Key size</li> <li>✓ Number of rounds</li> <li>✓ Sub key generation algorithm</li> <li>✓ Round function</li> <li>✓ Fast software encryption / decryption</li> <li>✓ Ease of analysis</li> </ul>
4	<b>Define Product cipher.</b> BTL1 Product Cipher means two or more basic cipher are combined together and produces the resultant cipher which is called the ‘product cipher’.
5	<b>Explain Avalanche effect.</b> BTL1 A desirable property of any encryption algorithm is that a small change in either the plaintext or the key produce a significant change in the ciphertext

6	<p><b>Define Diffusion &amp; Confusion.</b> BTL1</p> <p><b>Diffusion:</b></p> <ul style="list-style-type: none"> <li>✓ In <b>diffusion</b>, the statistical structure of the plaintext is dissipated into long-range statistics of the ciphertext.</li> <li>✓ This is achieved by having each plaintext digit affect the value of many ciphertext digits; generally, this is equivalent to having each ciphertext digit be affected by many plaintext digits</li> </ul> <p><b>Confusion:</b></p> <p>It can be achieved by substitution algorithm. It is the relationship between cipher text and key.</p>				
7	<p><b>Give the five modes of operation of Block cipher.</b> BTL2</p> <ul style="list-style-type: none"> <li>✓ Electronic Codebook(ECB)</li> <li>✓ Cipher Block Chaining(CBC)</li> <li>✓ Cipher Feedback(CFB)</li> <li>✓ Output Feedback(OFB)</li> <li>✓ Counter(CTR)</li> </ul>				
8	<p><b>State advantages of counter mode.</b> BTL2</p> <ul style="list-style-type: none"> <li>✓ Hardware efficiency</li> <li>✓ Software efficiency</li> <li>✓ Preprocessing</li> <li>✓ Random access</li> <li>✓ Provable security</li> <li>✓ Simplicity</li> </ul>				
9	<p><b>Define Multiple Encryption</b> BTL2</p> <p>Multiple Encryption is a technique in which the encryption is used multiple times. Eg: Double DES, Triple DES</p>				
10	<p><b>Specify the design criteria of block cipher.</b> BTL4</p> <ul style="list-style-type: none"> <li>✓ Number of rounds</li> <li>✓ Design of the function F</li> <li>✓ Key scheduling</li> </ul>				
11	<p><b>Define Reversible mapping.</b> BTL5</p> <p>Each plain text is maps with the unique cipher text. This transformation is called reversible mapping</p>				
12	<p><b>Specify the basic task for defining a security service.</b> BTL6</p> <p>A service that enhances the security of the data processing systems and the information transfer of an organization. The services are intended to counter security attack, and they make use of one or more security mechanism to provide the service.</p>				
13	<p><b>What is the difference between link and end to end encryption?</b> BTL2</p> <table border="1" data-bbox="230 1634 1454 1848"> <thead> <tr> <th data-bbox="230 1634 687 1733"><b>Link Encryption</b></th><th data-bbox="687 1634 1454 1733"><b>End to End Encryption</b></th></tr> </thead> <tbody> <tr> <td data-bbox="230 1733 687 1848"> <ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul> </td><td data-bbox="687 1733 1454 1848"> <ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul> </td></tr> </tbody> </table>	<b>Link Encryption</b>	<b>End to End Encryption</b>	<ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul>	<ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul>
<b>Link Encryption</b>	<b>End to End Encryption</b>				
<ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul>	<ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul>				
6	<p><b>Define Diffusion &amp; Confusion.</b> BTL1</p> <p><b>Diffusion:</b></p> <ul style="list-style-type: none"> <li>✓ In <b>diffusion</b>, the statistical structure of the plaintext is dissipated into long-range statistics of the ciphertext.</li> <li>✓ This is achieved by having each plaintext digit affect the value of many ciphertext digits; generally, this is equivalent to having each ciphertext digit be affected by many plaintext digits</li> </ul> <p><b>Confusion:</b></p> <p>It can be achieved by substitution algorithm. It is the relationship between cipher text and key.</p>				
7	<p><b>Give the five modes of operation of Block cipher.</b> BTL2</p> <ul style="list-style-type: none"> <li>✓ Electronic Codebook(ECB)</li> <li>✓ Cipher Block Chaining(CBC)</li> <li>✓ Cipher Feedback(CFB)</li> <li>✓ Output Feedback(OFB)</li> <li>✓ Counter(CTR)</li> </ul>				
8	<p><b>State advantages of counter mode.</b> BTL2</p> <ul style="list-style-type: none"> <li>✓ Hardware efficiency</li> <li>✓ Software efficiency</li> <li>✓ Preprocessing</li> <li>✓ Random access</li> <li>✓ Provable security</li> <li>✓ Simplicity</li> </ul>				
9	<p><b>Define Multiple Encryption</b> BTL2</p> <p>Multiple Encryption is a technique in which the encryption is used multiple times. Eg: Double DES, Triple DES</p>				
10	<p><b>Specify the design criteria of block cipher.</b> BTL4</p> <ul style="list-style-type: none"> <li>✓ Number of rounds</li> <li>✓ Design of the function F</li> <li>✓ Key scheduling</li> </ul>				
11	<p><b>Define Reversible mapping.</b> BTL5</p> <p>Each plain text is maps with the unique cipher text. This transformation is called reversible mapping</p>				
12	<p><b>Specify the basic task for defining a security service.</b> BTL6</p> <p>A service that enhances the security of the data processing systems and the information transfer of an organization. The services are intended to counter security attack, and they make use of one or more security mechanism to provide the service.</p>				
13	<p><b>What is the difference between link and end to end encryption?</b> BTL2</p> <table border="1" data-bbox="230 1634 1454 1848"> <thead> <tr> <th data-bbox="230 1634 687 1733"><b>Link Encryption</b></th><th data-bbox="687 1634 1454 1733"><b>End to End Encryption</b></th></tr> </thead> <tbody> <tr> <td data-bbox="230 1733 687 1848"> <ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul> </td><td data-bbox="687 1733 1454 1848"> <ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul> </td></tr> </tbody> </table>	<b>Link Encryption</b>	<b>End to End Encryption</b>	<ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul>	<ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul>
<b>Link Encryption</b>	<b>End to End Encryption</b>				
<ul style="list-style-type: none"> <li>✓ With link encryption, each vulnerable communication link is</li> </ul>	<ul style="list-style-type: none"> <li>✓ With end to end encryption, encryption process is carried out at the two end systems</li> </ul>				

	equipped on both ends with an encryption device	
	✓ Message exposed in sending host and in intermediate nodes	✓ Message encrypted in sending and intermediate nodes
	✓ Transparent to user	✓ User applies encryption
	✓ Host maintains encryption facility	✓ Users must determine algorithm
	✓ One facility for all users	✓ Users selects encryption scheme
14	<b>What is traffic Padding? What is its purpose?</b> BTL2 Traffic padding produces ciphertext output continuously, the purpose of padding is that even in the absence of the plain text, a continuous random data stream is generated.	
15	<b>List the evaluation criteria defined by NIST for AES?</b> BTL5 The evaluation criteria for AES is as follows: ✓ Security ✓ Cost ✓ Algorithm and implementation characteristics	
16	<b>What is Triple Encryption? How many keys are used in triple encryption?</b> BTL4 Triple Encryption is a technique in which encryption algorithm is performed three times using three keys.	
17	<b>List the schemes for the distribution of public keys.</b> BTL3 ✓ Public announcement ✓ Publicly available directory ✓ Public key authority ✓ Public key certificates	
18	<b>Drawback of 3-DES.</b> BTL3 ✓ Algorithm is sluggish in software ✓ The number of rounds is thrice as that of DES ✓ 3DES uses 64 bit block size ✓ To have higher efficiency and security a larger block size is needed.	
19	<b>List out an evaluation criteria for round 2.</b> BTL1 ✓ General security ✓ Software implementation ✓ Hardware implementation ✓ Attacks ✓ Encryption Vs Decryption ✓ Key ability-Ability to change keys quickly with minimum of resources.	
20	<b>List out the attacks to RSA.</b> BTL2 ✓ <b>Brute force</b> - Trying all possible private keys. ✓ <b>Mathematical attacks</b> - The approaches to factor the product of two prime numbers.	

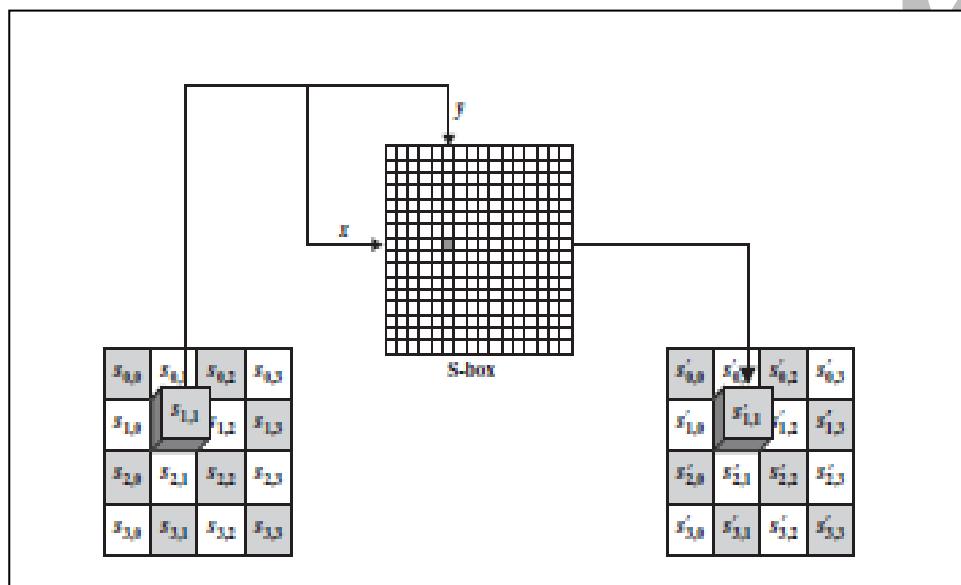
	<p>✓ <b>Timing attack</b> - Depends on the running time of the decryption algorithm</p>
21	<p><b>What is Primality Test? List the types of Primality Testing.</b> BTL1</p> <p>A primality test is an algorithm for determining whether an input number is prime or not.</p> <p>Types of Primality Test:</p> <p>Fermat Primality Test.</p> <p>Miller-Rabin Primality Test.</p> <p>Solovay-strassenPrimality Test.</p>
22	<p><b>What is Factoring ?</b>BTL1</p> <p>Factoring is the decomposition of an object into a product of other objects, or factors, which when multiplied together give the original.</p>
23	<p><b>Define RC4.</b> BTL2</p> <p>RC4 is a stream cipher designed in 1987 by Ron Rivest for RSA Security. RC4 is used in the SSL/TLS (Secure Sockets Layer/Transport Layer Security) standards that have been defined for communication between Web browsers and servers. It is also used in the WEP (Wired Equivalent Privacy) protocol and the newer WiFi Protected Access (WPA) protocol that are part of the IEEE 802.11 wireless LAN standard.</p>
24	<p><b>What is the meet in the middle attack?</b> BTL1</p> <p>This is the cryptanalytic attack that attempts to find the value in each of the range and domain of the composition of two functions such that the forward mapping of one through the first function is the same as the inverse image of the other through the second function-quite literally meeting in the middle of the composed function.</p>
25.	<p><b>List Four possible approaches to attack the RSA Algorithm.</b> BTL2</p> <p>Brute Force</p> <p>Mathematical Attacks</p> <p>Timing attacks</p> <p>Chosen Cipher text attacks</p>

**PART \* B**

- 1 **Explain Block cipher design principles and modes of operation. (Apr/May2014)**  
**(13M) BTL4**

**Answer:**Page no.:116 to 118 in William Stallings

- ✓ DES Design criteria (2M)
- ✓ Criteria for permutation(2M)
- ✓ Number of rounds(2M)
- ✓ S-Box design Diagram(2M)

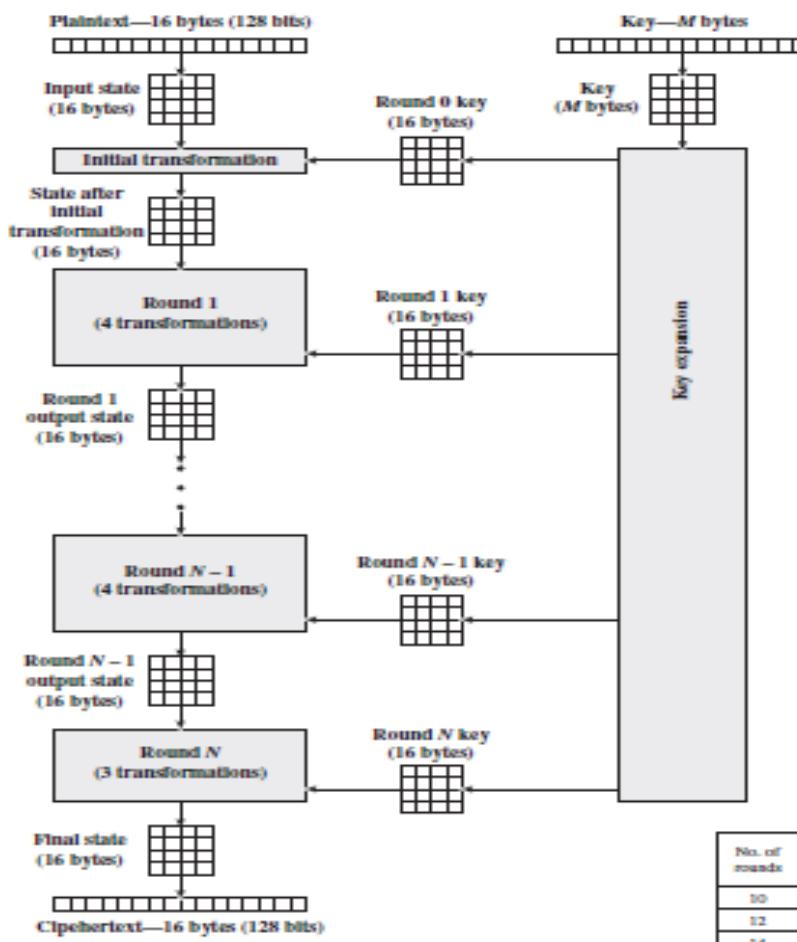


- ✓ Key schedule algorithm(2M)
- ✓ Electronic code book(2M)
- ✓ Cipher block chaining(1M)

- 2 **Explain about Advance Encryption Standard. (13M) BTL4**

**Answer:**Page no.:174 to 176 in William Stallings

- ✓ Evaluation criteria for AES (2M)
- ✓ Security(1M)
- ✓ Cost(1M)
- ✓ Implementation(1M)
- ✓ AES cipher(2M)
- ✓ AES Encryption and decryption(2M)



- ✓ Inverse substitution bytes(1M)
- ✓ Mix column(1M)
- ✓ Key expansion in AES 192 and AES 256 (2M)

3 Explain Triple DES. (13M) BTL4

**Answer:**Pageno.:101 to 103 in William Stallings

- ✓ Triple DES with 2 keys(3M)
- ✓ Diagram with formulas(4M)

$$\begin{aligned} C &= E(K_1, D(K_2, E(K_1, P))) \\ P &= D(K_1, E(K_2, D(K_1, C))) \end{aligned}$$

- ✓ Triple DES with 3 keys(4M)
- ✓ Diagram with operation formulas(2M)

$$C = E(K_3, D(K_2, E(K_1, P)))$$

4 Explain about RC4 algorithm. (13M) BTL4 (May/June/2012)

**Answer:**Page no.PPT given covering all topics

	<ul style="list-style-type: none"> <li>✓ Explain RC4 algorithm (2M)</li> <li>✓ Parameters (2M)</li> <li>✓ Primitive operators(1M)</li> <li>✓ Characteristics(1M)</li> <li>✓ Key expansion(2M)</li> <li>✓ Encryption(1M)</li> <li>✓ Decryption(1M)</li> <li>✓ RC4 modes (3M)</li> </ul>
5	<p><b>Explain public key algorithm. (13M) BTL4</b></p> <p><b>Answer:Page no.:293 to 295 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Public key algorithm(2M)</li> <li>✓ Characteristics of public key cryptography(2M)</li> <li>✓ Six ingredients(3M)</li> <li>✓ Decryption algorithm diagram(2M)</li> <li>✓ Steps to create public key(2M)</li> <li>✓ Diagrammatical representation(2M)</li> </ul>
6	<p>i)<b>Explain RSA algorithm. (6M) BTL4 (Apr/May 2011,Nov/Dec 2011,2012)</b></p> <p><b>Answer:Page no.:301 to 309 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Explain the process in mathematical fact(2M)</li> <li>✓ Choose, select, encrypt, transfer cipher text, decrypt.(3M)</li> <li>✓ Discuss with an example(1M)</li> </ul> <p>ii)<b>Explain blowfish encryption algorithm. (7M) BTL4</b></p> <p><b>Answer:Page no.:119 to 120 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Algorithm(1M)</li> <li>✓ Feistel network(3M)</li> <li>✓ Working methodology(2M)</li> <li>✓ Example with diagram(1M)</li> </ul>
7	<p><b>Explain Diffie Hellman key exchange. (13M) BTL4</b></p> <p><b>Answer:Page no.:325 to 327 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Key management techniques(3M)</li> <li>✓ Explain Diffie Hellman algorithm with steps (3M)</li> <li>✓ Provide diagrammatical explanation with example(4M)</li> <li>✓ Process explanation with steps and diagram(3M)</li> </ul>
8	<p><b>Explain Elliptical Curve cryptography. (13M) BTL4</b></p> <p><b>Answer:Page no.:341 to 343 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Elliptical curve over <math>Z_p</math>(4M)</li> </ul>

	<ul style="list-style-type: none"> <li>✓ Equation of elliptical curve over <math>Zp</math>(3M)</li> <li>✓ Elliptic curves over <math>GF(2^m)</math> (3M)</li> <li>✓ Elliptic curve cryptography(3M)</li> </ul>
<b>PART * C</b>	
1	<p><b>Explain block cipher principles and modes of operation. (15M) BTL4</b></p> <p><b>Answer:</b>Page no.:216 to 218 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ DES design (3M)</li> <li>✓ CBC mode (3M)</li> <li>✓ AES (3M)</li> <li>✓ Triple DES (3M)</li> <li>✓ RC 5 Algorithm(3M)</li> </ul>
2	<p><b>Explain Public Key cryptography. (15M) BTL4</b></p> <p><b>Answer:</b>Page no.:290 to 292 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ Public Key Cryptography (2M)</li> </ul> <p><b>Public-key cryptography</b>, or <b>asymmetric cryptography</b>, is any cryptographic system that uses pairs of keys:<i>public keys</i> which may be disseminated widely, and <i>private keys</i> which are known only to the owner</p> <ul style="list-style-type: none"> <li>✓ Characteristics (3M)</li> <li>✓ Six ingredients with explanation (5M)</li> <li>✓ Diagrams (2M)</li> <li>✓ Steps (3M)</li> </ul>
3	<p><b>Explain DES in detail . (15M) BTL4</b></p> <p><b>Answer:</b>Pageno.:101 to 108 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ Definition (3M)</li> </ul> <p><b>DES</b> key length and brute-force attacks. The Data Encryption Standard is a block cipher, <b>meaning</b> a cryptographic key and algorithm are applied to a block of data simultaneously rather than one bit at a time</p> <ul style="list-style-type: none"> <li>✓ Structure (6M)</li> <li>✓ Diagrams (6M)</li> </ul>

4	<p><b>Evaluate encryption and decryption using RSA algorithm for the following. <math>p=7</math>, <math>q=11</math>; <math>e=17</math>; <math>m=8</math>. (15M) BTL6</b></p> <p><b>Answer:</b>Page no.:247 to 249 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ computing their system modulus <math>N=p \cdot q</math> (2M)</li> <li>✓ note <math>\phi(N)=(p-1)(q-1)</math></li> <li>✓ STEPS(3M) <ul style="list-style-type: none"> <li>✓ to encrypt a message <math>M</math> the sender:</li> <li>✓ obtains public key of recipient <math>KU=\{e, N\}</math></li> <li>✓ computes: <math>C=M^e \bmod N</math>, where <math>0 \leq M &lt; N</math></li> <li>✓ to decrypt the ciphertext <math>C</math> the owner:</li> <li>✓ uses their private key <math>KR=\{d, p, q\}</math></li> <li>✓ computes: <math>M=C^d \bmod N</math></li> <li>✓ Encryption(5M)</li> <li>✓ Decryption(5M)</li> </ul> </li> </ul>
5	<p><b>Evaluate using Diffie-Hellman key exchange technique. Users A and B use a common prime <math>q=11</math> and a primitive root <math>\alpha=7</math>.</b></p> <p>(i) If user A has private key <math>X_A=3</math>. What is A's public key <math>Y_A</math>?  (ii) If user B has private key <math>X_B=6</math>. What is B's public key <math>Y_B</math>?  (iii) What is the shared secret key? Also write the algorithm. (15M) BTL6</p> <p><b>Answer:</b>Page no.:208 to 211in William Stallings</p> <ul style="list-style-type: none"> <li>✓ prime <math>p</math>, element <math>g \in Z_p^*</math> (5M) <math display="block">h_A = g^x \bmod p</math> <math display="block">h_B = g^y \bmod p</math> </li> <li>✓ Encryption(5M)</li> </ul>

	<ul style="list-style-type: none"> <li>✓ Decryption(5M)</li> </ul>
6	<p><b>Estimate the encryption and decryption values for the RSA algorithm parameters. P=3, Q=11, E=7, d=? , M=5. (15M) BTL6</b></p> <p><b>Answer:</b>Page no.:247 to 249 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ computing their system modulus <math>N=p \cdot q</math> (2M)</li> <li>✓ note <math>\phi(N)=(p-1)(q-1)</math></li> <li>✓ STEPS(3M)</li> <li>✓ to encrypt a message M the sender:</li> <li>✓ obtains public key of recipient <math>KU=\{e,N\}</math></li> <li>✓ computes: <math>C=M^e \text{ mod } N</math>, where <math>0 \leq M &lt; N</math></li> <li>✓ to decrypt the ciphertext C the owner:</li> <li>✓ uses their private key <math>KR=\{d,p,q\}</math></li> <li>✓ computes: <math>M=C^d \text{ mod } N</math></li> <li>✓ Encryption(5M)</li> <li>✓ Decryption(5M)</li> </ul>
7	<p><b>Implement RSA Algorithm for the given values, trace the sequence of calculations in RSA. P=7,q=13,e=5 and M=10. (15M) BTL5</b></p> <p><b>Answer:</b>Page no.:247 to 248 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ computing their system modulus <math>N=p \cdot q</math> (2M)</li> <li>✓ note <math>\phi(N)=(p-1)(q-1)</math></li> <li>✓ STEPS(3M)</li> <li>✓ to encrypt a message M the sender:</li> <li>✓ obtains public key of recipient <math>KU=\{e,N\}</math></li> <li>✓ computes: <math>C=M^e \text{ mod } N</math>, where <math>0 \leq M &lt; N</math></li> <li>✓ to decrypt the ciphertext C the owner:</li> <li>✓ uses their private key <math>KR=\{d,p,q\}</math></li> <li>✓ computes: <math>M=C^d \text{ mod } N</math></li> <li>✓ Encryption(5M)</li> <li>✓ Decryption(5M)</li> </ul>
8	<p><b>Users Alice and Bob use the Diffie Hellman Key exchange technique with a common prime <math>q=83</math> and primitive root <math>\alpha = 5</math>.</b></p> <ol style="list-style-type: none"> <li><b>if Alice has a private key <math>X_A=6</math>,what is the Alice's public key <math>Y_A</math>?</b></li> <li><b>If Bob has a private key <math>X_B=10</math>, what is Bob's public key <math>Y_B</math>?</b></li> <li><b>what is the shared secret key? (15M) BTL6</b></li> </ol> <p><b>Answer:</b>Page no.:325 to 329 in William Stallings</p> <ul style="list-style-type: none"> <li>✓ prime <math>p</math>, element <math>g \in \mathbb{Z}_p^*</math> (5M) <ul style="list-style-type: none"> <li><math>h_A = g^x \text{ mod } p</math></li> <li><math>h_B = g^y \text{ mod } p</math></li> </ul> </li> <li>✓ Encryption(5M)</li> <li>✓ Decryption(5M)</li> </ul>

--	--

### UNIT 3- HASH FUNCTIONS AND DIGITAL SIGNATURES

**Authentication requirement – Authentication function – MAC – Hash function – Security of hash function and MAC –MD5 - SHA - HMAC – CMAC - Digital signature and authentication protocols – DSS – EI Gamal – Schnorr**

#### PART \* A

1	<b>What is message authentication? BTL1</b>  It is a procedure that verifies whether the received message comes from assigned source has not been altered. It uses message authentication codes, hash algorithms to authenticate the message
2	<b>Define the classes of message authentication function. BTL1</b>  Message encryption: The entire cipher text would be used for authentication. Message Authentication Code: It is a function of message and secret key produce a fixed length value  Hash function: Some function that map a message of any length to fixed length which serves as authentication
3	<b>What are the requirements for message authentication? BTL1</b> The requirements for message authentication are <ul style="list-style-type: none"> <li>✓ Disclosure</li> <li>✓ Traffic analysis</li> <li>✓ Content modification</li> <li>✓ Sequence modification</li> <li>✓ Masquerade</li> <li>✓ Timing modification</li> <li>✓ Source repudiation</li> <li>✓ Destination repudiation</li> </ul>
4	<b>What do you mean by hash function? BTL4</b>  Hash function accept a variable size message M as input and produces a fixed size hash code H(M) called as message digest as output. It is the variation on the message authentication code
5	<b>Differentiate MAC and Hash function. BTL3</b>  MAC: In Message Authentication Code, the secret key shared by sender and receiver. The MAC is appended to the message at the source at a time which the message is assumed or known to be correct.  Hash Function:

	The hash value is appended to the message at the source at time when the message is assumed or known to be correct. The hash function itself not considered to be secret				
6	<b>Give any three hash algorithm.</b> BTL4 <ul style="list-style-type: none"> <li>✓ MD5 (Message Digest version 5) algorithm.</li> <li>✓ SHA_1 (Secure Hash Algorithm).</li> <li>✓ RIPEMD_160 algorithm.</li> </ul>				
7	<b>What are the requirements of the hash function?</b> BTL3 <ul style="list-style-type: none"> <li>✓ H can be applied to a block of data of any size.</li> <li>✓ H produces a fixed length output.</li> <li>✓ <math>H(x)</math> is relatively easy to compute for any given <math>x</math>, making both hardware and software implementations practical.</li> </ul>				
8	<b>What do you mean by MAC?</b> BTL3 MAC is Message Authentication Code. It is a function of message and secret key which produce a fixed length value called as MAC. $MAC = C_k(M)$ <p style="text-align: center;">Where <math>M</math> = variable length message</p> <p style="text-align: center;">K = secret key shared by sender and receiver.</p> <p style="text-align: center;"><math>CK(M)</math> = fixed length authenticator.</p>				
9	<b>Differentiate internal and external error control.</b> BTL3 <b>Internal error control:</b> In internal error control, an error detecting code also known as frame check sequence or checksum. <b>External error control:</b> In external error control, error detecting codes are appended after encryption.				
10	<b>What is the meet in the middle attack?</b> BTL2 This is the cryptanalytic attack that attempts to find the value in each of the range and domain of the composition of two functions such that the forward mapping of one through the first function is the same as the inverse image of the other through the second function-quite literally meeting in the middle of the composed function.				
11	<b>What is the role of compression function in hash function?</b> BTL2 The hash algorithm involves repeated use of a compression function $f$ , that takes two inputs and produce a $n$ -bit output. At the start of hashing the chaining variable has an initial value that is specified as part of the algorithm. The final value of the chaining variable is the hash value usually $b > n$ ; hence the term compression.				
12	<b>What is the difference between weak and strong collision resistance?</b> BTL2 <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; padding: 5px;"><b>Weak collision resistance</b></th> <th style="text-align: center; padding: 5px;"><b>Strong resistance collision</b></th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">For any given block <math>x</math>, it is computationally infeasible to fine <math>y \neq x</math> wit <math>H(y)=H(x)</math>.</td> <td style="padding: 5px;">It is computationally infeasible to find any pair <math>(x,y)</math> such that <math>H(x)=H(y)</math>.</td> </tr> </tbody> </table>	<b>Weak collision resistance</b>	<b>Strong resistance collision</b>	For any given block $x$ , it is computationally infeasible to fine $y \neq x$ wit $H(y)=H(x)$ .	It is computationally infeasible to find any pair $(x,y)$ such that $H(x)=H(y)$ .
<b>Weak collision resistance</b>	<b>Strong resistance collision</b>				
For any given block $x$ , it is computationally infeasible to fine $y \neq x$ wit $H(y)=H(x)$ .	It is computationally infeasible to find any pair $(x,y)$ such that $H(x)=H(y)$ .				

	It is proportional to $2^n$	It is proportional to $2^{n/2}$																																
13	<b>Compare MD5, SHA1 and RIPEMD-160 algorithm.BTL4</b>	<table border="1"> <thead> <tr> <th></th><th>MD5</th><th>SHA-1</th><th>RIPEMD160</th></tr> </thead> <tbody> <tr> <td>Digest length</td><td>128 bits</td><td>160 bits</td><td>160 bits</td></tr> <tr> <td>Basic unit of processing</td><td>512 bits</td><td>512 bits</td><td>512 bits</td></tr> <tr> <td>No of steps</td><td>64(4 rounds of 16)</td><td>80(4 rounds of 20)</td><td>160(5 pairs rounds of 16)</td></tr> <tr> <td>Maximum message size</td><td>infinity</td><td><math>2^{64}-1</math> bits</td><td><math>2^{64}-1</math> bits</td></tr> <tr> <td>Primitive logical function</td><td>4</td><td>4</td><td>5</td></tr> <tr> <td>Additive constants Used</td><td>64</td><td>4</td><td>9</td></tr> <tr> <td>Endianess</td><td>Little endian</td><td>Big endian</td><td>Little endian</td></tr> </tbody> </table>		MD5	SHA-1	RIPEMD160	Digest length	128 bits	160 bits	160 bits	Basic unit of processing	512 bits	512 bits	512 bits	No of steps	64(4 rounds of 16)	80(4 rounds of 20)	160(5 pairs rounds of 16)	Maximum message size	infinity	$2^{64}-1$ bits	$2^{64}-1$ bits	Primitive logical function	4	4	5	Additive constants Used	64	4	9	Endianess	Little endian	Big endian	Little endian
	MD5	SHA-1	RIPEMD160																															
Digest length	128 bits	160 bits	160 bits																															
Basic unit of processing	512 bits	512 bits	512 bits																															
No of steps	64(4 rounds of 16)	80(4 rounds of 20)	160(5 pairs rounds of 16)																															
Maximum message size	infinity	$2^{64}-1$ bits	$2^{64}-1$ bits																															
Primitive logical function	4	4	5																															
Additive constants Used	64	4	9																															
Endianess	Little endian	Big endian	Little endian																															

14	<b>Distinguish between direct and arbitrated digital signature. BTL 3</b>	
	<b>Direct digital signature</b>	<b>Arbitrated Digital Signature</b>
	<ul style="list-style-type: none"> <li>✓ The direct digital signature involves only the communicating parties</li>   <li>✓ This may be formed by encrypting the entire message with the sender's private key</li> </ul>	<ul style="list-style-type: none"> <li>✓ The arbiter plays a sensitive and crucial role in this digital signature</li>   <li>✓ Every signed message from a sender x to a receiver y goes first to an arbiter A, who subjects the message and its signature to a number of test to check its origin and content</li> </ul>
15	<b>What are the properties a digital signature should have? BTL1</b>	
	<ul style="list-style-type: none"> <li>✓ It must verify the author and the date and time of signature.</li> <li>✓ It must authenticate the contents at the time of signature.</li> <li>✓ It must be verifiable by third parties to resolve disputes.</li> </ul>	
16	<b>What are the applications in RC4 algorithm? BTL1</b>	
	<ul style="list-style-type: none"> <li>✓ WEP Protocol</li> <li>✓ LAN Networks</li> </ul>	
17	<b>How is the security of a MAC function expressed? BTL3</b>	
	<ul style="list-style-type: none"> <li>✓ Variable input size</li> <li>✓ Fixed output size</li> <li>✓ Efficiency</li> <li>✓ Preimage resistant (one-way property)</li> <li>✓ Second preimage resistant (weak Collision resistant)</li> <li>✓ Collision Resistant (Strong Collision Resistant)</li> <li>✓ Pseudorandomness</li> </ul>	
18	<b>Mention the significance of Signature function in DSS. BTL4</b>	
	<p>The signature function also depends on the sender's private key (<math>PRa</math>) and a set of parameters known to a group of communicating principals. The signature function is such that only the sender, with knowledge of the private key, could have produced the valid signature.</p>	
19	<b>What is Elliptic curve? BTL1</b>	
	<p>An elliptic curve is defined by an equation in two variables with coefficients. For cryptography, the variables and coefficients are restricted to elements in a finite field, which results in the definition of a finite abelian group.</p>	
20	<b>What are the two approaches of digital signatures? BTL1</b>	
	<ul style="list-style-type: none"> <li>✓ It must verify the author and the date and time of the signature.</li> <li>✓ It must authenticate the contents at the time of the signature.</li> <li>✓ It must be verifiable by third parties, to resolve disputes.</li> </ul>	
21	<b>What are the uses of RC4? BTL1</b>	

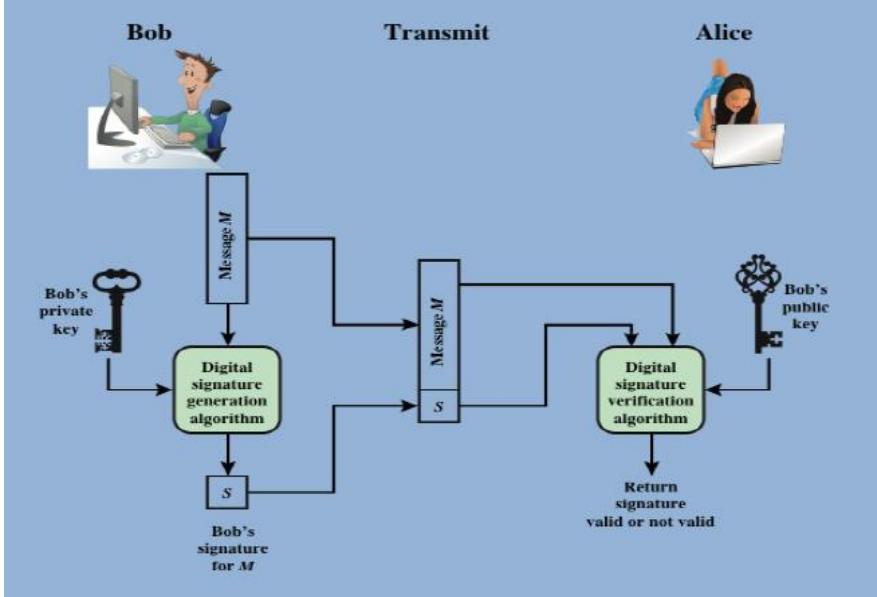
	<ul style="list-style-type: none"> <li>✓ Remarkably Simple And Quite Easy To Explain</li> <li>✓ RC4 Is Used In The Wifi Protected Access (WPA) Protocol That Are Part Of The Ieee 802.11 Wireless Lan Standard</li> <li>✓ RC4 Was Kept As A Trade Secret By RSA Security.</li> </ul>
22	<b>What are the security services provided by Digital Signature? BTL1</b> <ul style="list-style-type: none"> <li>✓ MD5</li> <li>✓ SHA</li> </ul>
23	<b>What is Direct Digital Signature? BTL1</b> The term <b>direct digital signature</b> refers to a digital signature scheme that involves only the communicating parties (source, destination). It is assumed that the destination knows the public key of the source.
24	<b>What are the requirements of Digital Signature? BTL1</b> <ul style="list-style-type: none"> <li>✓ The signature must be a bit pattern</li> <li>✓ The signature must use some information</li> <li>✓ Signature must be relatively easy to produce the digital signature.</li> <li>✓ Signature must be relatively easy to recognize and verify the digital signature.</li> <li>✓ Signature must be computationally infeasible to forge a digital signature.</li> <li>✓ Signature must be practical to retain a copy of the digital signature in storage.</li> </ul>
25	<b>What is Schnorr Digital Signature Scheme? BTL1</b> The Schnorr signature scheme is based on discrete logarithms [SCHN89, SCHN91]. The Schnorr scheme minimizes the message-dependent amount of computation required to generate a signature. The main work for signature generation does not depend on the message and can be done during the idle time of the processor.

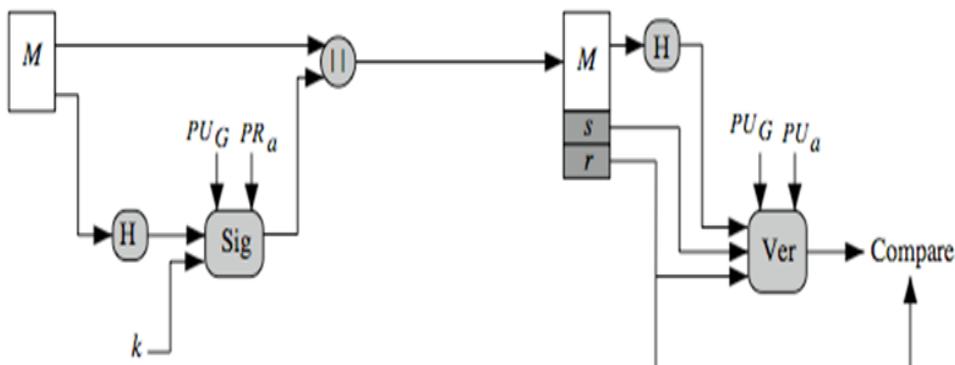
### PART \*B

1	<b>Explain Hash function. (13M) BTL4 (AU Nov/Dec 2012 )</b> <b>Answer:Page no.:351 to 353 in William Stallings</b> <ul style="list-style-type: none"> <li>✓ Authentication function (2M) is a short piece of information used to authenticate a message—in other words, to confirm that the message came from the stated sender (its authenticity) and has not been changed.</li> <li>✓ Hash Function function (2M) A hash function maps a variable-length data block or message into a fixed-length value called a hashcode. A variation on the message authentication code is the one way hash function. As with MAC, a hash function accepts a variable size message M as input and produces a fixed-size output, referred to as hash code H(M). <ul style="list-style-type: none"> <li>✓ Write in detail about MAC(2M)</li> <li>✓ Derive the steps(1 M)</li> <li>✓ Diagrams and cases(2M)</li> </ul> </li> </ul>
---	---

	<p>✓ Derive the steps(2M) ✓ Diagrams and cases(2M)</p>
2	<b>Explain MD5 algorithm. (13M) BTL4 (AU May/June 2012, Apr/May 2011)</b> Answer: Pageno.:353 to 355 in William Stallings <p>✓ Basic properties of MD5 algorithm(2M)  ✓ Padding(2M)  ✓ Append value(1M)  ✓ Divide input into 512 bit blocks(1M)  ✓ Initializing chaining variables(2M)  ✓ Process blocks(2M)</p> <p>✓ Processing of rounds (3M)</p>
3	<b>Explain Secure Hash algorithm. (13M) BTL4 (Nov/Dec 2014, April/May 2013)</b> Answer: Pageno.:366 to 368 in William Stallings <p>✓ Elaboration of Secure Hash algorithm(3M) <ul style="list-style-type: none"> <li>SHA was designed by NIST &amp; NSA in 1993, revised 1995 as SHA-1</li> <li>US standard for use with DSA signature scheme</li> <li>standard is FIPS 180-1 1995, also Internet RFC3174</li> <li><b>note:</b> the algorithm is SHA, the standard is SHS</li> <li>produces 160-bit hash values</li> </ul> </p>

	<ul style="list-style-type: none"> <li>✓ Obtain original message(2M)</li> <li>✓ Find same message digest in SHA512(2M)</li> <li>✓ Explain the algorithm with steps(3M)</li> <li>✓ Example with structural diagram(3M)</li> </ul>
4	<p><b>Explain Hash Based Message Authentication Code and CMAC .(13M) BTL4</b></p> <p><b>Answer:</b>Pageno.:399 to 401in William Stallings</p> <ul style="list-style-type: none"> <li>✓ Structural diagram of HMAC(2M)  </li> <li>✓ Expand all the functions(2M)</li> <li>✓ Explain with steps(1M)</li> <li>✓ Creating length of two phases(3M)</li> <li>✓ Appending (1M)</li> <li>✓ Producing b-bit block(2M)</li> </ul>

	<ul style="list-style-type: none"> <li>✓ Security of HMAC (2M)</li> </ul>
5	<p><b>Explain Digital Signatures.(13M) BTL3 (AU Nov/Dec 2011, May/June 2014)</b>  <b>Answer:Page no.:420 to 422 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Digital signature mechanism (2M)</li> </ul>  <ul style="list-style-type: none"> <li>✓ Requirements of Digital signature (3M)</li> <li>✓ Types of Approaches(2M)</li> <li>✓ Digital signature types (3M)</li> <li>✓ Direct digital signature(1M)</li> <li>✓ Arbitrated digital signature (2M)</li> </ul>
6	<p><b>Explain Authentication protocol. (13M) BTL4</b>  <b>Answer: Page no.:386 to 389 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Mutual authentication(3M)</li> <li>✓ Examples of Replay attack(1M)</li> <li>✓ Symmetric encryption Approach(3M)</li> <li>✓ Time stamps(1M)</li> <li>✓ One say authentication(2M)</li> <li>✓ Public key encryption approach (3M)</li> </ul>
7	<p><b>Explain digital signature standard. (13M) BTL4 (AU May/June 2014)</b>  <b>Answer: Page no.:427 to 429 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Digital signature algorithm designed to provide digital signature (3M)</li> <li>✓ Diagrammatical expansion(3M)</li> <li>✓ Digital signature algorithm(4M)</li> <li>✓ Diagrammatical expansion of algorithm with various stages (3M)</li> </ul>



8 Explain in detail about EL-GAMAL Algorithm. (13M) BTL4 (AU Nov/Dec 2013, May/June 2015)

**Answer:**Page no.:424to 426 in William Stallings

- ✓ Public key crypto system based on concept of Diffie-Hellman key management(1M)
- ✓ Components(2M)
- ✓ El Gamal Key generation and steps to generate private and public keys(3M)
- ✓ El Gamal Encryption(3M)
- ✓ El Gamal Decryption(3M)
- ✓ Proof of Decryption(1M)

### PART \* C

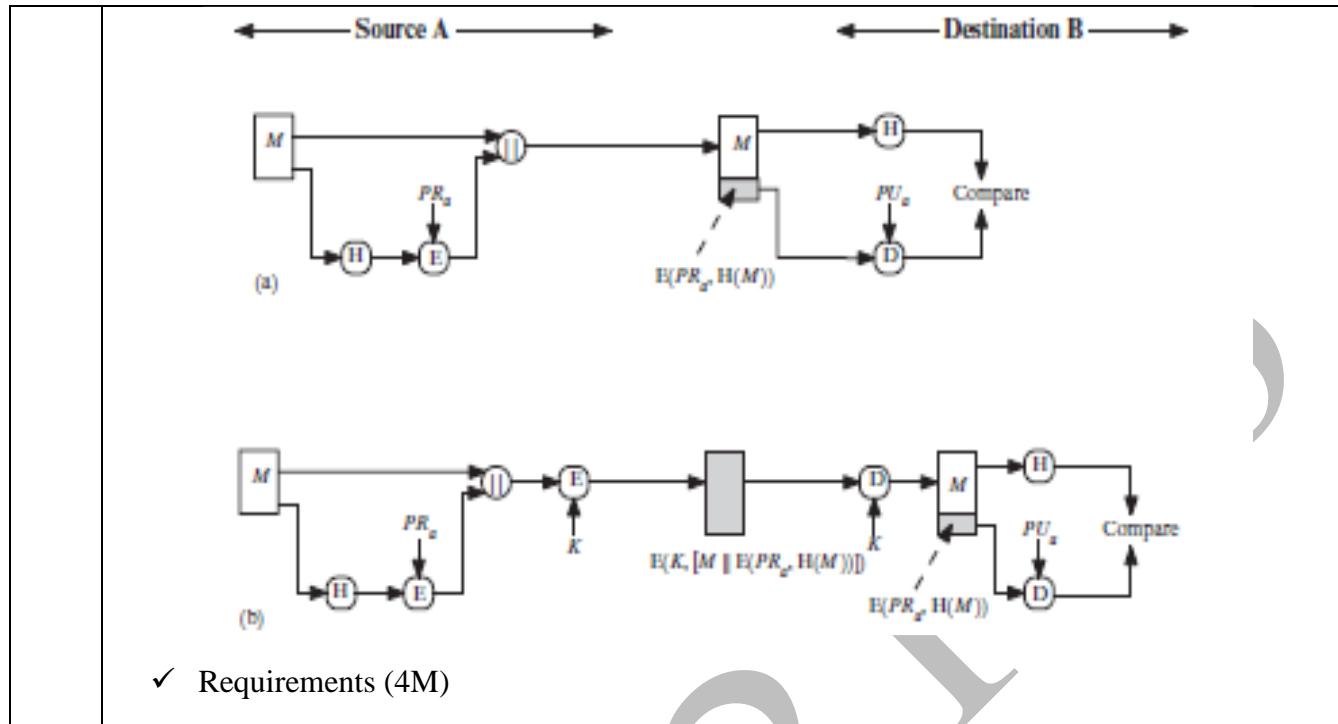
1 Explain message authentication requirement and its functions. (15M) BTL4  
**Answer:**Pageno.:386 to 389in William Stallings

- ✓ Message requirement (3M)
- ✓ Message functions (3M)
- ✓ Security of MAC (3M)
- ✓ Key terms and reviews (3M)
- ✓ Authenticated encryption (3M)

2 Explain Hash function and its Security. (15M) BTL4  
**Answer:**Pageno.:351 to 354in William Stallings

- ✓ Definition (2M)

	<p>A <b>hash function</b> <math>H</math> accepts a variable-length block of data <math>M</math> as input and produces a fixed-size hash value <math>h = H(M)</math>. A “good” hash function has the property that the results of applying the function to a large set of inputs will produce outputs that are evenly distributed and apparently random.</p> <ul style="list-style-type: none"> <li>✓ Applications (3M) <ul style="list-style-type: none"> <li>• Used to create a one-way password file.</li> <li>• Used for intrusion detection and virus detection</li> <li>• Used to construct a pseudorandom function (PRF) or a pseudorandom number generator (PRNG).</li> </ul> </li> <li>✓ Simple Hash Functions(3M)</li> <li>✓ Requirements and Security (3M)</li> </ul>																
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; background-color: #f2f2f2;">Requirement</th><th style="text-align: center; background-color: #f2f2f2;">Description</th></tr> </thead> <tbody> <tr> <td>Variable input size</td><td><math>H</math> can be applied to a block of data of any size.</td></tr> <tr> <td>Fixed output size</td><td><math>H</math> produces a fixed-length output.</td></tr> <tr> <td>Efficiency</td><td><math>H(x)</math> is relatively easy to compute for any given <math>x</math>, making both hardware and software implementations practical.</td></tr> <tr> <td>Preimage resistant (one-way property)</td><td>For any given hash value <math>h</math>, it is computationally infeasible to find <math>y</math> such that <math>H(y) = h</math>.</td></tr> <tr> <td>Second preimage resistant (weak collision resistant)</td><td>For any given block <math>x</math>, it is computationally infeasible to find <math>y \neq x</math> with <math>H(y) = H(x)</math>.</td></tr> <tr> <td>Collision resistant (strong collision resistant)</td><td>It is computationally infeasible to find any pair <math>(x, y)</math> with <math>x \neq y</math>, such that <math>H(x) = H(y)</math>.</td></tr> <tr> <td>Pseudorandomness</td><td>Output of <math>H</math> meets standard tests for pseudorandomness.</td></tr> </tbody> </table>	Requirement	Description	Variable input size	$H$ can be applied to a block of data of any size.	Fixed output size	$H$ produces a fixed-length output.	Efficiency	$H(x)$ is relatively easy to compute for any given $x$ , making both hardware and software implementations practical.	Preimage resistant (one-way property)	For any given hash value $h$ , it is computationally infeasible to find $y$ such that $H(y) = h$ .	Second preimage resistant (weak collision resistant)	For any given block $x$ , it is computationally infeasible to find $y \neq x$ with $H(y) = H(x)$ .	Collision resistant (strong collision resistant)	It is computationally infeasible to find any pair $(x, y)$ with $x \neq y$ , such that $H(x) = H(y)$ .	Pseudorandomness	Output of $H$ meets standard tests for pseudorandomness.
Requirement	Description																
Variable input size	$H$ can be applied to a block of data of any size.																
Fixed output size	$H$ produces a fixed-length output.																
Efficiency	$H(x)$ is relatively easy to compute for any given $x$ , making both hardware and software implementations practical.																
Preimage resistant (one-way property)	For any given hash value $h$ , it is computationally infeasible to find $y$ such that $H(y) = h$ .																
Second preimage resistant (weak collision resistant)	For any given block $x$ , it is computationally infeasible to find $y \neq x$ with $H(y) = H(x)$ .																
Collision resistant (strong collision resistant)	It is computationally infeasible to find any pair $(x, y)$ with $x \neq y$ , such that $H(x) = H(y)$ .																
Pseudorandomness	Output of $H$ meets standard tests for pseudorandomness.																
3	<p>✓ Cipher block chaining (4M)</p> <p><b>Explain Digital Signature and functions.(15M) BTL4</b>  <b>Answer: Pageno.:420 to 422 in William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Explanation (2M) <p>The operation of the digital signature is similar to that of the MAC. In the case of the digital signature, the hash value of a message is encrypted with a user's private key.</p> </li> <li>✓ Properties (3M)</li> <li>✓ Attacks and forgeries (2M)</li> <li>✓ Diagrams (4M)</li> </ul>																



## UNIT-4 SECURITY PRACTICE & SYSTEM SECURITY

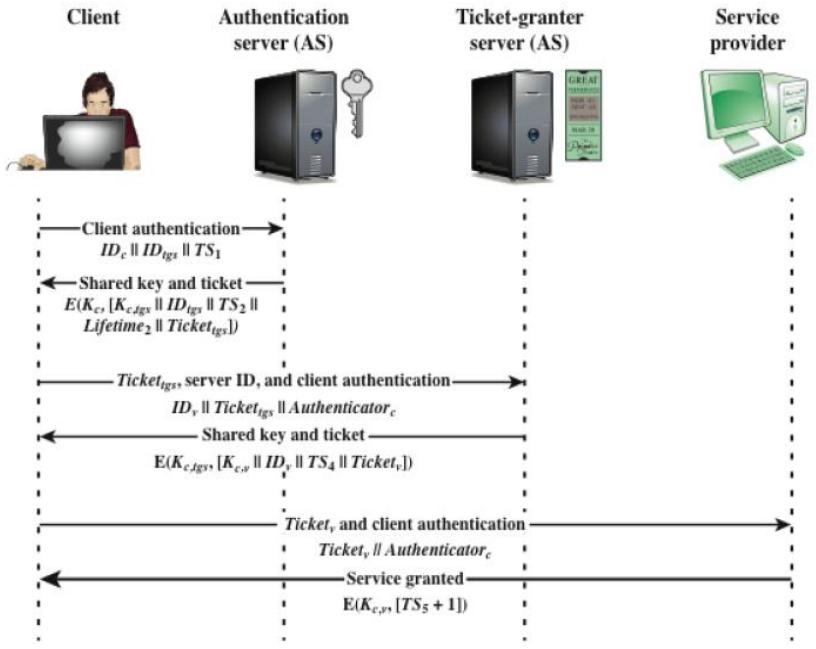
**Authentication applications – Kerberos – X.509 Authentication services - Internet Firewalls for Trusted System: Roles of Firewalls – Firewall related terminology- Types of Firewalls - Firewall designs - SET for E-Commerce Transactions. Intruder – Intrusion detection system – Virus and related threats – Countermeasures – Firewalls design principles – Trusted systems – Practical implementation of cryptography and security**

### PART \* A

1	<b>Define Kerberos.</b> BTL1 Kerberos is an authentication service developed as part of project Athena at MIT. The problem that Kerberos address is, assume an open distributed environment in which users at work stations wish to access services on servers distributed throughout the network.
2	<b>What is Kerberos? Write its uses.</b> BTL2 Kerberos is an authentication service developed as a part of project Athena at MIT. Kerberos provides a centralized authentication server whose functions are to authenticate servers.
3	<b>What are the requirements defined by Kerberos?</b> BTL1 <ul style="list-style-type: none"><li>✓ Secure</li><li>✓ Reliable</li><li>✓ Transparent</li><li>✓ Scalable</li></ul>
4	<b>In the content of Kerberos, What is realm?</b> BTL1 <ul style="list-style-type: none"><li>✓ A full service Kerberos environment consisting of a Kerberos server, a no. of clients, no.of application server requires the following</li><li>✓ The Kerberos server must have user ID and hashed password of all participating users in its database.</li><li>✓ The Kerberos server must share a secret key with each server. Such an environment is referred to as “Realm”.</li></ul>
5	<b>What is the purpose of X.509 standard?</b> BTL1 X.509 defines framework for authentication services by the X.500 directory to its users.X.509 defines authentication protocols based on public key certificates.
6	<b>List the 3 classes of intruder.</b> BTL2 <ul style="list-style-type: none"><li>✓ Masquerader</li><li>✓ Misfeasor</li><li>✓ Clandestine user</li></ul>
7	<b>Define virus. Specify the types of viruses.</b> BTL1 A virus is a program that can infect other program by modifying them the modification includes a copy of the virus program, which can then go on to infect other program. Types: <ul style="list-style-type: none"><li>✓ Parasitic virus</li><li>✓ Memory-resident virus</li><li>✓ Boot sector virus</li></ul>

	<ul style="list-style-type: none"> <li>✓ Stealth virus</li> <li>✓</li> <li>✓ Polymorphic virus</li> </ul>								
8	<p><b>What is application level gateway? BTL2</b></p> <p>An application level gateway also called a proxy server; act as a relay of application-level traffic. The user contacts the gateway using a TCP/IP application, such as Telnet or FTP, and the gateway asks the user for the name of the remote host to be accessed.</p>								
9	<p><b>List the design goals of firewalls. BTL1</b></p> <ul style="list-style-type: none"> <li>✓ All traffic from inside to outside, and vice versa, must pass through the firewall.</li> <li>✓ Only authorized traffic, as defined by the local security policy, will be allowed to pass.</li> <li>✓ The firewall itself is immune to penetration</li> </ul>								
10	<p><b>What are the steps involved in SET Transaction? BTL2</b></p> <p>The customer opens an account      The customer receives a certificate      Merchants have their own certificate      The customer places an order.      The merchant is verified.      The order and payment are sent.      The merchant requests payment authorization.      The merchant confirms the order.      The merchant provides the goods or services.      The merchant requests payment.</p>								
11	<p><b>What is dual signature? Write its purpose. BTL2</b></p> <p>The purpose of the dual signature is to link two messages intended for two different recipients. To avoid misplacement of orders.</p>								
12	<p><b>What is the need for authentication applications? BTL1</b></p> <ul style="list-style-type: none"> <li>✓ Security for E-mail</li> <li>✓ Internet protocol security</li> <li>✓ IP address security.</li> </ul>								
13	<p><b>Differentiate public key encryption and conventional encryption. BTL3</b></p> <table border="1"> <thead> <tr> <th>Conventional encryption</th> <th>Public key encryption</th> </tr> </thead> <tbody> <tr> <td>Same algorithm with same key used for encryption and decryption</td> <td>Same algorithm is used for encryption and decryption with a pair of keys</td> </tr> <tr> <td>Sender and receiver must share the algorithm and key</td> <td>Sender and receiver have one of the matched pair keys</td> </tr> <tr> <td>Key must be kept secret.</td> <td>Any one of the keys must be kept secretly.</td> </tr> </tbody> </table>	Conventional encryption	Public key encryption	Same algorithm with same key used for encryption and decryption	Same algorithm is used for encryption and decryption with a pair of keys	Sender and receiver must share the algorithm and key	Sender and receiver have one of the matched pair keys	Key must be kept secret.	Any one of the keys must be kept secretly.
Conventional encryption	Public key encryption								
Same algorithm with same key used for encryption and decryption	Same algorithm is used for encryption and decryption with a pair of keys								
Sender and receiver must share the algorithm and key	Sender and receiver have one of the matched pair keys								
Key must be kept secret.	Any one of the keys must be kept secretly.								
14	<p><b>What is message authentication? BTL2</b></p> <p>Message authentication is a process that verifies whether the received message comes from</p>								

	assigned source has not been altered.
15	<p><b>Specify the requirements for message authentication.</b> BTL3</p> <ul style="list-style-type: none"> <li>✓ Disclosure</li> <li>✓ Traffic analysis</li> <li>✓ Masquerade</li> <li>✓ Content modification</li> <li>✓ Sequence modification</li> <li>✓ Timing modification</li> <li>✓ Repudiation.</li> </ul>
16	<p><b>Specify the four categories of security threats.</b> BTL3</p> <ul style="list-style-type: none"> <li>✓ Interruption</li> <li>✓ Interception</li> <li>✓ Modification</li> <li>✓ Fabrication</li> </ul>
17	<p><b>What do you mean by SET? What are the features of SET?</b> BTL2</p> <p>SET is an open encryption and security specification designed to protect credit card transaction on the Internet.</p>
18	<p><b>Write any 3 hash algorithm.</b> BTL2</p> <ul style="list-style-type: none"> <li>✓ MD5 algorithm</li> <li>✓ SHA-I</li> <li>✓ RIPEMD-160 algorithm.</li> </ul>
19	<p><b>What is worm?</b> BTL2</p> <p>A worm is a program that can replicate itself and send copies from computer to computer across network connections</p>
20	<p><b>What is Bastion host?</b> BTL2</p> <p>Bastion host is a system identified by firewall administrator as critical strong point in network security</p>
21	<p><b>Write the four general techniques of firewall.</b> BTL3</p> <ul style="list-style-type: none"> <li>✓ Security control</li> <li>✓ Direction control</li> <li>✓ User control</li> <li>✓ Behavior control</li> </ul>
22	<p><b>Write the three types of firewall.</b> BTL3</p> <ul style="list-style-type: none"> <li>✓ Packet filter</li> <li>✓ Application level gateway</li> <li>✓ Circuit level gateway</li> </ul>
23	<p><b>List approaches for intrusion detection.</b> BTL1</p> <ul style="list-style-type: none"> <li>✓ Statistical anomaly detection</li> <li>✓ Rule based detection</li> </ul>
24	<p><b>What is meant by SET? What are the features of SET?</b> BTL2</p> <p>Secure Electronic Transaction (SET) is an open encryption and security specification designed to protect credit card transaction on the internet.</p> <p>Features are:</p>

	<ul style="list-style-type: none"> <li>✓ Confidentiality of information</li> <li>✓ Integrity of data</li> <li>✓ Cardholder account authentication</li> <li>✓ Merchant authentication</li> </ul>
25	<p><b>What is Zombie? BTL2</b>  Zombie is a program that securely takes over another internet-attached computer and then uses that computer to launch attacks are difficult to trace the Zombie's creator.</p>
	<b>PART * B</b>
1	<p><b>Explain Authentication applications. (13M) BTL4 (May/June 2015, May/June 2014)</b>  <b>Answer : Page : 476 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Kerberos  It is an authentication service developed as part of project Athena at MIT. The problem that Kerberos address is, assume an open distributed environment in which users at work stations wish to access services on servers distributed throughout the network. It is an authentication protocol (2M)</li> <li>✓ How Kerberos works  It differs from username authentication methods because instead of authenticating each user to each network service, it uses symmetric encryption and a trusted third party to authenticate users to a suite of network services (2M)</li> </ul>  <pre> graph TD     Client[Client] -- "Client authentication" --&gt; AS[Authentication server (AS)]     AS -- "Shared key and ticket" --&gt; TGS[Ticket-granter server (AS)]     TGS -- "Ticket, server ID, and client authentication" --&gt; SP[Service provider]     SP -- "Ticket and client authentication" --&gt; TGS     TGS -- "Shared key and ticket" --&gt; SP     SP -- "Service granted" --&gt; TGS     TGS -- "Ticket and client authentication" --&gt; Client     Client -- "Ticket and client authentication" --&gt; SP     </pre> <p>✓ Kerberos Message Exchanges (3M)</p>

(1)  $C \rightarrow AS \quad ID_c \parallel ID_{tgs} \parallel TS_1$   
(2)  $AS \rightarrow C \quad E(K_c, [K_{c,tgs} \parallel ID_{tgs} \parallel TS_2 \parallel Lifetime_2 \parallel Ticket_{tgs}])$   
 $Ticket_{tgs} = E(K_{tgs}, [K_{c,tgs} \parallel ID_C \parallel AD_C \parallel ID_{tgs} \parallel TS_2 \parallel Lifetime_2])$

(a) Authentication Service Exchange to obtain ticket-granting ticket

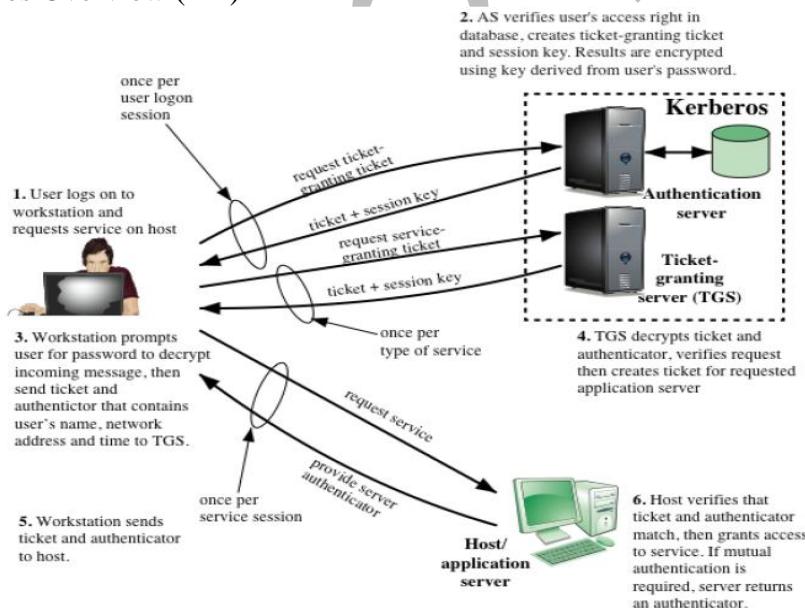
(3)  $C \rightarrow TGS \quad ID_v \parallel Ticket_{tgs} \parallel Authenticator_c$   
(4)  $TGS \rightarrow C \quad E(K_{c,tgs}, [K_{c,v} \parallel ID_v \parallel TS_4 \parallel Ticket_v])$   
 $Ticket_{tgs} = E(K_{tgs}, [K_{c,tgs} \parallel ID_C \parallel AD_C \parallel ID_{tgs} \parallel TS_2 \parallel Lifetime_2])$   
 $Ticket_v = E(K_v, [K_{c,v} \parallel ID_C \parallel AD_C \parallel ID_v \parallel TS_4 \parallel Lifetime_4])$   
 $Authenticator_c = E(K_{c,tgs}, [ID_C \parallel AD_C \parallel TS_3])$

(b) Ticket-Granting Service Exchange to obtain service-granting ticket

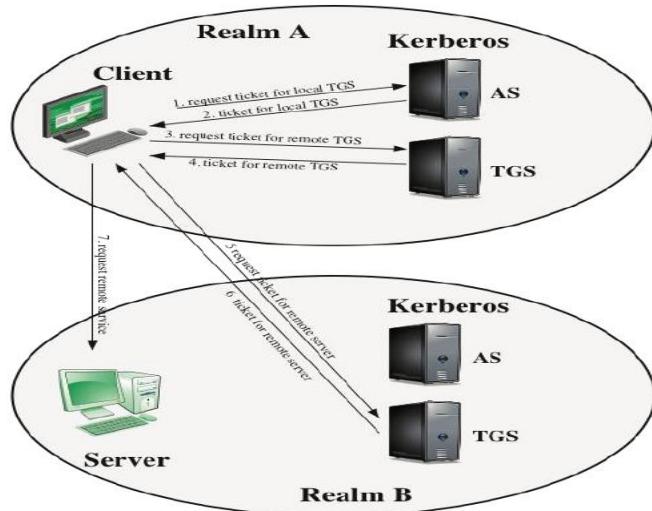
(5)  $C \rightarrow V \quad Ticket_v \parallel Authenticator_c$   
(6)  $V \rightarrow C \quad E(K_{c,v}, [TS_5 + 1])$  (for mutual authentication)  
 $Ticket_v = E(K_v, [K_{c,v} \parallel ID_C \parallel AD_C \parallel ID_v \parallel TS_4 \parallel Lifetime_4])$   
 $Authenticator_c = E(K_{c,v}, [ID_C \parallel AD_C \parallel TS_5])$

(c) Client/Server Authentication Exchange to obtain service

✓ Kerberos Overview (2M)



✓ Kerberos Realm- A full-service Kerberos environment consisting of a Kerberos server, a number of clients, and a number of application servers. (2M)



- ✓ Difference between Kerberos 4 and 5
  - ✓ Kerberos Version 5 Message Exchanges

(1M)  
(1M)

(1) **C → AS** *Options || IDc || Realmc || IDtgs || Times || Nonce1*

(2) AS → C  $Realmc \parallel IDC \parallel Tickettgs \parallel E(Kc, [Kc, tgs \parallel Times \parallel Nonce1 \parallel Realmtgs \parallel IDtgs])$

$$Ticketngs = E(Ktgs, [Flags \parallel Kc,tgs \parallel Realmc \parallel IDC \parallel ADC \parallel Times])$$

**(a) Authentication Service Exchange to obtain ticket-granting ticket**

(3) C → TGS   *Options || IDv || Times || || Nonce2 || Tickettgs || Authenticatorc*

(4)  $TGS \rightarrow C \text{ } Realmc \parallel IDC \parallel Ticketv \parallel E(Kc,tgs, [Kc,v \parallel Times \parallel Nonce2 \parallel Realmv \parallel IDv])$

*Tickettgs = E(Ktgs, [Flags || Kc,tgs || Realmc || IDC || ADC || Times])*

*TicketIV = E(Kv, [Flags || Kv-y || Realmc || IDC || ADC || Times])*

*Authenticatorc = E(Kc-tes, [IDC || Realmc || TS])*

(b) Ticket-Granting Service Exchange to obtain service-granting ticket

(b) Proper training will also encourage a team to self-govern their work.

(5)  $C \Rightarrow V$ . Options || Ticket- || Authenticator-

(6)  $N \rightarrow G$ ,  $E = \{TS_1\} \cup \{S_{12}\} \cup S_{13}\#1$

[*13½* || *Subkey* || *Seq#* ]

(6)  $V \rightarrow C \quad E_{K_C, v} [ TS_2 \parallel Subkey \parallel Seq\# ]$

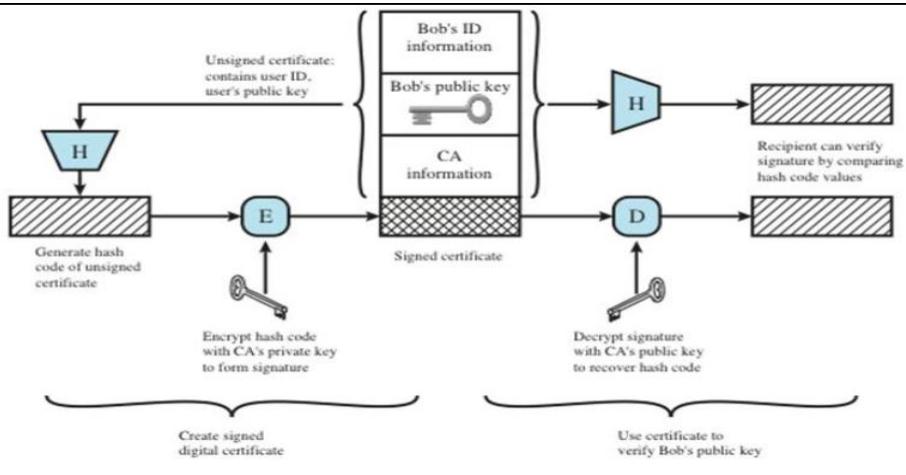
$$\text{Ticketv} = E(Kv, [\text{Flags} \parallel Kc,v \parallel \text{Realmc} \parallel IDC \parallel ADC \parallel \text{Times}])$$

$$Authenticator_c = E(K_c, v, [IDC \parallel Realm_c \parallel TS2 \parallel Subkey \parallel Seq\#])$$

2 Explain in detail about X.509. (13M) BTL4 (May/June 2013)

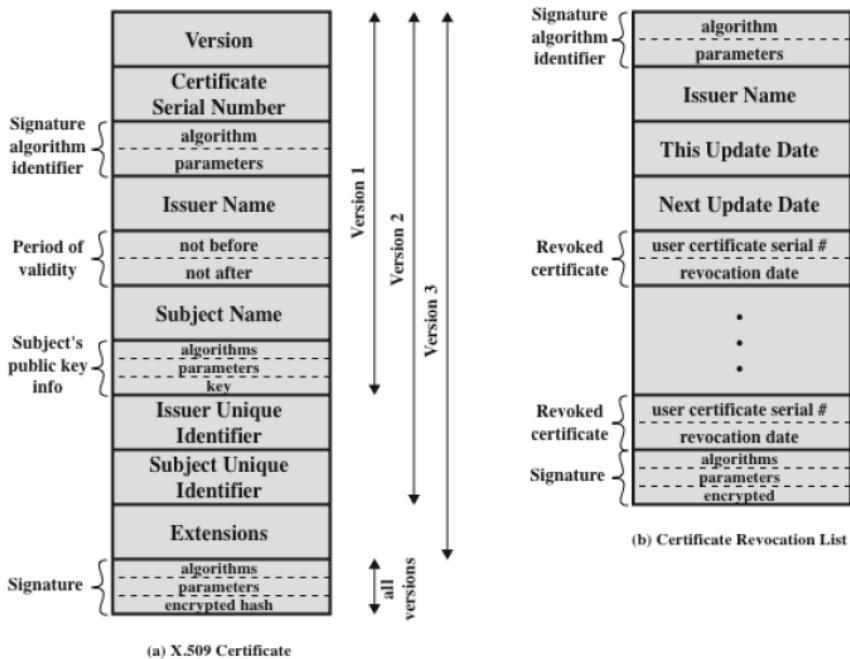
**Answer : Page : 453 – William Stallings**

- ✓ X.509 Authentication service defines the structure of digital certificates  
X.509 defines framework for authentication services by the X.500 directory to its users.X.509 defines authentication protocols based on public key certificates. (2M)
  - ✓ X.509 Framework- Was initially issued in 1988 with the latest revision in 2000 .Based on the use of public-key cryptography and digital signatures .Does not dictate the use of a specific algorithm but recommends RSA.Does not dictate a specific hash algorithm (2M)
  - ✓ Public Key Certificate Use



(3M)

✓ X.509 Format



(3M)

(2M)

(1M)

- ✓ Obtaining Certificate
- ✓ Certificate Revocation

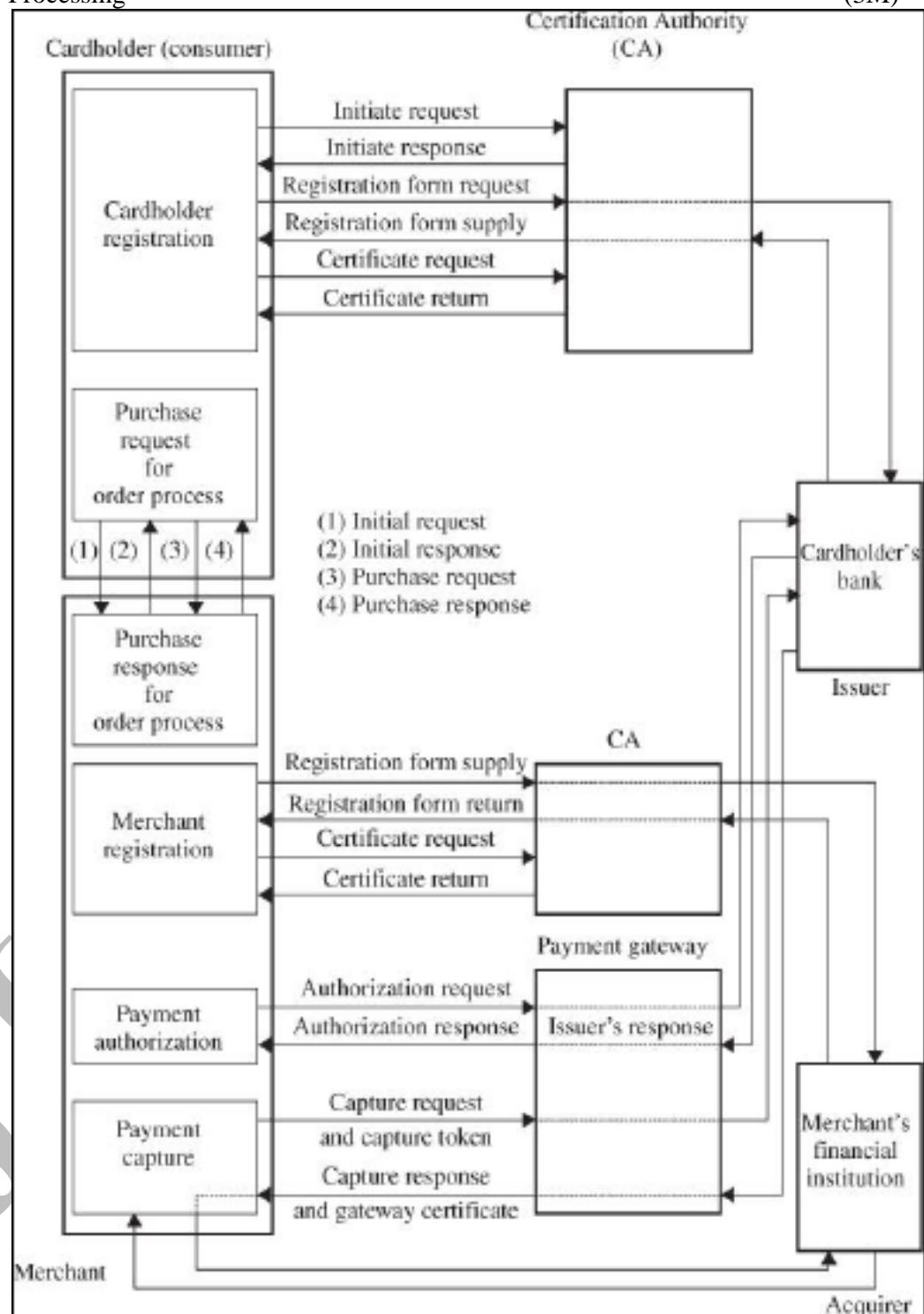
3 Explain Secure Electronic Transaction. (13M) BTL4 (Apr/May 2011, Nov/Dec 2011, Nov/Dec 2012)

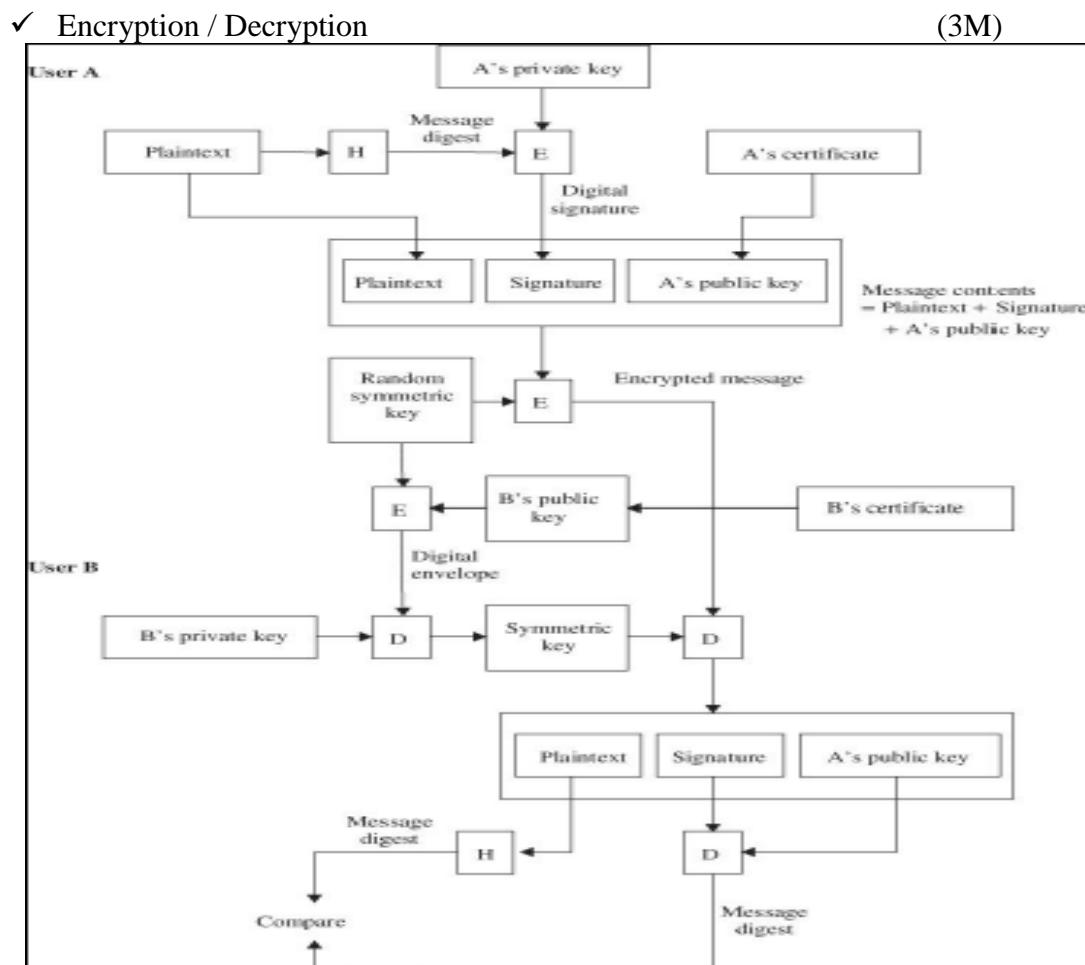
Answer : Page : 627 – William Stallings

SET(2M)

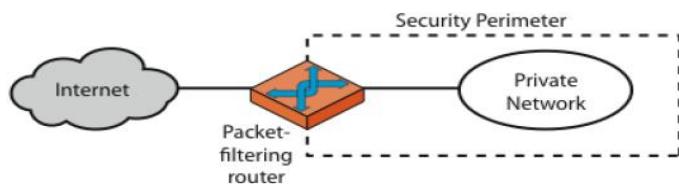
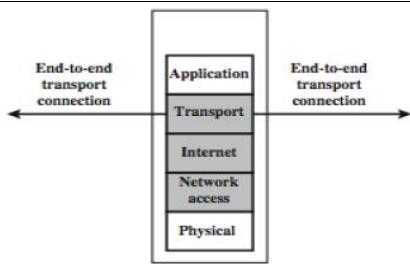
- ✓ Set of security protocol and formats enable user to employ existing infrastructure  
Secure Electronic Transaction (SET) is an open encryption and security specification designed to protect credit card transaction on the internet (2M)
- ✓ Services provided by SET (1M)
- ✓ Key features (1M)

- ✓ Business Requirements for SET (1M)
- ✓ SET System Participants (2M)
- ✓ Processing (3M)





- 4 Explain about Firewall and types of Firewall and Design of Firewall. (13M) BTL4 (Apr/May 2011, Nov/Dec 2011, May/June 2012, Nov/Dec 2012, Nov/Dec 2013, May/June 2015)  
**Answer : Page : 630 – William Stallings**
- ✓ Firewall (1M)
  - A **choke point** of control and monitoring, interconnects networks with differing trust, imposes restrictions on network services, only authorized traffic is allowed, auditing and controlling access, can implement alarms for abnormal behavior
  - ✓ Types of firewall
  - (i) Packet filtering router firewall  
 Protecting internal users from the external network threats is to implement this type of security (1M)
  - ✓ Diagram(2M)

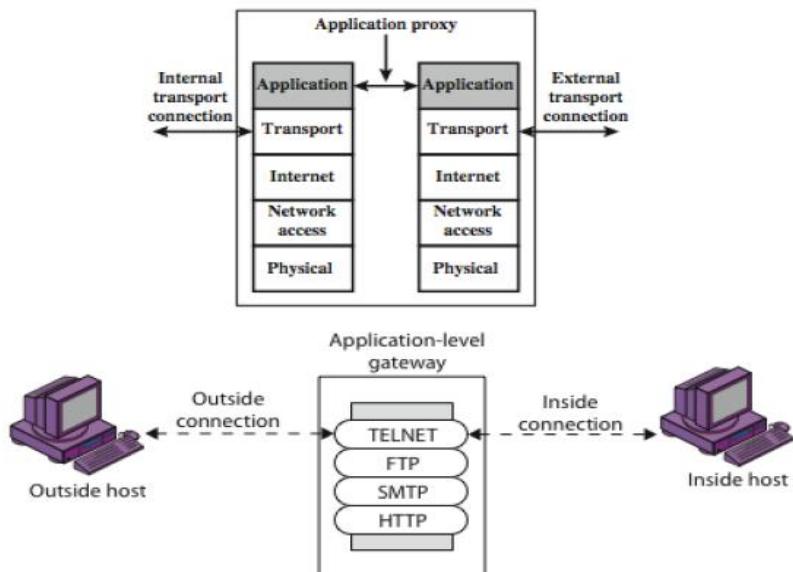


✓ Advantages and Disadvantages(1M)

(ii) Application Level Gateway

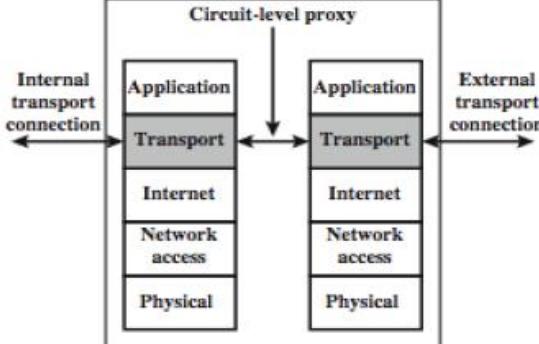
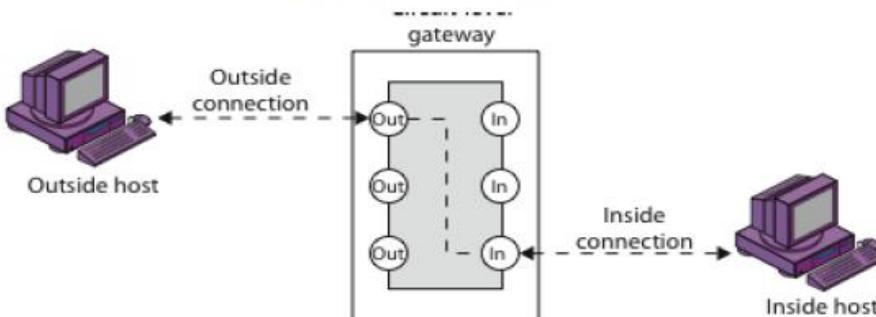
Have application specific gateway / proxy, has full access to protocol, user requests service from proxy ,proxy validates request as legal,then actions request and returns result to user,can log / audit traffic at application level (1M)

✓ Diagram(2M)



✓ Advantages and Disadvantages(1M)

(iii) Circuit Level Gateway- relays two TCP connections, imposes security by limiting which such connections are allowed ,once created usually relays traffic without examining contents.(1M)

	<ul style="list-style-type: none"> <li>✓ Diagram(2M)</li> <li>✓ Advantages and Disadvantages(1M)</li> </ul>  <p style="text-align: center;">(e) Circuit-level proxy firewall</p> 
5	<p><b>Explain in detail about Intrusion detection system. (13M) BTL4 (Nov/Dec 2011, May/June 2014)</b></p> <p><b>Answer : Page : 645 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Intrusion Detection Systems look for attack signatures, which are specific patterns that usually indicate malicious or suspicious intent. (1M)</li> <li>✓ Statistical anomaly detection Involves the collection of data relating to the behavior of legitimate users over a period of time (2M)</li> <li>✓ Rule based detection Involves as a set of rules can be used to decide given behavior of an intruder (2M)</li> <li>✓ Distributed intrusion detection System should detect a substantial percentage of intrusion while keeping the false alarm rate at acceptable level. (2M)</li> <li>✓ Rule based penetration identification (2M)</li> <li>✓ Architecture for distribution intrusion detection – diagram (2M)</li> </ul>

	<p>         ✓ LAN Monitor agent module          Operates same as a host agent module except that it analyzes LAN traffic and reports the results to the central manager. (1M)       </p> <p>         ✓ Honey Pot          Relatively recent innovation in intrusion detection technology (1M)       </p>
6	<p><b>Explain about Malicious software viruses. (13M) BTL4 (May/June 2012, Nov/Dec 2012, May/June 2013, Nov/Dec 2013, May/June 2014, May/June 2015)</b></p> <p><b>Answer : Page : 645 &amp; 650 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Malicious programs</li> <li>Overall taxonomy of software threats (1M)</li> <li>✓ Virus</li> <li>It is a program that can infect other programs by modifying them (1M)</li> <li>✓ Four phases</li> <li>Dormant phase</li> <li>Propagation phase</li> <li>Triggering phase</li> <li>Execution phase (1M)</li> <li>✓ Virus structure</li> <li>Virus can be postponed to an executable program (1M)</li> <li>✓ Types of virus</li> <li>Macro virus</li> <li>E mail virus</li> <li>Morris virus</li> <li>Worm (1M)</li> <li>✓ Macro virus</li> <li>Platform independent virus infect the documents and easily spread (2M)</li> <li>✓ E mail virus</li> <li>It spreads through mails, use of MS embedded in attachment (2M)</li> <li>✓ Worm</li> <li>It seeks out more machines to infect and every machine that is infected serves as a launch site</li> </ul>

	automatic padding for attacks on other machines ✓ Virus counter measures	(2M) (2M)
7.	<p><b>Explain about Various types of Configurations or Firewall Designs in Firewall. (13M)</b></p> <p>BTL4</p> <p><b>Answer : Page : 630 – William Stallings</b></p> <p>The three basic firewall designs are considered: <b>a single-homed bastion host</b>, <b>a dual-homed bastion host</b>, and <b>a screened subnet firewall</b>.(1M)</p> <p>(i) <b>Screened Host Firewall (Single-Homed Bastion Host)</b></p> <p>Uses a single-homed bastion host plus a packet-filtering router. Single-homed bastion hosts can be configured as either circuit-level or application-level gateways. When using either of these two gateways, each of which is called a <i>proxy server</i>, the bastion host can hide the configuration of the internal network.(1M)</p> <p>✓ Diagram(2M)</p> <p>✓ Advantages and Disadvantages(1M)</p> <p>(ii) <b>Screened Host Firewall (Dual-Homed Bastion Host)</b></p> <p>The configuration of the screened host firewall using a dual-homed bastion host adds significant security, compared with a single-homed bastion host. Dual-homed bastion host has two network interfaces. (1M)</p> <p>✓ Diagram(2M)</p> <p>✓ Advantages and Disadvantages(1M)</p> <p>(iii) <b>Screened Subnet Firewall</b></p> <p>The third implementation of a firewall is the screened subnet, which is also known as a <i>DMZ</i>. This firewall is the most secure one among the three implementations, simply because it uses a bastion host to support both circuit- and application-level gateways. All publicly accessible devices, including modem and server, are placed inside the DMZ. (1M)</p>	

	<ul style="list-style-type: none"> <li>✓ Diagram(2M)</li> </ul> <ul style="list-style-type: none"> <li>✓ Advantages and Disadvantages(1M)</li> </ul>
	<b>PART * C</b>
1	<p><b>Explain Authentication applications. (15M) BTL4</b>  <b>Answer : Page : 468 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Kerberos           <p>It is an authentication service developed as part of project Athena at MIT. The problem that Kerberos address is, assume an open distributed environment in which users at work stations wish to access services on servers distributed throughout the network. It is an authentication protocol (3M)</p> </li> <li>✓ How Kerberos works           <p>It differs from username authentication methods because instead of authenticating each user to each network service, it uses symmetric encryption and a trusted third party to authenticate users to a suite of network services (3M)</p> </li> <li>✓ Steps with protocol (5M)           <ul style="list-style-type: none"> <li>User client logon</li> <li>Client authentication</li> <li>Client service authentication</li> <li>Client service request</li> </ul> </li> <li>✓ Diagram</li> </ul> <div style="border: 1px solid black; padding: 10px; margin-top: 10px;"> <p>The diagram illustrates the six-step Kerberos authentication process:</p> <ol style="list-style-type: none"> <li>1 Request TGT</li> <li>2 TGT + Session Key</li> <li>3 Request Ticket + Auth</li> <li>4 Ticket + Session Key</li> <li>5 Request Service + Auth</li> <li>6 Server Authentication</li> </ol> <p>Alice's client sends step 1 (Request TGT) to the KDC. The KDC responds with step 2 (TGT + Session Key). Alice sends step 3 (Request Ticket + Auth) to the KDC. The KDC responds with step 4 (Ticket + Session Key). Alice sends step 5 (Request Service + Auth) to the Resource server. The Resource server performs step 6 (Server Authentication).</p> </div> <p style="text-align: right;">(2M)</p>

	<p>✓ Explanation of working methodology with diagram</p> <pre> graph TD     Client -- "1. Logon" --&gt; KDC     KDC -- "2. Return Kerberos Token" --&gt; Client     Client -- "3. Request" --&gt; Tomcat     Tomcat -- "4. Authenticate" --&gt; KDC     KDC -- "5. Request with Token" --&gt; Tomcat     Tomcat -- "10. 200 OK Response" --&gt; Client     Tomcat -- "7. Verification Response" --&gt; LDAP     LDAP -- "6. Verify Token" --&gt; Tomcat     Tomcat -- "8. Get Roles" --&gt; LDAP     LDAP -- "9. Return Roles" --&gt; Tomcat   </pre> <p>(2M)</p>
2	<p><b>Explain Internet Firewall and its related terminology. (15M) BTL4</b></p> <p><b>Answer : Page : CHAPTER 22 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Need for firewall It is a mechanism that protects and isolates internal network (3M)</li> <li>✓ Characteristics Service control Direction control User control Behavior control</li> <li>✓ Types of firewall Packet filter Application level gateway Circuit level gateway (3M)</li> <li>✓ Firewall The host is a system identified by the firewall administrator as a critical strong point in the network security (3M)</li> <li>✓ Firewall location and configuration The first type of firewall is a screened host which uses a single homes bastion host plus a packet filtering router. It uses two or more network interfaces It is a network architecture that uses single firewall with 3 network interface.(3M)</li> </ul>

## UNIT 5- E-MAIL, IP & WEB SECURITY

**E-mail Security:** Security Services for E-mail-attacks possible through E-mail - establishing keys privacy-authentication of the source-Message Integrity-Non-repudiation-Pretty Good Privacy-S/MIME. **IP Security:** Overview of IPSec - IP and IPv6-Authentication Header-Encapsulation Security Payload (ESP)-Internet Key Exchange (Phases of IKE, ISAKMP/IKE Encoding). **Web Security:** SSL/TLS Basic Protocol-computing the keys- client authentication-PKI as deployed by SSL Attacks fixed in v3- Exportability-Encoding-Secure Electronic Transaction (SET).

### PART \* A

1	<b>Define key Identifier - BTL1</b> PGP assigns a key ID to each public key that is very high probability unique with a user ID. It is also required for the PGP digital signature. The key ID associated with each public key consists of its least significant 64bits.								
2	<b>List the limitations of SMTP/RFC 822? – BTL1</b> <ul style="list-style-type: none"> <li>• SMTP cannot transmit executable files or binary objects.</li> <li>• It cannot transmit text data containing national language characters.</li> <li>• SMTP servers may reject mail message over certain size.</li> <li>• SMTP gateways cause problems while transmitting ASCII and EBCDIC.</li> <li>• SMTP gateways to X.400 E-mail network cannot handle non textual data included in X.400 messages.</li> </ul>								
3	<b>Define S/MIME. BTL2</b> Secure/Multipurpose Internet Mail Extension(S/MIME) is a security enhancement to the MIME Internet E-mail format standard, based on technology from RSA Data Security.								
4	<b>What are the different between SSL version 3 and TLS? BTL1</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%;">SSL</th><th style="width: 50%;">TLS</th></tr> </thead> <tbody> <tr> <td>In SSL the minor version is 0 and major version is 3</td><td>In TLS, the major version is 3 and the minor version is 1</td></tr> <tr> <td>SSL use HMAC alg., except that the padding bytes concatenation</td><td>TLS makes use of the same alg</td></tr> <tr> <td>SSL supports 12 various alert codes</td><td>TLS supports all of the alert codes defined in SSL3 with the exception of no certificate</td></tr> </tbody> </table>	SSL	TLS	In SSL the minor version is 0 and major version is 3	In TLS, the major version is 3 and the minor version is 1	SSL use HMAC alg., except that the padding bytes concatenation	TLS makes use of the same alg	SSL supports 12 various alert codes	TLS supports all of the alert codes defined in SSL3 with the exception of no certificate
SSL	TLS								
In SSL the minor version is 0 and major version is 3	In TLS, the major version is 3 and the minor version is 1								
SSL use HMAC alg., except that the padding bytes concatenation	TLS makes use of the same alg								
SSL supports 12 various alert codes	TLS supports all of the alert codes defined in SSL3 with the exception of no certificate								
5	<b>What are the services provided by PGP services? BTL1</b> <ul style="list-style-type: none"> <li>• Digital signature</li> <li>• Message encryption</li> <li>• Compression</li> <li>• E-mail compatibility</li> <li>• Segmentation</li> </ul>								

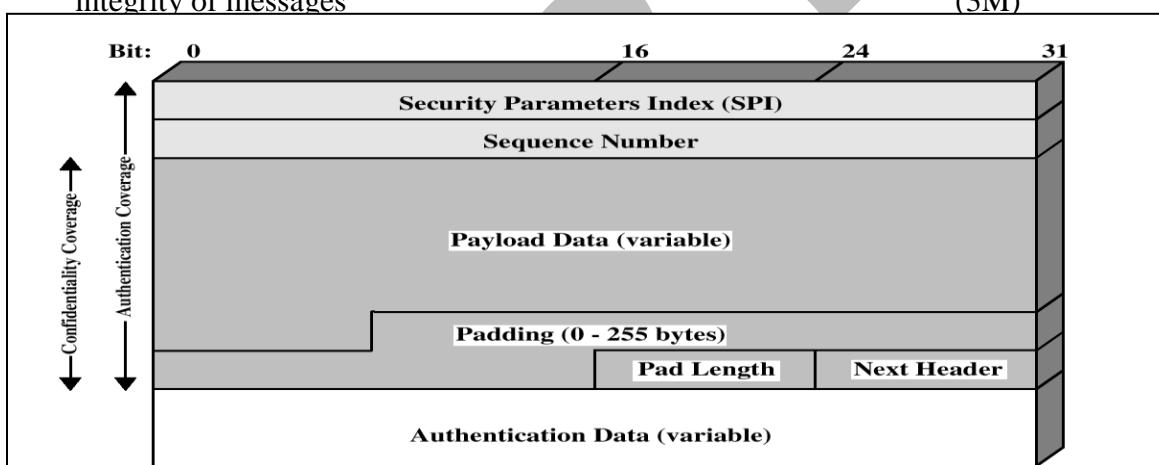
6	<b>Why E-mail compatibility function in PGP needed? BTL2</b> Electronic mail systems only permit the use of blocks consisting of ASCII text. To accommodate this restriction PGP provides the service converting the raw 8-bit binary stream to a stream of printable ASCII characters. The scheme used for this purpose is Radix-64 conversion
7	<b>Name any cryptographic keys used in PGP. BTL3</b> <ul style="list-style-type: none"> <li>✓ One-time session conventional keys.</li> <li>✓ Public keys.</li> <li>✓ Private keys.</li> <li>✓ Pass phrase based conventional keys.</li> </ul>
8	<b>Define S/MIME .BTL1</b> Secure / Multipurpose Internet Mail Extension(S/MIME) is a security enhancement to the MIME internet E-mail format standard, based on technology from RSA Data security.
9	<b>What are the services provided by PGP services? BTL2</b> <ul style="list-style-type: none"> <li>✓ Digital signature</li> <li>✓ Compression</li> <li>✓ Segmentation</li> <li>✓ Message encryption</li> <li>✓ E-mail compatibility</li> </ul>
10	<b>Name any cryptographic keys used in PGP. BTL3</b> <ul style="list-style-type: none"> <li>✓ One time session conventional keys</li> <li>✓ Public keys</li> <li>✓ Private keys</li> <li>✓ Pass phrase based conventional keys.</li> </ul>
11	<b>What is security association? BTL2</b> A security association (SA) is the establishment of shared security attributes between two network entities to support secure communication.
12	<b>What does Internet key management in IPSec? BTL2</b> Internet key exchange (IKE) is a key management protocol standard used in conjunction with the Internet Protocol Security (IPSec) standard protocol. It provides security for Virtual Private Networks (VPNs) negotiations and network access to random hosts.
13	<b>List out the IKE hybrid protocol dependence. BTL1</b> <ul style="list-style-type: none"> <li>✓ ISAKMP - Internet Security Association and Key Management Protocols.</li> <li>✓ Oakley</li> </ul>
14	<b>What does IKE hybrid protocol mean? BTL2</b> Internet Key Exchange (IKE) is a key management protocol standard used in conjunction with the internet protocol security (IPSec) standard protocol. It provides security for Virtual Private Networks (VPNs) negotiations and network access to random hosts.
15	<b>What are the two security services provided by IPSec? BTL2</b> <ul style="list-style-type: none"> <li>✓ Authentication Header (AH)</li> <li>✓ Encapsulating Security Payload (ESP).</li> </ul>

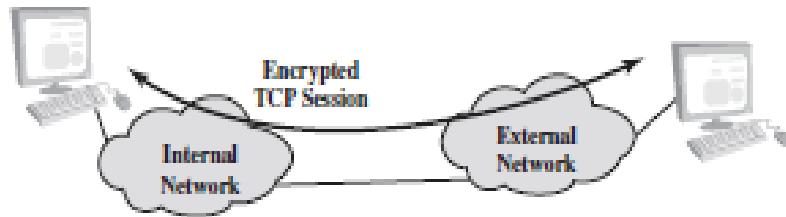
16	<b>What are the fields available in AH header? BTL2</b> <ul style="list-style-type: none"> <li>✓ Next header</li> <li>✓ Payload length</li> <li>✓ Reserved</li> <li>✓ Security parameter</li> <li>✓ Sequence number Integrity check value</li> </ul>
17	<b>What is virtual private network? BTL2</b> VPN means virtual private network, a secure tunnel between two devices.
18	<b>What is ESP? BTL2</b> ESP-encapsulating security payload provides authentication, integrity and confidentiality, which protect against data tempering and provide message content protection
19	<b>What is Behavior-Blocking Software (BBS)? BTL2</b> BBS integrates with the OS of a host computer and monitors program behavior in real time for malicious actions.
20	<b>List password selection strategies. BTL1</b> <ul style="list-style-type: none"> <li>✓ User education</li> <li>✓ Reactive password checking</li> <li>✓ Computer-generated password.</li> <li>✓ Proactive password checking.</li> </ul>

**Part \* B**

1	<b>Explain about Email Security. (13M) BTL4</b> <b>Answer : Page : 591 – William Stallings</b> <ul style="list-style-type: none"> <li>✓ Modes of Operation <ul style="list-style-type: none"> <li>Authentication</li> <li>Confidentiality</li> <li>Compression</li> <li>e-mail compatibility</li> </ul> <span style="float: right;">(2M)</span> </li> <li>✓ Tunnel mode <ul style="list-style-type: none"> <li>It provides the protection to the entire IP Packet</li> </ul> <span style="float: right;">(1M)</span> </li> <li>✓ Transport mode <ul style="list-style-type: none"> <li>It provides protection primarily for upper layer protocols</li> </ul> <span style="float: right;">(1M)</span> </li> <li>✓ Internet key exchange protocol <ul style="list-style-type: none"> <li>Manual</li> <li>Automated</li> </ul> <span style="float: right;">(2M)</span> </li> <li>✓ Security Association</li> </ul>
---	--

	<p>It provides a framework for internet key management and provides the specific protocol support including formats (1M)</p> <ul style="list-style-type: none"> <li>✓ Contents of SAD</li> <li>✓ It represent a specification of security services offered to traffic carried through a in-directional channel from one node to another (1M)</li> <li>✓ Authentication Header</li> <li>✓ It is used to provide connectionless integrity and data origin authentication for IP datagrams (1M)</li> <li>✓ Authentication Header fields           <ul style="list-style-type: none"> <li>Access control</li> <li>Connectionless integrity</li> <li>Data origin authentication</li> <li>Confidentiality</li> </ul> </li> <li>✓ Anti Replay Attacks           <p>It is a sub protocol of IPsec that is part of Internet engineering task force. The main goal is to avoid hackers injecting or making changes in packets that travel from a source to destination (1M)</p> </li> <li>✓ Values in sliding window (2M)</li> </ul>
2	<p><b>Explain in detail about IPSEC. (13M) BTL4</b></p> <p><b>Answer : Page : 640 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ IP v4           <p>Specifies an IPv4 address or range of addresses that are authorized senders for a domain (2M)</p> </li> <li>✓ IP v6           <p>Specifies an IPv6 address or range of addresses that are authorized senders for a domain. (2M)</p> </li> <li>✓ AH tunnel modes           <p>It authenticates the entire inner IP packet selected portion of outer IP header (1M)</p> </li> <li>✓ IP header           <p>An IP header is header information at the beginning of an IP packet which contains information about IP version, source IP address, destination IP address, time-to-live (2M)</p> </li> </ul>

	<ul style="list-style-type: none"> <li>✓ New IP header</li> <li>✓ TCP header</li> </ul> <p>TCP is the primary transport protocol used to provide reliable, full-duplex connections. The most common use of TCP is to exchange TCP data encapsulated in an IP datagram.</p> <ul style="list-style-type: none"> <li>✓ Original data</li> </ul> <p>It refers to any data object that hasn't undergone thorough processing, either manually or through automated computer software.</p> <ul style="list-style-type: none"> <li>✓ Original IP header</li> </ul>	(1M)
3	<b>Explain Encapsulating security payload. (13M) BTL4</b> <b>Answer : Page : 651 – William Stallings</b> <ul style="list-style-type: none"> <li>✓ ESP consists of an encapsulating header and trailer used to provide encryption or combined encryption/ authentication. The current specification is RFC 4303, IP Encapsulating Security Payload (ESP).The purpose is to provide confidentiality and integrity of messages</li> </ul> 	(3M)
	<ul style="list-style-type: none"> <li>✓ ESP Transport mode</li> </ul> <p>Transport mode ESP is used to encrypt and optionally authenticate the data carried by IP</p> <ul style="list-style-type: none"> <li>✓ Operation of ESP Transport mode</li> </ul>	(1M)
		(4M)



(a) Transport-level security

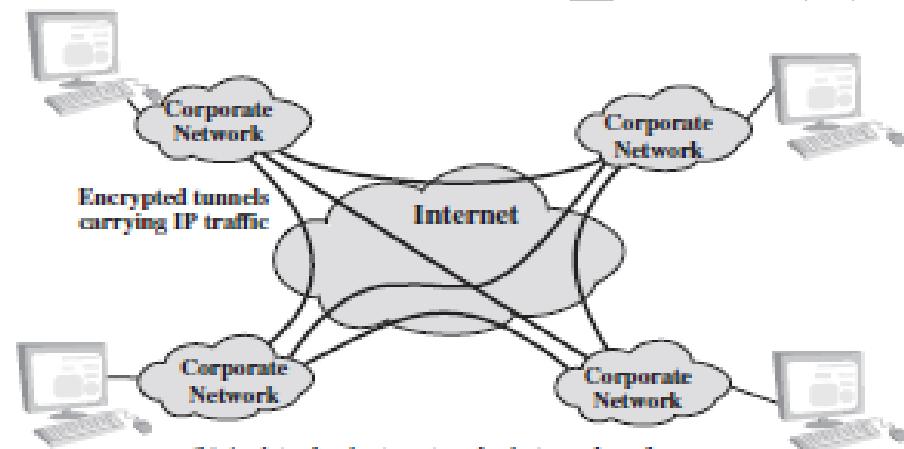
- ✓ ESP tunnel mode

Tunnel mode ESP is used to encrypt an entire IP packet. For this mode, the ESP header is prefixed to the packet and then the packet plus the ESP trailer is encrypted. This method can be used to counter traffic analysis

(1M)

- ✓ Operation of ESP tunnel mode

(4M)



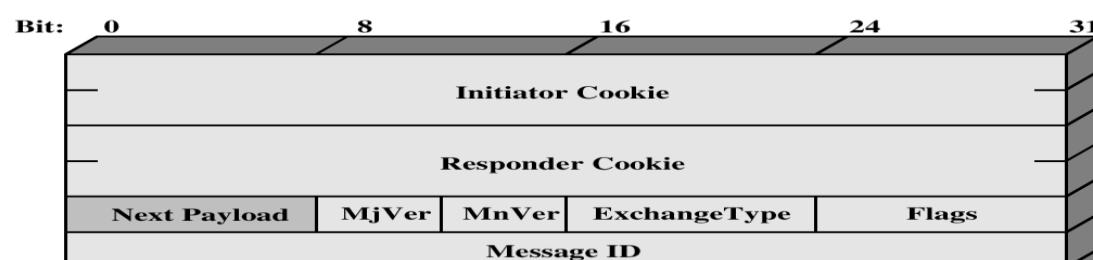
(b) A virtual private network via tunnel mode

4

**Explain Internet security Association and Key Management Protocol. (13M) BTL4**

**Answer : Page : 663 – William Stallings**

- ✓ Procedures and formats for establishing maintaining and deleting Security Association information (1M)
- ✓ IP Seckey management (2M)
- ✓ Initiator cookie (1M)
- ✓ Responder cookie (1M)
- ✓ Major and minor version (3M)
- ✓ Message ID (1M)
- ✓ Length (1M)
- ✓ Payloads (2M)



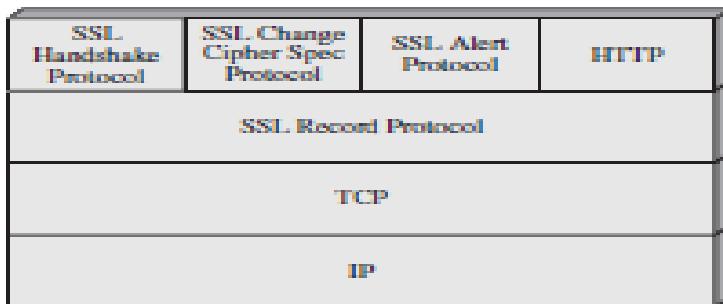
- ✓ Exchange types (1M)

5 Explain about WEB security. (13M) BTL4 (Apr/May 2011, Nov/Dec 2011, Nov/Dec 13, May/June 2015)

Answer : Page : 510 – William Stallings

- ✓ Secure socket layer

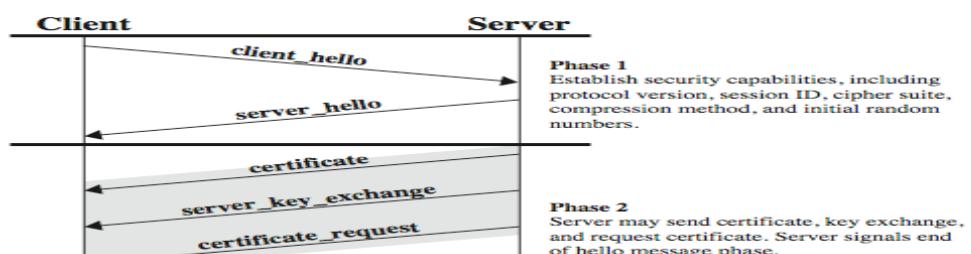
It is the standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data passed between the web server and browsers remain private and integral. (2M)



- ✓ How SSL works

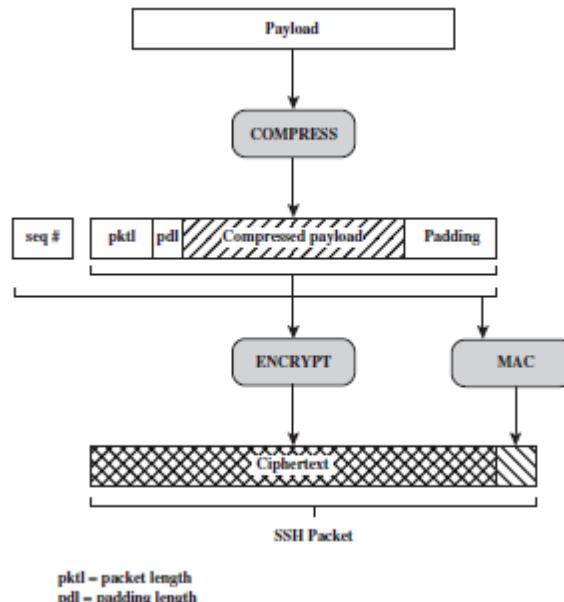
After connection is made, the session key is used to encrypt all transmitted data. Browser connects to a web server secured with SSL and request the server identity. Server sends a copy of its SSL certificate (2M)

- ✓ Handshake protocol (2M)



- ✓ Change cipher spec protocol (1M)
- ✓ Record protocol (1M)
- ✓ Alert protocol (1M)
- ✓ Fatal alerts (2M)
- ✓ Transport layer security (2M)

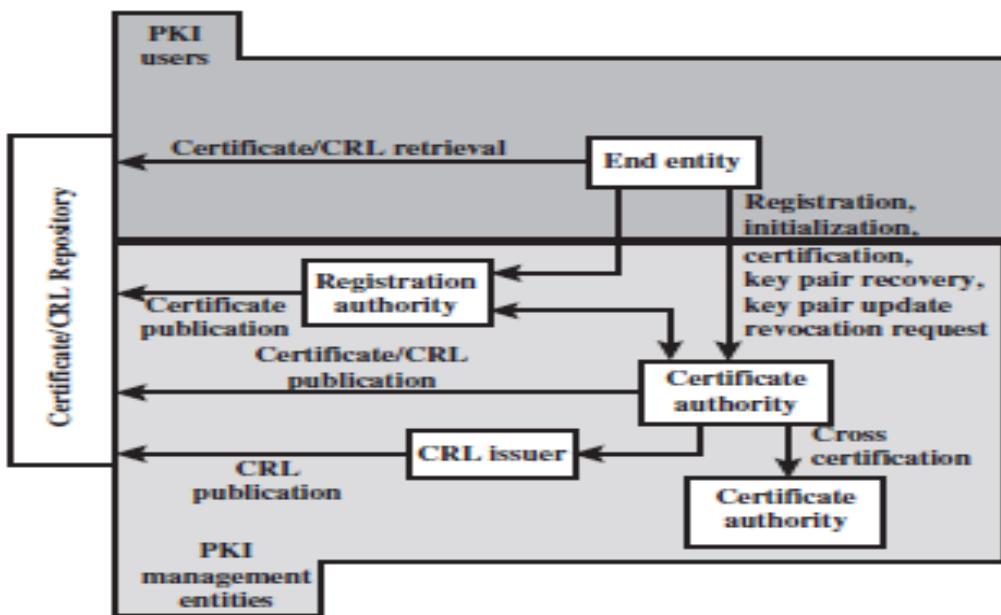
Definition  
Architecture  
Parameter  
Architecture Diagram



6

**Explain public key infrastructure. (13M) BTL4 (Nov/Dec 2013)**

	<b>Answer : Page : 515 – William Stallings</b>
	<ul style="list-style-type: none"> <li>✓ It is a model for creating, distributing and revoking certificates based on X.509. A set of policies, processes, server platforms, software and workstations used for the purpose of administering certificates and public-private key pairs, including the ability to issue, maintain, and revoke public key certificates (1M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ End entity A generic term used to denote end users, devices (e.g., servers, routers), or any other entity that can be identified in the subject field of a public-key certificate. End entities typically consume and/or support PKI-related services (1M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ Certification authority The issuer of certificates and (usually) certificate revocation lists (CRLs). It may also support a variety of administrative functions, although these are often delegated to one or more Registration Authorities. (2M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ Registration authority An optional component that can assume a number of administrative functions from the CA. The RA is often associated with the end entity registration process but can assist in a number of other areas as well. (2M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ CRI issuer An optional component that a CA can delegate to publish CRLs (1M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ Repository A generic term used to denote any method for storing certificates and CRLs so that they can be retrieved by end entities. (1M)</li> </ul>
	<ul style="list-style-type: none"> <li>✓ PKI Architecture – Diagram (2M)</li> </ul>



- ✓ PKI management functions
  - Registration
  - Initialization
  - Certification
  - Key pair recovery
  - Key pair update
  - Revocation request
  - Cross certification
- ✓ PKI management protocols

(2M)

(1M)

### PART \* C

1 Explain E-mail security. (15M) BTL4

Answer : Page : 591 – William Stallings

- ✓ Security Services for E-mail (2M)
- ✓ Possible Attacks through E-mail (2M)
- ✓ Establishing Keys privacy (2M)
- ✓ Authentication of source (2M)
- ✓ Message Integrity (2M)

	<ul style="list-style-type: none"> <li>✓ NonRepudiation (1M)</li> <li>✓ PGP (2M)</li> <li>✓ S/MIME (2M)</li> </ul>
2	<p><b>Explain IP Security. (15M) BTL4</b>  <b>Answer : Page : 639 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Security policy (3M) <ul style="list-style-type: none"> <li>Security Parameters Index (SPI)</li> <li>IP Destination Address</li> <li>Security Protocol Identifier</li> </ul> </li> <li>✓ Encapsulating security payload (3M) <ul style="list-style-type: none"> <li>Diagram</li> <li>Format</li> <li>Algorithm</li> </ul> </li> <li>✓ Internet key exchange (3M) <ul style="list-style-type: none"> <li>Manual</li> <li>Automated</li> <li>OKDP</li> <li>ISAKMP</li> </ul> </li> <li>✓ Cryptographic suites (3M) <ul style="list-style-type: none"> <li>ESP encryption</li> <li>ESP integrity</li> <li>IKE encryption</li> <li>IKE PRF</li> <li>IKE Integrity</li> <li>IKE DH group</li> </ul> </li> <li>✓ Diagrams (3M)</li> </ul>

	<pre> sequenceDiagram     participant Initiator     participant Responder     Initiator-&gt;&gt;Responder: HDR, SAi1, KEi, Ni     Responder-&gt;&gt;Initiator: HDR, SAr1, KER, Nr, [CERTREQ]     Initiator-&gt;&gt;Responder: HDR, SK {IDi, [CERT,] [CERTREQ,] [IDr,] AUTH, SAi2, TSi, TSr}     Responder-&gt;&gt;Initiator: HDR, SK {IDr, [CERT,] AUTH, SAr2, TSi, TSr}     note over Initiator: (a) Initial exchanges     Initiator-&gt;&gt;Responder: HDR, SK {[N], SA, Ni, [KEi], [TSi, TSr]}     Responder-&gt;&gt;Initiator: HDR, SK {SA, Nr, [KER], [TSi, TSr]}     note over Initiator: (b) CREATE_CHILD_SA Exchange     Initiator-&gt;&gt;Responder: HDR, SK {[N,] [D,] [CP, ...]}     Responder-&gt;&gt;Initiator: HDR, SK {[N,] [D,] [CP, ...]}     note over Initiator: (c) Informational Exchange   </pre>			
3	<p><b>Explain Web Security. (15M) BTL4</b>  <b>Answer : Page : 510 – William Stallings</b></p> <ul style="list-style-type: none"> <li>✓ Secure socket layer (3M) <ul style="list-style-type: none"> <li>It is designed to make use of TCP to provide a reliable end-to-end secure service.</li> <li>Connection</li> <li>Session</li> </ul> </li> <li>✓ Protocols and its working (3M) <ul style="list-style-type: none"> <li>Confidentiality</li> <li>Message Integrity</li> <li>fragmentation</li> <li>compression</li> <li>message authentication code</li> </ul> </li> <li>✓ differentiation of SSL and TSL (4M)</li> </ul> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">BASIS FOR COMPARISON</td> <td style="padding: 5px;">SSL</td> <td style="padding: 5px;">TLS</td> </tr> </table>	BASIS FOR COMPARISON	SSL	TLS
BASIS FOR COMPARISON	SSL	TLS		

	Cryptography secret	Uses message digest of the pre-master secret for creating master secret.	Uses a pseudorandom function to create master secret.	
	Record protocol	Uses MAC (Message Authentication Code)	Uses HMAC (Hashed MAC)	
	Alert protocol	The "No certificate" alert message is included.	It eliminates alert description (No certificate) and adds a dozen other values.	
	Message authentication	Ad hoc	Standard	
	key material authentication	Ad hoc	Pseudorandom function	
	Certificate verify	Complex	Simple	
	Finished	Ad hoc	Pseudorandom function	

✓ Diagrams (5M)

**SSH Protocol Exchange**

- Establish Authenticated Transport Layer Connection
- Open a channel
- Data transfer (multiple SSH\_MSG\_CHANNEL\_DATA messages)
- Close a channel

**SSL Handshake Protocol**

- Phase 1: client\_hello, server\_hello, certificate, server\_key\_exchange, certificate\_request, server\_hello\_done
- Phase 2: certificate, server\_key\_exchange, certificate\_request, server\_hello\_done
- Phase 3: certificate, client\_key\_exchange, certificate\_verify
- Phase 4: change\_cipher\_spec, finished

Note: Shaded transfers are optional or situation-dependent messages that are not always sent.

**IT6702 - DATA WAREHOUSING AND DATA MINING**

L T P C  
3003

### **OBJECTIVES:**

The student should be made to:

Be familiar with the concepts of data warehouse and data mining.

Be acquainted with the tools and techniques used for Knowledge Discovery in Databases.

<b>UNIT I</b>	<b>DATA WAREHOUSING</b>	<b>9</b>
Data warehousing Components –Building a Data warehouse -- Mapping the Data Warehouse to a Multiprocessor Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata		
<b>UNIT II</b>	<b>BUSINESS ANALYSIS</b>	<b>9</b>
Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cognos Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OLAP Guidelines – Multidimensional versus Multirelational OLAP – Categories of Tools – OLAP Tools and the Internet.		
<b>UNIT III</b>	<b>DATA MINING</b>	<b>9</b>
Introduction – Data – Types of Data – Data Mining Functionalities – Interestingness of Patterns – Classification of Data Mining Systems – Data Mining Task Primitives – Integration of a Data Mining System with a Data Warehouse – Issues –Data Preprocessing.		
<b>UNIT IV</b>	<b>ASSOCIATION RULE MINING AND CLASSIFICATION</b>	<b>9</b>
Mining Frequent Patterns, Associations and Correlations – Mining Methods – Mining various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining – Classification and Prediction - Basic Concepts - Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction.		
<b>UNIT V</b>	<b>CLUSTERING AND TRENDS IN DATA MINING</b>	<b>9</b>
Cluster Analysis - Types of Data – Categorization of Major Clustering Methods – K-means– Partitioning Methods – Hierarchical Methods - Density-Based Methods –Grid Based Methods – Model-Based Clustering Methods – Clustering High Dimensional Data - Constraint – Based Cluster Analysis – Outlier Analysis – Data Mining Applications.		
<b>TOTAL: 45</b>		<b>PERIODS</b>

## **OUTCOMES:**

**After completing this course, the student will be able to:**  
Apply data mining techniques and methods to large data sets  
Use data mining toolsCompare and contrast the various classifiers.

## **TEXT BOOKS:**

1. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining and OLAP", Tata McGraw – Hill Edition, Thirteenth Reprint 2008.
  2. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.

**Subject Code: IT6702****Subject Name: DATA WAHREHOUSING AND DATA MINING****Year/Semester: IV/07****Subject Handler: Ms.Sonia Jenifer****Rayen****UNIT -1- DATA WAREHOUSING**

**Data warehousing Components –Building a Data warehouse -- Mapping the Data Warehouse to a Multiprocessor Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata.**

**PART \* A**

Q.N O	QUESTIONS
1.	<p><b>What is data warehouse? (May/June 2010) BTL1</b>  A data warehouse is a repository of multiple heterogeneous data sources organized under a unified schema at a single site to facilitate management decision making. A <b>data warehouse</b> is a subject-oriented, time-variant and nonvolatile collection of data in support of management's decision-making process.</p>
2.	<p><b>Define Data mart. (May/June 2012) BTL1</b>  Data mart is a data store that is subsidiary to a data warehouse of integrated data. The data mart is directed at a partition of data that is created for the use of a dedicated group of users.</p>
3.	<p><b>What is data warehouse metadata?(Apr/May 2008) (NOV/DEC 2018) BTL1</b>  Metadata are data about data. When used in a data warehouse, metadata are the data that define warehouse objects. Metadata are created for the data names and definitions of the given warehouse. Additional metadata are created and captured for time stamping any extracted data, the source of the extracted data, and missing fields that have been added by data cleaning or integration processes.</p>
4.	<p><b>In the context of data warehousing what is data transformation? (May/June 2009) BTL2</b>  In <i>data transformation</i>, the data are transformed or consolidated into forms appropriate for mining. Data transformation can involve the following:</p> <ul style="list-style-type: none"> <li>✓ Smoothing</li> <li>✓ Aggregation</li> <li>✓ Generalization</li> <li>✓ Normalization</li> <li>✓ Attribute construction</li> </ul>
5.	<p><b>List the characteristics of a data warehouse. (Nov/Dec 2009) BTL1</b>  There are four key characteristics which separate the data warehouse from other major operational systems:</p> <ul style="list-style-type: none"> <li>✓ Subject Orientation: Data organized by subject</li> <li>✓ Integration: Consistency of defining parameters</li> <li>✓ Non-volatility: Stable data storage medium</li> <li>✓ Time-variance: Timeliness of data and access terms</li> </ul>

6.	<b>Mention the various sources for data warehouse? (Nov/Dec 2009) BTL4</b> Handling of relational and complex types of data: Because relational databases and data warehouses are widely used, the development of efficient and effective data mining systems for such data is important. Mining information from heterogeneous databases and global information systems: Local- and wide-area computer networks (such as the Internet) connect many sources of data, forming huge, distributed, and heterogeneous databases.
7.	<b>What is bitmap indexing?(Nov/Dec 2009) BTL1</b> The bitmap indexing method is popular in OLAP products because it allows quick searching in data cubes. The bitmap index is an alternative representation of the <i>record ID (RID)</i> list.
8.	<b>Differentiate fact table and dimension table.(May/June 2010) BTL5</b> <ul style="list-style-type: none"> <li>✓ Fact table contains the name of facts (or) measures as well as keys to each of the related dimensional tables.</li> <li>✓ A dimension table is used for describing the dimension. (e.g.) A dimension table for item may contain the attributes item_name, brand and type.</li> </ul>
9.	<b>Briefly discuss the schemas for multidimensional databases. (NOV/DEC 2018) (May/June 2010) BTL4</b> <ul style="list-style-type: none"> <li>✓ Stars schema: The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension.</li> <li>✓ Snowflakes schema: The snowflake schema is a variant of the star schema model, where some dimension tables are <i>normalized</i>, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.</li> <li>✓ Fact Constellations: Sophisticated applications may require multiple fact tables to <i>share</i> dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation.</li> </ul>
10.	<b>How is a data warehouse different from a database? How are they similar? (Nov/Dec 2007, Nov/Dec 2010, May/June 2012) BTL5</b> Data warehouse is a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision-making. A relational database is a collection of tables, each of which is assigned a unique name. Each table consists of a set of attributes (columns or fields) and usually stores a large set of tuples (records or rows). Each tuple in a relational table represents an object identified by a unique key and described by a set of attribute values. Both are used to store and manipulate the data.
11.	<b>List out the functions of OLAP servers in the data warehouse architecture. (Nov/Dec 2010) BTL2</b> The OLAP server performs multidimensional queries of data and stores the results in its multidimensional storage. It speeds the analysis of fact tables into cubes, stores the cubes until needed, and then quickly returns the data to clients.
12.	<b>Differentiate data mining and data warehousing.(Nov/Dec 2011) BTL5</b> <ul style="list-style-type: none"> <li>✓ Data mining refers to extracting or “mining” knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named “knowledge mining from data.”</li> <li>✓ A data warehouse is usually modeled by a multidimensional database structure, where each dimension corresponds to an attribute or a set of attributes in the schema, and each cell stores the value of some aggregate measure, such as count or sales amount.</li> </ul>

13.	<p><b>List out the logical steps needed to build a Data warehouse. BTL3</b></p> <ul style="list-style-type: none"> <li>✓ Collect and analyze business requirements.</li> <li>✓ Create a data model and a physical design for the Data warehouse.</li> <li>✓ Define data source</li> <li>✓ Choose the database technology and platform for the warehouse.</li> <li>✓ Extract the data from the operational databases, transform it, clean it up and load it into the database.</li> <li>✓ Choose database access and reporting tool.</li> <li>✓ Choose database connectivity software.</li> <li>✓ Choose data analysis and presentation software.</li> <li>✓ Update the data warehouse</li> </ul>
14.	<p><b>Write note on shared-nothing architecture.BTL1</b></p> <ul style="list-style-type: none"> <li>✓ The data is partitioned across all disks and the DBMS is partitioned across multiple servers</li> <li>✓ Each of which resides on individual nodes of the parallel system and has an ownership of its disk and thus its own database partition.</li> <li>✓ A shared-nothing RDBMS parallelizes the execution of a SQL query across multiple processing nodes.</li> <li>✓ Each processor has its own memory and disk and communicates with other processors by exchanging messages and data over the interconnection network.</li> </ul>
15.	<p><b>What are the access tools groups available?BTL1</b></p> <ul style="list-style-type: none"> <li>✓ Data query and reporting tools</li> <li>✓ Application development tools</li> <li>✓ Executive information system(EIS) tools</li> <li>✓ On-line analytical processing tools</li> <li>✓ Data mining tools</li> </ul>
16.	<p><b>Write down the applications of data warehousing(Apr/May 2008). BTL5</b></p> <ul style="list-style-type: none"> <li>✓ Financial services</li> <li>✓ Banking services</li> <li>✓ Customer goods</li> <li>✓ Retail sectors</li> <li>✓ Controlled manufacturing</li> </ul>
17.	<p><b>What are the applications of querying tools?(Apr/May 2011)BTL1</b></p> <ul style="list-style-type: none"> <li>✓ Multidimensional analysis</li> <li>✓ Decision making</li> <li>✓ In-depth analysis such as data classification, clustering</li> </ul>
18.	<p><b>List the two different types of reporting tools. (May/June 2014) BTL1</b></p> <ol style="list-style-type: none"> <li>1. Production reporting tools – companies generate regular operational reports or support high volume batch jobs, such as calculating and printing paychecks.</li> <li>2. Report writers – are inexpensive desktop tools designed for end users.</li> </ol>
19.	<p><b>List the two ways the parallel execution of the tasks with in SQL statements can be done. (Nov/Dec 2012)BTL2</b></p> <ul style="list-style-type: none"> <li>✓ Horizontal parallelism – which means that the DB is partitioned across multiple disks and parallel processing within a specific task.</li> <li>✓ Vertical parallelism – which occurs among different tasks – all component query operations (SCAN, JOIN, SORT) are in parallel in a pipelined fashion.</li> </ul>

20.	<b>What are the technologies included in data warehousing? BTL2</b> ✓ Relational and multi-dimensional database management systems																					
	✓ Client/server architecture ✓ Meta data modeling and repositories ✓ Graphical user interfaces and much more																					
21.	<b>Give alternative terms for data mining? (MAY/JUNE 2013) BTL1</b> ✓ Knowledge extraction ✓ Data/pattern analysis ✓ Data archaeology ✓ Data dredging.																					
22.	<b>How OLAP different from OLTP? BTL5 (NOV/DEC 2018)</b> <table border="1"> <tr> <td>Feature</td><td>OLTP</td><td>OLAP</td></tr> <tr> <td>Characteristics</td><td>Operational Processing</td><td>Information Processing</td></tr> <tr> <td>Orientation</td><td>Transaction</td><td>Analysis</td></tr> <tr> <td>User</td><td>Clerk, DBS, Database Professional</td><td>Knowledge worker ( eg: manager, executive, analyst)</td></tr> <tr> <td>Function</td><td>Day-today operations</td><td>Long-term informational requirements decision support</td></tr> <tr> <td>DB design</td><td>ER-based, application oriented</td><td>Star/snowflake, subject-Oriented</td></tr> <tr> <td>Data</td><td>Current, guaranteed up-to date</td><td>Historic, accuracy maintained over time</td></tr> </table>	Feature	OLTP	OLAP	Characteristics	Operational Processing	Information Processing	Orientation	Transaction	Analysis	User	Clerk, DBS, Database Professional	Knowledge worker ( eg: manager, executive, analyst)	Function	Day-today operations	Long-term informational requirements decision support	DB design	ER-based, application oriented	Star/snowflake, subject-Oriented	Data	Current, guaranteed up-to date	Historic, accuracy maintained over time
Feature	OLTP	OLAP																				
Characteristics	Operational Processing	Information Processing																				
Orientation	Transaction	Analysis																				
User	Clerk, DBS, Database Professional	Knowledge worker ( eg: manager, executive, analyst)																				
Function	Day-today operations	Long-term informational requirements decision support																				
DB design	ER-based, application oriented	Star/snowflake, subject-Oriented																				
Data	Current, guaranteed up-to date	Historic, accuracy maintained over time																				
23.	<b>What is the need for backend process in data warehouse design? (MAY/JUNE 2014) BTL4</b> Data warehouse systems use back-end tools and utilities to populate and refresh their data. These tools and facilities include the following functions like extraction , load refresh and transform																					
24.	<b>What are the advantages of dimensional modeling? (MAY/JUNE 2014) BTL3</b> The basic concepts of dimensional modeling are facts, dimensions and measures. A data warehouse, requires a concise, subject-oriented schema that facilitates online data analysis																					
25.	<b>Define Star Schema. (NOV/DEC 2016) BTL1</b> Each dimension in a star schema is represented with only one-dimension table. This dimension table contains the set of attributes																					

26	<b>How data warehouse different from database? (MAY/JUNE 2013) BTL5</b> <ul style="list-style-type: none"> <li>✓ A database is used for Online Transactional Processing (OLTP) but can be used for other purposes such as Data Warehousing.</li> <li>✓ A data warehouse is used for Online Analytical Processing (OLAP). This reads the historical data for the Users for business decisions.</li> <li>✓ In a database the tables and joins are complex since they are normalized for RDMS. This reduces redundant data and saves storage space.</li> <li>✓ In data warehouse, the tables and joins are simple since they are de-normalized. This is done to reduce the response time for analytical queries.</li> <li>✓ Relational modeling techniques are used for RDMS database design, whereas modeling techniques are used for the Data Warehouse design.</li> <li>✓ A database is optimized for write operation, while a data warehouse is optimized for read operations.</li> <li>✓ In a database, the performance is low for analysis queries, while in a data warehouse, there is high performance for analytical queries.</li> <li>✓ A data warehouse is a step ahead of a database. It includes a database in its structure.</li> </ul>
27.	<b>Nine decisions to design a data warehouse. (NOV/DEC 2016) BTL3</b> <ul style="list-style-type: none"> <li>✓ The processes, and hence the identity of the fact tables</li> <li>✓ The grain of each fact table</li> <li>✓ The dimensions of each fact table</li> <li>✓ The facts, including precalculated facts.</li> <li>✓ The dimension attributes with complete descriptions and proper terminology</li> <li>✓ How to track slowly changing dimensions</li> <li>✓ The aggregations, heterogeneous dimensions, minidimensions, query models and other physical storage decisions</li> <li>✓ The historical duration of the database</li> <li>✓ The urgency with which the data is extracted and loaded into the data warehouse</li> </ul>
<b>PART-B</b>	
1.	<b>Explain the three tier architecture of a data warehouse (MAY/JUNE 2012) BTL1 (13 M)</b>  <b>Answer: Page: 8 - Poonkuzhali</b> <ul style="list-style-type: none"> <li>✓ The bottom tier is a warehouse database server - relational database system-Back-end tools and utilities - feed data into- bottom tier- from operational databases (e.g., customer profile information provided by external consultants) ( 5M)</li> <li>✓ The middle tier is an OLAP server - typically implemented using either (1) a relational OLAP (ROLAP) model (i.e., an extended relational DBMS that maps operations on multidimensional data to standard relational operations); or (2) a multidimensional OLAP (MOLAP) model (i.e., a special-purpose server that directly implements multidimensional data and operations). ( 4 M)</li> <li>✓ The top tier is a front-end client layer, - contains query and reporting tools, analysis tools, and/or data mining tools (e.g., trend analysis, prediction, and so on). ( 4 M)</li> </ul>
2	<b>Explain about snow flake schema with examples (13 M) (MAY/JUNE 2012) (NOV/DEC 2018) BTL1 Answer Page : 29-35 - Poonkuzhali</b> <p>Explanation – The snowflake schema is a variant of the star schema model, where some dimension tables are normalized- thereby further splitting the data into additional tables. (6 M)</p> <p>Diagram- (4 M)</p> <p>Example- Example of all student database (3 M)</p>

3	<p><b>Explain in detail about Data warehouse architecture and component (13 M) (MAY/JUNE 2013) (NOV/DEC 2016) BTL1</b></p> <p><b>Answer Page : 2- 10 - Poonkuzhali</b></p> <p>Diagram-(2 M)</p> <p>Explanation -A data warehouse is usually modeled by a multidimensional database structure, where each dimension corresponds to an attribute or a set of attributes in the schema, and each cell stores the value of some aggregate measure, such as count or sales amount. -(5 M)</p> <p>Explain each components and its functionalities (6 M)</p>
4	<p><b>Explain about Design and construction of data warehouse using schemas (13 M) ( MAY/JUNE 2013) BTL3</b></p> <p><b>Answer Page : 29-35- Poonkuzhali</b></p> <p>Explanation - Stars schema- which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension.</p> <p>Snowflakes schema:- The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. The Fact Constellations: - Sophisticated applications may require multiple fact tables to share</p>
	<p>dimension tables. (6 M)</p> <p>Diagram(4 M)</p> <p>Example – Any schemas for of your own profile (3 M)</p>
5	<p><b>Explain the mapping of data warehouse to multiprocessor architecture (13 M) (MAY/JUNE 2014) BTL2</b></p> <p><b>Answer Page: 1.22- 1.30 - Poonkuzhali</b></p> <p>Introduction - To manage large number of client requests efficiently- database vendor's designed parallel hardware architectures -by multiserver and multithreaded systems. This is called interquery parallelism - different server threads handle multiple requests at the same time. (1 M)</p> <p>Data warehouse can be mapped into different type of architectures as follows:</p> <ul style="list-style-type: none"> <li>✓ Shared memory architecture – shares the memory - diagram(4 M)</li> <li>✓ Shared disk architecture- shares the disk - diagram(4 M)</li> <li>✓ Shared nothing architecture-shared nothing- diagram(4 M)</li> </ul>
6	<p><b>Discuss about data warehouse meta data (13 M) (MAY/JUNE 2014) (NOV/DEC 2018) BTL2</b></p> <p><b>Answer: Page : 1.56 – 1.63-Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Definition – data about data (2 M)</li> <li>✓ Meta data Interchange initiative – interchange format- goal of charter- standard specification(2 M)</li> <li>✓ Meta data repository- defines the no of components(3 M)</li> <li>✓ Mata data management- managing the metadata (3 M)</li> <li>✓ Examples : Platinum repository(3M)</li> </ul>
7	<p><b>Discuss DBMS schemas for decision support. Discuss performance problem with decision schemas (13 M) (NOV/DEC 2016) BTL2</b></p> <p><b>Answer Page : 1.35 – 1.40 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Data layout for best access – need to access large number of record(2M)</li> <li>✓ Multidimensional data model- various dimensions of the business- to meet the business query(2M)</li> <li>✓ Star schema</li> <li>Diagram (2M)</li> <li>Explanation – fact table- dimension table (3M)</li> <li>✓ Bitmap indexing – indexing in the table for access(4M)</li> </ul>

8	<p><b>Discuss in detail about access to legacy data. (13 M) BTL2</b></p> <p><b>Answer : Page : 1.45 – 1.47 -Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Definition – scalability and manageability (2M)</li> <li>✓ Explanation- data layer- process layer- user layer (5M)</li> <li>✓ Model I – virtual database – host interface and business transaction (3M)</li> <li>✓ Model II – terminal emulation- peer to peer communication(3M)</li> </ul>
9	<p><b>Describe in detail about data extraction?(13 M) BTL1</b></p> <p><b>Answer : Page : 1.43 – 1.55 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Introduction - extraction of information form the data(1 M)</li> <li>✓ Tools to perform the task – data transformation – data consolidation and integration- meta data synchronization – metadata management(4 M)</li> <li>✓ Vendor Approaches- Separate Products- single integrated solution(4 M)</li> <li>✓ Access Legacy data (4 M)</li> </ul>
	<b>PART C</b>
1	<p><b>(i).Suppose that a data warehouse consists of four dimensions customer, product, salesperson and sales time, and the three measure sales Amt(in rupees), VAT(in rupees) and payment_type(in rupees). Draw the different classes of schemas that are popularly used for modeling data warehouses and explain it. BTL6(15 M)</b></p> <p><b>Answer: Page : 29-35 -Poonkuzhali</b></p> <p>Star Schema - The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension. (5 M)</p> <p>A snowflakes schema - The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake. (5 M)</p> <p>Fact Constellation - Sophisticated applications may require multiple fact tables to share dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation (5 M)</p>
2	<p><b>How would you explain Metadata implementation with examples? (May'13 May'14) (15 M) BTL2</b></p> <p><b>Answer: Page : 1.56 – 1.63 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Definition – data about data (2 M)</li> <li>✓ Meta data Interchange initiative – interchange format- goal of charter- standard specifcation(2 M)</li> <li>✓ Meta data repository- defines the no of components(3 M)</li> <li>✓ Mata data management- managing the metadata (3 M)</li> <li>✓ Examples : Platinum repository(3M)</li> </ul>

3	<p><b>Explain the steps in building a data warehouse. BTL2 (15 M)</b></p> <p><b>Answer Page : 2-10 - Poonkuzhalai</b></p> <p>Introduction -constructing the data ware house(3 M)</p> <p>A top down approach - centralized repository -to house corporate wide business data. This repository is called Enterprise Data Warehouse (EDW). -The data in the EDW is stored in a normalized form in order to avoid redundancy. (4 M)</p> <p>A bottom up approach- Incremental approach to build a data warehouse. Here we build the data marts separately at different points of time as and when the specific subject area requirements are clear. (4 M)</p> <p>Combination of both(4 M)</p>
---	--

## UNIT -II- BUSINESS ANALYSIS

**Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cognos Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OLAP Guidelines – Multidimensional versus Multirelational OLAP – Categories of Tools – OLAP Tools and the Internet.**

### PART \* A

Q. NO	QUESTIONS
1	<p><b>What are production reporting tools? Give examples. (May/June2013) BTL1</b>          Production reporting tools will let companies generate regular operational reports or support high-volume batch jobs. Such as calculating and printing pay checks.</p> <p><b>Examples:</b></p> <ul style="list-style-type: none"> <li>✓ Third generation languages such as COBOL</li> <li>✓ Specialized fourth generation languages such as Information builders, Inc's Focus</li> <li>✓ High-end client/server tools such as MITI's SQL.</li> </ul>
2	<p><b>Define data cube. (May/June2013) BTL1</b>          Data cube consists of a large set of facts or measures and a number of dimensions. Facts are numerical measures that are quantities by which we can analyze the relationship between dimensions. Dimensions are the entities or perspectives with respect to an organization for keeping records and are hierarchical nature.</p>
3	<p><b>What is a Reporting tool? List out the two different types of reporting tools. (May/June 2014, Nov/Dec 2012) BTL1</b>          Reporting tools are software applications that make data extracted in a query accessible to the user. That is it used for to generate the various types of reports.</p> <p>It can be divided into 2 types:</p> <ul style="list-style-type: none"> <li>✓ Production reporting tools</li> <li>✓ Desktop reporting tools</li> </ul>
4	<p><b>DefineOLAP. (May/June2014) (NOV/DEC 2018) BTL1</b></p> <ul style="list-style-type: none"> <li>✓ OLAP (online analytical processing) is computer processing that enables a user to easily and selectively extract and view data from different points of view.</li> <li>✓ OLAP is becoming an architecture that an increasing number of enterprises are implementing to support analytical applications.</li> </ul>
5	<p><b>Briefly discuss the schemas for multidimensional databases. (NOV/DEC 2018) (May/June 2010, Nov/Dec 2014, May/June 2011) BTL4</b></p> <ul style="list-style-type: none"> <li>✓ <b>Stars schema:</b> The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension.</li> <li>✓ <b>Snowflakes schema:</b> The snowflake schema is a variant of the star schema model, where some dimension tables are <i>normalized</i>, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.</li> <li>✓ <b>Fact Constellations:</b> Sophisticated applications may require multiple fact tables to <i>share</i> dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation.</li> </ul>
6	<p><b>Define the categories of tools in business analysis.(Nov/Dec2014) BTL1</b>          There are 5 categories of tools in business analysis.</p>

	<ul style="list-style-type: none"> <li>✓ Reporting tools – it can be used to generate there ports.</li> <li>✓ Managed query tools – it can be used to SQL queries for accessing the databases.</li> <li>✓ Executive information systems – It allow developers to build customized, graphical decision support applications or “briefing books”.</li> <li>✓ On-line analytical processing – these tools aggregate data along common business subjects or dimensions and then let users navigate the hierarchies and dimensions with the click of a mousebutton.</li> <li>✓ <b>Data mining</b> – It use a variety of statistical and artificial intelligence algorithm to analyze the correlation of variables in the data and extract interesting patterns and relationship to investigate.</li> </ul>									
7	<p><b>Differentiate MOLAP, ROLAP and HOLAP.(Nov/Dec2013) BTL2</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;"><b>MOLAP</b></th><th style="text-align: center;"><b>ROLAP</b></th><th style="text-align: center;"><b>HOLAP</b></th></tr> </thead> <tbody> <tr> <td>MOLAP stands for Multidimensional Online Analytical Processing</td><td>ROLAP stands for Relational Online Analytical Processing</td><td>HOLAP stands for Hybrid Online Analytical Processing</td></tr> <tr> <td>The MOLAP storage mode causes the aggregations of the partition and a copy of its source data to be stored in a multidimensional structure in Analysis Services when the partition is processed.</td><td>The ROLAP storage mode causes the aggregations of the partition to be stored in indexed views in the relational database that was specified in the partition's data source.</td><td>The HOLAP storage mode combines attributes of both MOLAP and ROLAP. Like MOLAP, HOLAP causes the aggregations of the partition to be stored in a multidimensional structure in an SQL Server Analysis Services instance.</td></tr> </tbody> </table>	<b>MOLAP</b>	<b>ROLAP</b>	<b>HOLAP</b>	MOLAP stands for Multidimensional Online Analytical Processing	ROLAP stands for Relational Online Analytical Processing	HOLAP stands for Hybrid Online Analytical Processing	The MOLAP storage mode causes the aggregations of the partition and a copy of its source data to be stored in a multidimensional structure in Analysis Services when the partition is processed.	The ROLAP storage mode causes the aggregations of the partition to be stored in indexed views in the relational database that was specified in the partition's data source.	The HOLAP storage mode combines attributes of both MOLAP and ROLAP. Like MOLAP, HOLAP causes the aggregations of the partition to be stored in a multidimensional structure in an SQL Server Analysis Services instance.
<b>MOLAP</b>	<b>ROLAP</b>	<b>HOLAP</b>								
MOLAP stands for Multidimensional Online Analytical Processing	ROLAP stands for Relational Online Analytical Processing	HOLAP stands for Hybrid Online Analytical Processing								
The MOLAP storage mode causes the aggregations of the partition and a copy of its source data to be stored in a multidimensional structure in Analysis Services when the partition is processed.	The ROLAP storage mode causes the aggregations of the partition to be stored in indexed views in the relational database that was specified in the partition's data source.	The HOLAP storage mode combines attributes of both MOLAP and ROLAP. Like MOLAP, HOLAP causes the aggregations of the partition to be stored in a multidimensional structure in an SQL Server Analysis Services instance.								
8	<p><b>List any four tools for performing OLAP.(Nov/Dec2013) BTL2</b></p> <ul style="list-style-type: none"> <li>✓ Arbor Essbase Web</li> <li>✓ Information advantage webOLAP</li> <li>✓ Micro strategy DSSweb</li> <li>✓ Brio technology</li> </ul>									
9	<p><b>Classify OLAP tools. (Apr/May2011) BTL1</b></p> <ul style="list-style-type: none"> <li>✓ MOLAP – Multidimensional Online AnalyticalProcessing</li> <li>✓ ROLAP – Multirelational Online Analytical Processing</li> <li>✓ MQE – Managed QueryEnvironment</li> </ul>									
10	<p><b>How the complex aggregation at multiple granularities is achieved using multi-feature cubes?(May/June2012) BTL1</b></p> <p>Multi-feature cubes, which compute complex queries involving multiple dependent aggregates at multiple granularity. These cubes are very useful in practice. Many complex data mining queries can be answered by multi-feature cubes without any significant increase in computational cost, in comparison to cube computation for simple queries with standard data cubes.</p>									

11	<b>Give examples for managed query tools.(Nov/Dec2012) BTL4</b> <ul style="list-style-type: none"> <li>✓ IQ software's IQobjects</li> <li>✓ Andyne Computing Ltd's GQL</li> <li>✓ IBM's Decisionserver</li> <li>✓ Oracle Corp's Discoverer/2000</li> </ul>																								
12	<b>What is Apex cuboid? (Apr/May 2011,Nov/Dec 2011) BTL1</b> <ul style="list-style-type: none"> <li>✓ Apex cuboid or 0-D cuboid which holds the highest level of summarization.</li> <li>✓ The Apex cuboid is typically denoted by all.</li> </ul>																								
13	<b>What is multidimensional database? (Nov/Dec2011) BTL4</b> <p>Data warehouses and OLAP tools are based on a multidimensional data model. This model is used for the design of corporate data warehouses and department data marts. This model contains a star schema, snowflake schema and fact constellation schemas. The core of multidimensional model is the data cube.</p>																								
14	<b>What are the applications of query tools? (Nov/Dec2014) BTL1</b> <p>The applications of query tools are</p> <ul style="list-style-type: none"> <li>✓ Multidimensional analysis</li> <li>✓ Decision making</li> <li>✓ In-depth analysis such as data classification</li> <li>✓ Clustering.</li> </ul>																								
15	<b>Compare OLTP and OLAP. (Apr/May 2008,May/June 2010) (NOV/DEC 2018) BTL5</b> <table border="1" data-bbox="285 925 1444 1917"> <thead> <tr> <th><b>Data Warehouse (OLAP)</b></th><th><b>Operational Database (OLTP)</b></th></tr> </thead> <tbody> <tr> <td>Involves historical processing of information.</td><td>Involves day-to-day processing.</td></tr> <tr> <td>OLAP systems are used by knowledge workers such as executives, managers and</td><td>OLTP systems are used by clerks, DBAs, or database professionals.</td></tr> <tr> <td>Useful in analyzing the business.</td><td>Useful in running the business.</td></tr> <tr> <td>It focuses on Information out.</td><td>It focuses on Data in.</td></tr> <tr> <td>Based on Star Schema, Snowflake, Schema and Fact Constellation Schema.</td><td>Based on Entity Relationship Model.</td></tr> <tr> <td>Contains historical data.</td><td>Contains current data.</td></tr> <tr> <td>Provides summarized and consolidated data.</td><td>Provides primitive and highly detailed data.</td></tr> <tr> <td>Provides summarized and multidimensional view of data.</td><td>Provides detailed and flat relational view of data.</td></tr> <tr> <td>Number of users is in hundreds.</td><td>Number of users is in thousands.</td></tr> <tr> <td>Number of records accessed is in millions.</td><td>Number of records accessed is in tens.</td></tr> <tr> <td>Database size is from 100 GB to 1 TB</td><td>Database size is from 100 MB to 1 GB.</td></tr> </tbody> </table>	<b>Data Warehouse (OLAP)</b>	<b>Operational Database (OLTP)</b>	Involves historical processing of information.	Involves day-to-day processing.	OLAP systems are used by knowledge workers such as executives, managers and	OLTP systems are used by clerks, DBAs, or database professionals.	Useful in analyzing the business.	Useful in running the business.	It focuses on Information out.	It focuses on Data in.	Based on Star Schema, Snowflake, Schema and Fact Constellation Schema.	Based on Entity Relationship Model.	Contains historical data.	Contains current data.	Provides summarized and consolidated data.	Provides primitive and highly detailed data.	Provides summarized and multidimensional view of data.	Provides detailed and flat relational view of data.	Number of users is in hundreds.	Number of users is in thousands.	Number of records accessed is in millions.	Number of records accessed is in tens.	Database size is from 100 GB to 1 TB	Database size is from 100 MB to 1 GB.
<b>Data Warehouse (OLAP)</b>	<b>Operational Database (OLTP)</b>																								
Involves historical processing of information.	Involves day-to-day processing.																								
OLAP systems are used by knowledge workers such as executives, managers and	OLTP systems are used by clerks, DBAs, or database professionals.																								
Useful in analyzing the business.	Useful in running the business.																								
It focuses on Information out.	It focuses on Data in.																								
Based on Star Schema, Snowflake, Schema and Fact Constellation Schema.	Based on Entity Relationship Model.																								
Contains historical data.	Contains current data.																								
Provides summarized and consolidated data.	Provides primitive and highly detailed data.																								
Provides summarized and multidimensional view of data.	Provides detailed and flat relational view of data.																								
Number of users is in hundreds.	Number of users is in thousands.																								
Number of records accessed is in millions.	Number of records accessed is in tens.																								
Database size is from 100 GB to 1 TB	Database size is from 100 MB to 1 GB.																								

		Highly flexible.	Provides high performance.	
16	<b>List out OLAP operations in multidimensional data model. (May/June2009) BTL1</b>	<ul style="list-style-type: none"> <li>✓ Roll-up - performs aggregation on a datacube</li> <li>✓ Drill-down - is the reverse operation of roll-up.</li> <li>✓ Slice and dice – Slice operation selects one particular dimension from a given cube and provides a new sub-cube. Dice selects two or more dimensions from a given cube and provides a new sub-cube.</li> <li>✓ Pivot (or) rotate - The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data.</li> </ul>		
17	<b>Mention the functions of OLAP servers in the data warehousing architecture. (Nov/Dec 2010) BTL4</b>	The OLAP server performs multidimensional queries of data and stores the results in its multidimensional storage. It speeds the analysis of fact tables into cubes, stores the cubes until needed, and then quickly returns the data to clients.		
18	<b>What is impromptu? BTL1</b>	Impromptu from Cognos Corporation is positioned as an enterprise Answer: for interactive database reporting that delivers 1 to 100+ seat scalability.		
19	<b>Mention some supported databases of Impromptu. BTL1</b>	<ul style="list-style-type: none"> <li>✓ ORACLE</li> <li>✓ Microsoft SQLServer</li> <li>✓ SYBASE</li> <li>✓ Omni SQLGateway</li> <li>✓ SYBASE NetGateway</li> </ul>		
20	<b>What is enterprise warehouse? BTL1</b>	An enterprise warehouse collects all the information's about subjects spanning the entire organization. It provides corporate-wide data integration, usually from one or more operational systems or external information providers. It contains detailed data as well as summarized data and can range in size from a few giga bytes to hundreds of giga bytes, tera bytes or beyond.		
21.	<b>Write a note on Report writers. BTL1</b>	<ul style="list-style-type: none"> <li>✓ Report writers are inexpensive desktop tools designed for endusers.</li> <li>✓ Report writers have graphical interfaces and built-in charting functions; they can pull groups of data from variety of data sources and integrate them in a single report.</li> <li>✓ Leading report writers include Crystal Reports, Actuate and Platinum technology, Inc's Inforeports.</li> </ul>		
22.	<b>What is the use of knowledge base? (MAY/JUNE 2013) BTL2</b>	Knowledge base: This is the domain knowledge that is used to guide the search or evaluate the interestingness of resulting patterns. Such knowledge can include concept hierarchies, used to organize attributes or attribute values into different levels of abstraction.		
23.	<b>Define concept hierarchy. (MAY/JUNE 2013) BTL1</b>	A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level, more general concepts		
24.	<b>Define OLAP. (MAY/JUNE 2014) BTL1</b>	OLAP is an acronym for Online Analytical Processing. OLAP performs multidimensional analysis of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling.		
25.	<b>List out the different types of reporting tools? (MAY/JUNE 2014) BTL1</b>	<ul style="list-style-type: none"> <li>✓ Production reporting tool</li> </ul>		

	<ul style="list-style-type: none"> <li>✓ Report writers</li> </ul>
26.	<p><b>OLAP Guidelines (NOV/DEC 2016) BTL1</b></p> <ul style="list-style-type: none"> <li>✓ <b>Multidimensional conceptual view:</b> The OLAP should provide an appropriate multidimensional Business model that suits the Business problems and Requirements.</li> <li>✓ <b>Transparency:</b> The OLAP tool should provide transparency to the input data for the users.</li> <li>✓ <b>Accessibility:</b> The OLAP tool should only access the data required only to the analysis needed.</li> <li>✓ <b>Consistent reporting performance:</b> The Size of the database should not affect in any way the performance.</li> </ul>
27.	<p><b>Comment on OLAP tools on internet (NOV/DEC 2016) BTL5</b></p> <ul style="list-style-type: none"> <li>✓ The internet provides connectivity between countries acting as a free resource.</li> <li>✓ The web eases administrative tasks of managing scattered locations.</li> <li>✓ The Web allows users to store and manage data and applications on servers that can be managed, maintained and updated centrally.</li> </ul>

### PART- B

1	<p><b>Enumerate the Features of Cognus impromptu (13 M)(MAY/JUNE 2012 ) (NOV/DEC 2018)BTL1</b></p> <p><b>Answer: Page : 2.5 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Introduction -Impromptu from Cognos Corporation is positioned as an enterprise (6 M)</li> <li>✓ Catalog- Impromptu information catalog(3 M)</li> <li>✓ Implementation -architecture -reporting (4 M)</li> </ul>															
2	<p><b>Explain in detail about OLAP operations (MAY/JUNE 2012) BTL4 (13 M)</b></p> <p><b>Answer: Page : 73 -Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Roll-up - performs aggregation on a datacube (3 M)</li> <li>✓ Drill-down - is the reverse operation of roll-up. (3 M)</li> <li>✓ Slice and dice – Slice operation selects one particular dimension from a given cube and provides a new sub-cube. Dice selects two or more dimensions from a given cube and provides a new sub-cube. (3 M)</li> <li>✓ Pivot (or) rotate - The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. (4 M)</li> </ul>															
3	<p><b>Write the difference between multi-dimensional OLAP and multi-relational OLAP (MAY/JUNE 2014) BTL2 (13 M)</b></p> <p><b>Answer: Page : 64 – Poonkuzhali</b></p> <p>Introduction (3 M)</p> <p>Table (10 M)</p> <table border="1"> <thead> <tr> <th>Characteristics</th> <th>ROLAP</th> <th>MOLAP</th> </tr> </thead> <tbody> <tr> <td>Schema</td> <td>Uses star Schema -Additional dimensions can be added dynamically.</td> <td>Uses Data cubes -Additional dimensions require recreation of data cube.</td> </tr> <tr> <td>Database Size</td> <td>Medium to large</td> <td>Small to medium</td> </tr> <tr> <td>Architecture</td> <td>Client/Server</td> <td>Client/Server</td> </tr> <tr> <td>Access</td> <td>Support ad-hoc requests</td> <td>Limited to pre-defined dimensions</td> </tr> </tbody> </table>	Characteristics	ROLAP	MOLAP	Schema	Uses star Schema -Additional dimensions can be added dynamically.	Uses Data cubes -Additional dimensions require recreation of data cube.	Database Size	Medium to large	Small to medium	Architecture	Client/Server	Client/Server	Access	Support ad-hoc requests	Limited to pre-defined dimensions
Characteristics	ROLAP	MOLAP														
Schema	Uses star Schema -Additional dimensions can be added dynamically.	Uses Data cubes -Additional dimensions require recreation of data cube.														
Database Size	Medium to large	Small to medium														
Architecture	Client/Server	Client/Server														
Access	Support ad-hoc requests	Limited to pre-defined dimensions														

	Characteristics	ROLAP	MOLAP	
Resources	High	Very High		
Flexibility	High	Low		
Scalability	High	Low		
Speed	Good with small data sets. Average for medium to large data set.	Faster for small to medium data sets. Average for large data sets.		

4 **Discuss different tool categories data warehouse business analysis (NOV/DEC 2016) BTL1 (13 M)**

**Answer: Page: 2.1- Poonkuzhali**

- ✓ Reporting Tool – reporting the status(3 M)
- ✓ Managed Query- managing the database(3 M)
- ✓ Executive Information System- used for executional part(3 M)
- ✓ OLAP- transactional data (2 M)
- ✓ Data Mining- extracted data from warehouse(2 M)

5 **Summarize the various OLAP operations in the Multidimensional Data Model. BTL3 (13 M)**

**Answer: Page : 73-Poonkuzhali**

- ✓ Roll-up - performs aggregation on a datacube (3 M)
- ✓ Drill-down - is the reverse operation of roll-up. (3 M)
- ✓ Slice and dice – Slice operation selects one particular dimension from a given cube and provides a new sub-cube. Dice selects two or more dimensions from a given cube and provides a new sub-cube. (3 M)
- ✓ Pivot (or) rotate - The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. (4 M)

6 **How would you describe in detail about reporting query? Describe in detail about application reporting query? BTL4 (13 M)**

**Answer: Page : 45 – 48-Poonkuzhali**

- ✓ Introduction (2 M)
- ✓ Reporting tools with explanation (8 M)
  - Rich, interactive display – Wide variety of tables, charts, graphs and other visual BI tools can be configured and linked to source data to generate interactive data visualizations
  - Share reports via a web browser – Interactive reports can be quickly shared through a web browser or any mobile device.
  - Unify disparate data sources – Use data from multiple sources in a single report, including data from Excel, text/CSV files, any database (SQL Server, Oracle, MySQL), and Google platforms
  - Automatic and manual data refresh – Reports can be refreshed manually or automatically at pre-defined intervals
  - Fast query response – Query response is in seconds, even when dealing with huge amounts of data or working off commodity hardware
- ✓ Applications (3 M)

### PART- C

1 **Explain the feature of Cognosimprontu business analysis tool. BTL1 (15 M)  
(NOV/DEC 2018) Answer: Page : 2.5 -Poonkuzhali**

- ✓ Introduction -Impromptu from Cognos Corporation is positioned as an enterprise (6 M)
- ✓ Catalog- Impromptu information catalog(5 M)

	<ul style="list-style-type: none"> <li>✓ Implementation -architecture -reporting (4 M)</li> </ul>
2	<p><b>(i)Examine how a client/server application can be constructed using Power Builder painters. (7 M) BTL6</b></p> <p><b>Answer: Page : 2.9- Poonkuzhali</b></p> <p><b>Applications</b></p> <ul style="list-style-type: none"> <li>✓ About power builder painters (2M)</li> <li>✓ Architecture (4 M)</li> <li>✓ Benefits(1 M)</li> </ul> <p><b>(ii)Illustrate the three functional components offers to support the life cycle of Forte. (8 M) (BTL6)</b></p> <p><b>Answer: Page no :2.15 -Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ About Forte(2M)</li> <li>✓ Architecture and components(5 M)</li> <li>Benefits(1 M)</li> </ul>
3	<p><b>Explain in detail about the need for the architectural styles, trends and major players of OLAP. BTL6 (15 M)</b></p> <p><b>Answer: Page : 2.18- Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ OLAP introduction OLAP On-Line Analytical Processing - Long transactions, usually complex queries(e.g., all statistics about all sales, grouped by dept and month) has “Data mining” operations and Infrequent update (3M)</li> <li>✓ OLAP working- transactional data (5 M)</li> <li>✓ Components of OLAP (5M)</li> <li>✓ Applications(2 M)</li> </ul>
4	<p><b>Examine the relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing. BTL5 (15 M)</b></p> <p><b>Answer: Page : 64-Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Multidimensional online analytical processing data (4M)</li> <li>✓ Multi relational online analytical processing. (4M)</li> </ul> <p>Table ( 7 M)</p>

<b>Characteristics</b>	<b>ROLAP</b>	<b>MOLAP</b>
Schema	Uses star Schema -Additional dimensions can be added dynamically.	Uses Data cubes -Addition dimensions require recreation of data cube.
Database Size	Medium to large	Small to medium
Architecture	Client/Server	Client/Server
Access	Support ad-hoc requests	Limited to pre-defined dimensions
Characteristics	ROLAP	MOLAP
Resources	High	Very High
Flexibility	High	Low
Scalability	High	Low
Speed	<ul style="list-style-type: none"> <li>• Good with small data sets.</li> <li>• Average for medium to large data set.</li> </ul>	<ul style="list-style-type: none"> <li>• Faster for small to medium data sets.</li> <li>• Average for large data sets.</li> </ul>

5	<p><b>Design a multidimensional data model for hospital data warehouse consists of three dimensions time, doctor and the patient and the two measures count and charge, where charge is a fee that a doctor charges a patient for a visit.</b></p> <ol style="list-style-type: none"> <li><b>1. Enumerate the different classes of schemas that are popularly used for modeling data warehouses.</b></li> <li><b>2. Draw a schema diagram for the above data warehouse using all of the schemas classes listed in 1. BTL6</b></li> </ol> <p><b>Answer: Page :1.35 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Stars schema: The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension.</li> <li>✓ Snowflakes schema: The snowflake schema is a variant of the star schema model, where some dimension tables are <i>normalized</i>, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.</li> <li>✓ Fact Constellations: Sophisticated applications may require multiple fact tables to <i>share</i> dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation. (3M)</li> <li>✓ Draw the three schemas for the given scenario (3* 4M – 12M)</li> </ul>
---	---

### UNIT 3- DATA MINING

**Introduction – Data – Types of Data – Data Mining Functionalities – Interestingness of Patterns – Classification of Data Mining Systems – Data Mining Task Primitives – Integration of a Data Mining System with a Data Warehouse – Issues –Data Preprocessing.**

#### PART A

<b>1. Define data mining. Give some alternative terms of data mining. BTL1</b>	<ul style="list-style-type: none"> <li>✓ <i>Data mining</i> refers to extracting or “mining” knowledge from large amounts of data. Data mining is a process of discovering interesting knowledge from large amounts of data stored either, in database, data warehouse or other information repositories.</li> </ul> <p><b>Alternative names are</b></p> <ul style="list-style-type: none"> <li>✓ Knowledge mining</li> <li>✓ Knowledge extraction</li> <li>✓ Data/pattern analysis</li> <li>✓ Data Archaeology</li> <li>✓ Data Dredging</li> </ul>
<b>2. What is KDD? What are the steps involved in KDD process? BTL1(NOV/DEC 2018)</b>	<p><i>Knowledge discovery in databases</i> (KDD) is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results.</p> <p><b>The steps involved in KDD process are</b></p> <ul style="list-style-type: none"> <li>✓ <b>Data Cleaning</b> – In this step, the noise and inconsistent data is removed</li> <li>✓ <b>Data Integration</b> – In this step, multiple data sources are combined</li> <li>✓ <b>Data Selection</b> – In this step, data relevant to the analysis task are retrieved from the database</li> <li>✓ <b>Data Transformation</b> – In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations</li> <li>✓ <b>Data Mining</b> – In this step, intelligent methods are applied in order to extract data patterns</li> <li>✓ <b>Pattern Evaluation</b> – In this step, to identify the truly interesting patterns representing knowledge based on some interestingness measures</li> <li>✓ <b>Knowledge Presentation</b> – In this step, visualization and knowledge representation techniques are used to present the mined knowledge to the user</li> </ul>
<b>3. What are the various forms of data preprocessing? (Apr/May 2008) BTL1</b>	<ul style="list-style-type: none"> <li>✓ Data cleaning</li> <li>✓ Data integration</li> <li>✓ Data transformation</li> <li>✓ Data reduction</li> </ul>
<b>4. State why preprocessing an important issue for data warehousing and data mining? (Apr/May 2011) BTL3</b>	<p>In real world data tend to be incomplete, noisy and inconsistent data. So preprocessing is an important issue for data warehousing and data mining.</p>
<b>5. Write the 2 measures of association rule. (Apr/May 2008) BTL5</b>	<ul style="list-style-type: none"> <li>✓ <i>Support</i> – It means how often X and Y occur together as a percentage of the total transaction</li> <li>✓ <i>Confidence</i> – It measures how much a particular item is dependent on another</li> </ul>
<b>6. What is descriptive and predictive data mining? (May/June 2014) BTL1</b>	<ul style="list-style-type: none"> <li>✓ The <i>descriptive data-mining</i> model is discover patterns in the data and understands the</li> </ul>

	<p>relationships between attributes represented by the data</p> <ul style="list-style-type: none"> <li>✓ In contrast, the <b><i>predictive data-mining</i></b> model predicts the future outcomes based on passed records present in the database or with known answers</li> </ul>				
7.	<p><b>What is data transformation? (Nov/Dec 2014) BTL1</b></p> <p>In <i>data transformation</i>, the data are transformed or consolidated into forms appropriate for mining. Strategies for data transformation include the following:</p> <ul style="list-style-type: none"> <li>✓ Smoothing</li> <li>✓ Aggregation</li> <li>✓ Generalization of the data</li> <li>✓ Normalization</li> <li>✓ Attribute construction</li> </ul>				
8.	<p><b>What is data cleaning? (Nov/Dec 2007,Nov/Dec 2011,May/June 2013) BTL1</b></p> <p><i>Data cleaning</i> is a process used to determine inaccurate, incomplete or unreasonable data and then improve the quality through correcting of detected errors and omissions. Generally data cleaning reduces errors and improves the data quality.</p>				
9.	<p><b>List some applications of Data Mining. (May/June 2008) BTL2</b></p> <ul style="list-style-type: none"> <li>✓ Financial Data Analysis</li> <li>✓ Retail Industry</li> <li>✓ Telecommunication Industry</li> <li>✓ Biological Data Analysis</li> <li>✓ Other Scientific Applications</li> <li>✓ Intrusion Detection</li> </ul>				
10.	<p><b>What is pattern evaluation? (May/June 2013) BTL1</b></p> <p>This is one of the steps in the KDD process. In this step, the patterns obtained in the data mining stage are converted in to knowledge based on some interestingness measures.</p>				
11.	<p><b>List the primitives that specify a data mining tasks. (May/June 2012) BTL2</b></p> <ul style="list-style-type: none"> <li>✓ Task-relevant data</li> <li>✓ Knowledge type to be mined</li> <li>✓ Background knowledge</li> <li>✓ Pattern interestingness measure</li> <li>✓ Visualization of discovered patterns</li> </ul>				
12.	<p><b>Differentiate between data characterization and discrimination. (Nov/Dec 2013) BTL4</b></p> <table border="1"> <thead> <tr> <th style="text-align: center;"><b>Data Characterization</b></th><th style="text-align: center;"><b>Data Discrimination</b></th></tr> </thead> <tbody> <tr> <td>Characterization is a Summarization of the general characteristics or features of a target class of data.</td><td>Discrimination is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.</td></tr> </tbody> </table>	<b>Data Characterization</b>	<b>Data Discrimination</b>	Characterization is a Summarization of the general characteristics or features of a target class of data.	Discrimination is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.
<b>Data Characterization</b>	<b>Data Discrimination</b>				
Characterization is a Summarization of the general characteristics or features of a target class of data.	Discrimination is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.				
13.	<p><b>State why concept hierarchies are useful in data mining. (Nov/Dec 2012) BTL5</b></p> <p><i>Concept hierarchies</i> define a sequence of mappings from a set of lower-level concepts to higher-level, more general concepts and can be represented as a set of nodes organized in a tree, in the form of a lattice, or as a partial order. They are useful in data mining because they allow the discovery of knowledge at multiple levels of abstraction and provide the structure on which data can be generalized (rolled-up) or specialized (drilled- down). Together, these operations allow users to view the data from different perspectives, gaining further insight into relationships hidden in the data. Generalizing has the advantage of compressing the data set, and mining on a compressed data set will require fewer I/O operations. This will be more efficient than mining on a large, uncompressed data set.</p>				

14.	<b>What do data mining functionalities include? (Apr/May 2011) BTL1</b> Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. Data mining tasks can be classified in to 2 categories: Descriptive and Predictive.
15.	<b>What is classification? (May/June 2011) BTL1</b> Classification involves finding rules that partition the data into disjoint groups. The input for the classification is the training data set, whose class labels are already known. Classification analyzes the training data set and contracts a model based on the class label and aims to assign a class label to the future unlabeled records.
16.	<b>Describe challenges to data mining regarding performance issues. BTL2</b> ✓ Efficiency and scalability of data mining algorithms ✓ Parallel, distributed and incremental mining algorithms
17.	<b>What is prediction? BTL1</b> Prediction is used to predict missing or unavailable data values rather than class labels. Prediction refers to both data value prediction and class label prediction.
18.	<b>What are outliers? BTL1</b> Data objects which differ significantly from the remaining data objects are referred to as outliers.
19.	<b>What are the two steps using in data cleaning as a process? BTL2</b> ✓ Discrepancy detection ✓ Data transformation
20.	<b>List the issues of data integration. BTL2</b> ✓ Schema integration and object matching ✓ Redundancy ✓ Detection and resolution of data value conflict
21.	<b>List out data mining functionalities. (MAY/JUNE 2012) BTL2</b> ✓ Classification ✓ Clustering ✓ Prediction ✓ Characterization
22.	<b>List out steps involved in association rule mining. (MAY/JUNE 2013) BTL1</b> In general, association rule mining can be viewed as a two-step process: ✓ Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min sup ✓ Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence
23.	<b>What is the purpose of Apriori algorithm? (MAY/JUNE 2013) BTL2</b> Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties.
24.	<b>What is legacy database? (MAY/JUNE 2014) BTL1</b> A legacy database is a group of <i>heterogeneous databases</i> that combines different kinds of data systems, such as relational or object-oriented databases, hierarchical databases, network databases, spreadsheets, multimedia databases, or file systems
25.	<b>What is descriptive and predictive data modeling? (MAY/JUNE 2014) BTL1</b> <b>Descriptive</b> Analytics, which use data aggregation and data mining techniques to provide insight into the past and answer: "What has happened?" <b>Predictive</b> Analytics, which use statistical models and forecasts techniques to understand the future and answer: "What could happen?"

26.	<p><b>List out steps involved in association rule mining. (MAY/JUNE 2015) BTL2</b></p> <p>In general, association rule mining can be viewed as a two-step process:</p> <ul style="list-style-type: none"> <li>✓ Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min sup</li> <li>✓ Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence</li> </ul>
27.	<p><b>Outlier analysis. (NOV/DEC 2016) BTL1</b></p> <p>An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism. The data objects that do not comply with the general behavior or model of the data and inconsistent with remaining set of data are called as outliers</p>
28.	<p><b>Steps in data transformation. (NOV/DEC 2016) BTL2</b></p> <ul style="list-style-type: none"> <li>✓ Smoothing</li> <li>✓ Aggregation</li> <li>✓ Generalization</li> <li>✓ Normalization</li> <li>✓ Attribute construction</li> </ul>
29.	<p><b>List the data reduction strategies. BTL1</b></p> <ul style="list-style-type: none"> <li>✓ Data cube aggregation</li> <li>✓ Attribute subset selection</li> <li>✓ Dimensionality reduction</li> <li>✓ Numerosity reduction</li> <li>✓ Discretization and concept hierarchy generation</li> </ul>

## PART B

1	<p><b>Explain about Data mining task primitives (MAY/JUNE 2012) BTL2 (13 M)</b></p> <p><b>Answer: Page : 3.41 - Poonkuzhali</b></p> <p>Introduction :Data mining - extracting or “mining” knowledge from large amounts of data - many other terms – such as knowledge mining from data- knowledge extraction- data/pattern analysis- data archaeology- data dredging (1 M)</p> <ul style="list-style-type: none"> <li>✓ The set of task-relevant data (2 M)</li> <li>✓ The kind of knowledge to be mined: (2 M)</li> <li>✓ The background knowledge (2 M)</li> <li>✓ The interestingness measures and thresholds (3 M)</li> <li>✓ The expected representation for visualizing (3 M)</li> </ul>
2	<p><b>Explain about the Steps in knowledge discovery from databases (MAY/JUNE 2012) BTL2 (13 M)</b></p> <p><b>Answer: Page : 3.3 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Introduction -Knowledge discovery as a process is depicted in following figure and consists of an iterative sequence of the following steps (1 M)</li> <li>✓ data cleaning : to remove noise or irrelevant data (2 M)</li> <li>✓ data integration: where multiple data sources may be combined (2 M)</li> <li>✓ data selection: where data relevant to the analysis task are retrieved from the database(2 M)</li> <li>✓ data transformation: where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations (2 M)</li> <li>✓ data mining :an essential process where intelligent methods are applied in order to extract data patterns (2 M)</li> <li>✓ Pattern evaluation: to identify the truly interesting patterns representing knowledge based on some</li> </ul>

	interestingness measures knowledge presentation: where visualization and knowledge representation techniques are used to present the mined knowledge to the user(2 M)
3	<p><b>What is the use of data mining task? What are the basic types of data mining tasks? Explain with examples. (MAY/JUNE 2014) (NOV/DEC 2018)BTL3 (13 M)</b></p> <p><b>Answer: Page : 3.5- Poonkuzhalai</b></p> <p>Introduction : Introduction :Data mining - extracting or “mining” knowledge from large amounts of data - many other terms – such as knowledge mining from data- knowledge extraction- data/pattern analysis- data archaeology- data dredging (1 M)</p> <ul style="list-style-type: none"> <li>✓ Relational databases (1 M)</li> <li>✓ Data warehouse (2 M)</li> <li>✓ Transactional database(2 M)</li> <li>✓ Advanced data bases(2 M)</li> <li>✓ Temporal data base(1 M)</li> <li>✓ Text and multimedia database (2 M)</li> <li>✓ Heterogeneous database(1 M)</li> <li>✓ Legacy databases(1 M)</li> </ul>
4	<p><b>What is interestingness of a pattern? What approach would you designed to mine interestingness of patterns? BTL3 (13 M)</b></p> <p><b>Answer: Page: 3.41 - Poonkuzhalai</b></p> <p>A data mining system/query may generate thousands of patterns, not all of them are interesting. (1 M)</p> <p><b>Interestingness measures (4 M)</b></p> <ul style="list-style-type: none"> <li>✓ Objective vs. subjective interestingness measures</li> <li>✓ Find all the interesting patterns: Completeness</li> <li>✓ Search for only interesting patterns: Optimization</li> <li>✓ Pattern evaluation: refers to interestingness of pattern</li> </ul> <p><b>Performance issues (4 M)</b></p> <ul style="list-style-type: none"> <li>✓ Efficiency and scalability of data mining algorithms</li> <li>✓ Parallel, distributed, and incremental updating algorithms</li> </ul> <p><b>Issues relating to the diversity of database types (4 M)</b></p> <ul style="list-style-type: none"> <li>• Handling of relational and complex types of data</li> </ul> <p>Mining information from heterogeneous databases and global information systems</p>
5	<p><b>Explain the various data mining functionalities in detail. BTL2 (13 M)</b></p> <p><b>Answer: Page : 3.8 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Concept and class description - Data mining functionalities - to specify the kind of patterns to be found in data mining tasks. In general, data mining tasks can be classified into two categories: Descriptive – the process to define each data record predictive- used for future prediction of data (2 M)</li> <li>✓ Association analysis – discovery of association rules (2 M)</li> <li>✓ Classification and Prediction – classification that partitions the data into disjoint sets – prediction- predicting missing or unavailable data (2 M)</li> <li>✓ Cluster analysis-grouping data into different groups (2 M)</li> <li>✓ Outlier analysis – identifying exceptions or rare groups(2 M)</li> <li>✓ Evolution analysis- predict by effective decision making (2 M)</li> <li>✓ Diagram, (1 M)</li> </ul>

	<pre> graph TD     DM[Data Mining] --&gt; P[Predictive]     DM --&gt; D[Descriptive]     P --&gt; C[Classification]     P --&gt; R[Regression]     P --&gt; T[Time Series Analysis]     P --&gt; Pd[Prediction]     D --&gt; Cl[Clustering]     D --&gt; Su[Summarization]     D --&gt; AR[Association Rules]   </pre>
6	<p><b>State and Explain the various classification of data mining systems with example? example? [Nov-Dec 2014] (8 M) BTL2 (13 M)</b></p> <p><b>Answer: Page: 3.11 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Classification according to the type of data source mined: (2 M)</li> <li>✓ Classification according to the data model drawn on: (2 M)</li> <li>✓ Classification according to the kind of knowledge discovered: (2 M)</li> <li>✓ Classification according to mining techniques used: (2 M)</li> </ul> <p><b>ii) describe the Five primitives for specifying a data mining task (5 M)</b></p> <p><b>Answer: Page : 3.41 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Task-relevant data (2 M)</li> <li>✓ Knowledge type to be mined (3 M)</li> <li>✓ Background knowledge (2 M)</li> <li>✓ Pattern interestingness measure (3 M)</li> <li>✓ Visualization of discovered patterns (3 M)</li> </ul>
7	<p><b>Explain various methods of data cleaning in detail. [May-Jun 2014] (13M) BTL2</b></p> <p><b>Answer: Page : 3.19 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Data Cleaning- clean inconsistent- noisy- incomplete data (2 M)</li> <li>✓ Missing values – fill in missing values(3 M)</li> <li>✓ Noisy Data –random error or variance in a measured values(3 M)</li> <li>✓ Inconsistent data – due to functional dependency and data entry(3 M)</li> <li>✓ Disadvantage of data cleaning process- ( 2 M)</li> </ul>
8	<p><b>Explain different strategies of data reduction [Nov-Dec 2016] BTL2 (13 M)</b></p> <p><b>Answer: Page : 3.41 - Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Data cube Aggregation – aggregation on data cube (3 M)</li> <li>✓ Attribute subset selection- reduces the dataset size by removing the irrelevant data (3 M)</li> <li>✓ Dimensionality reduction- compressed representation of original data (4 M)</li> <li>✓ Numerosity reduction- choosing smaller form of representation(3 M)</li> </ul>
9	<p><b>Describe data discretization and concept hierarchy generation. State which concept hierarchies are useful in data minng. [Nov-Dec 2016] BTL2 (13 M)</b></p> <p><b>Answer: Page : 3.38 – Poonkuzhalai</b></p> <ul style="list-style-type: none"> <li>✓ Introduction (1 M)</li> <li>✓ Data discretization – reduce the no of values for the given continuous attribute (3 M)</li> <li>✓ Concept hierarchy generation- for the numerical data describes the discretization of data (3 M)</li> <li>✓ Discretization and concept hierarchy generation for numerical data - generated automatically by a Data discretization (3 M)</li> <li>✓ Concept hierarchy generation for categorical data- these are discrete data (3 M)</li> </ul>

**PART C**

1	<p><b>Discuss whether or not each of the following activities is a data mining task. BTL6 (15 M)</b></p> <ul style="list-style-type: none"> <li>✓ Credit card fraud detection using transaction records</li> <li>✓ Dividing the customers of a company according to their gender</li> <li>✓ Computing the total sales of a company</li> <li>✓ Predicting the future stock price of a company using historical records</li> <li>✓ Monitoring seismic waves for earthquake activities</li> </ul> <p><b>Answer:</b></p> <p>Discuss each process in explain in terms of data mining</p> <ul style="list-style-type: none"> <li>✓ Credit card fraud detection using transaction records(3 M)</li> <li>✓ Dividing the customers of a company according to their gender(3 M)</li> <li>✓ Computing the total sales of a company(3 M)</li> <li>✓ Predicting the future stock price of a company using historical records(3 M)</li> <li>✓ Monitoring seismic waves for earthquake activities(3 M)</li> </ul>
2	<p><b>Discuss on descriptive and predictive data mining tasks with illustrations BTL2 (15 M)</b></p> <p><b>Answer: Page : 4.76, 4.80 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Explain descriptive Data mining (7 M)           <ul style="list-style-type: none"> <li>Estimating classifier accuracy</li> <li>Classifier accuracy measures</li> </ul> </li> <li>✓ predictive data mining (8 M)           <ul style="list-style-type: none"> <li>Linear Regression</li> <li>Non linear regression</li> <li>Other regression methods</li> </ul> </li> </ul>
3	<p><b>Suppose that the data for analysis include the attributed age. The age valuesfor the data tuples are 13,15,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70.</b></p> <ul style="list-style-type: none"> <li>✓ use smoothing by bin means to smooth the above data using a bin depth of 3. Illustrate your steps.</li> <li>✓ Classify the various methods for data smoothing. BTL3 (15 M)</li> </ul> <p><b>Answer: Page : 3.20 - Poonkuzhali</b></p> <p><b>Binning:</b> Binning methods smooth a sorted data value by consulting its “neighborhood,” that is, the values around it. The sorted values are distributed into a number of “buckets,” or <i>bins</i>. Because binning methods consult the neighborhood of values, they perform <i>local</i> smoothing. (5 M )</p> <p>Binning methods for data smoothing: (10 M)</p> <p>Sorted data for <i>price</i> (in dollars): 4, 8, 15, 21, 21, 24, 25, 28, 34</p> <p>Partition into (equal-frequency) bins:</p> <ul style="list-style-type: none"> <li>Bin 1: 4, 8, 15</li> <li>Bin 2: 21, 21, 24</li> <li>Bin 3: 25, 28, 34</li> </ul> <p>Smoothing by bin means:</p> <ul style="list-style-type: none"> <li>Bin 1: 9, 9, 9</li> <li>Bin 2: 22, 22, 22</li> <li>Bin 3: 29, 29, 29</li> </ul> <p>Smoothing by bin boundaries:</p> <ul style="list-style-type: none"> <li>Bin 1: 4, 4, 15</li> <li>Bin 2: 21, 21, 24</li> <li>Bin 3: 25, 25, 34</li> </ul>

4	<p><b>Distinguish between data generalization and characterizations. (ii). Sketch the various phases of data mining and explain the different steps involved in preprocessing with their significance before mining, Give an example for each process. BTL3 (15 M)</b></p> <p><b>Answer:</b> Page : 3.17- Poonkuzhalai</p> <p><b>Concept/Class Description: Characterization and Discrimination (10 marks)</b></p> <ul style="list-style-type: none"> <li>✓ data characterization, by summarizing the data of the class under study (often called the target class) in general terms, or</li> <li>✓ data discrimination, by comparison of the target class with one or a set of comparative classes (often called the contrasting classes), or</li> <li>✓ both data characterization and discrimination. Data characterization is a summarization of the general characteristics or features of a target class of data.</li> </ul> <p><b>Example 1 (5 M)</b></p> <p>Data characterization. A data mining system should be able to produce a description summarizing the characteristics of customers who spend more than \$1,000 a year at <i>AllElectronics</i>.</p> <p>Data discrimination - is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.</p>
5	<p><b>Using Equi-depth binning method, partition the data given below into 4 bins and perform smoothing according to the following methods.</b></p> <ol style="list-style-type: none"> <li>1. Smoothing by bin means</li> <li>2. Smoothing by bin median</li> <li>3. Smoothing by bin boundaries 24,25,26,27,28,56,67,70,70,75,78,89,89,90,91,94,95,96,100,102,103, 107,109,112.</li> </ol> <p><b>BTL3 (15 M)</b></p> <p><b>Answer:</b> Page : 3.20- Poonkuzhalai</p> <p>Binning: Binning methods smooth a sorted data value by consulting its “neighborhood,” that is, the values around it. The sorted values are distributed into a number of “buckets,” or <i>bins</i>. Because binning methods consult the neighborhood of values, they perform <i>local</i> smoothing.</p> <p>Binning methods for data smoothing:</p> <p>Sorted data for <i>price</i> (in dollars): 4, 8, 15, 21, 21, 24, 25, 28, 34</p> <p>Partition into (equal-frequency) bins:</p> <ul style="list-style-type: none"> <li>Bin 1: 4, 8, 15</li> <li>Bin 2: 21, 21, 24</li> <li>Bin 3: 25, 28, 34</li> </ul> <p>Smoothing by bin means:</p> <ul style="list-style-type: none"> <li>Bin 1: 9, 9, 9</li> <li>Bin 2: 22, 22, 22</li> <li>Bin 3: 29, 29, 29</li> </ul> <p>Smoothing by bin boundaries:</p> <ul style="list-style-type: none"> <li>Bin 1: 4, 4, 15</li> <li>Bin 2: 21, 21, 24</li> <li>Bin 3: 25, 25, 34</li> </ul>
6	<p><b>Design and discuss in detail about integration of data mining system with a data warehouse BTL3 (15 M)</b></p> <p><b>Answer:</b> Page : 1.9- Poonkuzhalai</p> <ul style="list-style-type: none"> <li>✓ Integrating data mining with data warehouse (5M)</li> <li>✓ Construction of data warehouse (10M)</li> </ul>

7	<p><b>Explain with diagrammatic illustration data mining as a confluence of multiple disciplines (15 M) BTL6</b></p> <p><b>Answer: Page : 3.01- Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ What is data mining- non trivial process of identifying novel data (3 M)</li> <li>✓ Need of data mining – to extract knowledge(3 M)</li> <li>✓ KDD vs Data mining – KDD is process of identifying potential useful information- data mining is a step in KDD(3 M)</li> <li>✓ Steps in KDD process- data selection- transformation – mining- evaluation - presentation (3 M)</li> </ul> <p>Diagram (3 M)</p>
8	<p><b>How is Mining single dimensional Boolean association rule from transaction databases is done?</b></p> <p><b>Explain BTL6</b></p> <p><b>Answer: Page :3.47- Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Mining multidimensional Boolean association rules form transactional database (4M)</li> <li>✓ Mining multilevel association rules from transactional database(4M)</li> <li>✓ Mining multidimensional association rules from transactional database an data warehouse(4M)</li> <li>✓ Constraint based association mining (3M)</li> </ul>

## UNIT-4 ASSOCIATION RULE MINING AND CLASSIFICATION

**Mining Frequent Patterns, Associations and Correlations – Mining Methods – Mining various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining – Classification and Prediction - Basic Concepts - Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction.**

### PART \* A

1.	<b>What is decision tree method? (Apr/May 2008) BTL1</b> A decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and leaf nodes represent classes or class distributions. The top most in a tree is the node.										
2.	<b>Distinguish between classification and clustering. (Apr/May2008) BTL4</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center; padding: 5px;"><b>classification</b></th><th style="text-align: center; padding: 5px;"><b>Clustering</b></th></tr> </thead> <tbody> <tr> <td style="padding: 5px;">Supervised learning</td><td style="padding: 5px;">Unsupervised learning</td></tr> <tr> <td style="padding: 5px;">Class label of each training sample is provided.</td><td style="padding: 5px;">Class label of each training sample is not known</td></tr> <tr> <td style="padding: 5px;">The set of classes are not known in advance.</td><td style="padding: 5px;">The number or set of classes to be learned advance.</td></tr> <tr> <td style="padding: 5px;">Learning by example.</td><td style="padding: 5px;">Learning by observation.</td></tr> </tbody> </table>	<b>classification</b>	<b>Clustering</b>	Supervised learning	Unsupervised learning	Class label of each training sample is provided.	Class label of each training sample is not known	The set of classes are not known in advance.	The number or set of classes to be learned advance.	Learning by example.	Learning by observation.
<b>classification</b>	<b>Clustering</b>										
Supervised learning	Unsupervised learning										
Class label of each training sample is provided.	Class label of each training sample is not known										
The set of classes are not known in advance.	The number or set of classes to be learned advance.										
Learning by example.	Learning by observation.										
3.	<b>List out the major strength of decision tree. (Apr/May2008) BTL2</b> <ul style="list-style-type: none"> <li>Construction of decision tree classifiers does not require any domain knowledge or parameter setting and therefore is appropriate for exploratory knowledge discovery.</li> <li>Handle high dimensional data.</li> <li>Simple and fast.</li> <li>Used for many applications such as medicine, manufacturing, financial analysis, astronomy, etc.</li> <li>Basis of several commercial role induction systems.</li> </ul>										
4.	<b>How do you choose the best split while constructing a decision tree.? (May/June2014) BTL3</b> By using <b>attribute selection measures</b> , we choose the best split while constructing a decision tree. The measures are <ul style="list-style-type: none"> <li>Information gain</li> <li>Gain ratio</li> <li>Gini index</li> </ul>										
5.	<b>What are the tree pruning methods? (Nov/Dec2011)BTL1</b> Tree pruning use statistical measures to remove the least reliable branches. Pruned trees tend to be smaller and less complex and easier to comprehend. The tree pruning										

	<p>methods are</p> <ul style="list-style-type: none"> <li>➤ Prepruning</li> <li>➤ Post pruning</li> <li>➤ Pessimistic pruning</li> </ul>
6.	<p><b>What is correlation analysis? Justify with an example (Nov/Dec 2011 ) (NOV/DEC 2018) BTL3</b></p> <p>A correlation measure can be used to augment the support and confidence but also for association rules. This leads to correlation rules of the form  <math>A \Rightarrow B</math> [support, confidence, correlation]</p>
7.	<p><b>Give examples for binary and multi-dimensional association rules.(Nov/Dec2013)</b></p> <p>BTL2</p> <p>Binary or Single dimensional association rule  <math>\text{buys}(X = \text{"computer"}) \Rightarrow \text{buys}(X = \text{"HPPrinter"})</math></p> <p>Multidimensional association rules  <math>\text{Age}(X, \text{"30..39"}) \wedge \text{income}(X, \text{"2000..40000"}) \Rightarrow \text{buys}(X, \text{"LCD TV"})</math></p>
8.	<p><b>List the 2 interesting measures of an association rule . (Nov/Dec 2012) BTL2</b></p> <p>There are 2 interesting measures od an association rule. They are</p> <ul style="list-style-type: none"> <li>✓ Support (<math>A \Rightarrow B</math>)=<math>P(A \cup B)</math></li> <li>✓ Confidence (<math>A \Rightarrow B</math>)=<math>P(B / A)</math></li> </ul>
9.	<p><b>What is a support vector machine? (Apr/May2011) BTL1</b></p> <p>Decision tree induction algorithms function recursively. First, an attribute must be selected as the root node. In order to create the most efficient (smallest) tree, the root node must effectively split the data. Each split attempts to pare down a set of instances (the actual data) until they all have the same classification. The split is the one that provides what is termed the most information gain.</p>
10.	<p><b>State the need for pruning phase in decision tree construction? (May/June2011)</b></p> <p>BTL2</p> <p>When a decision tree is built many of the branches will reflect anomalies in the training data to noise or outliers. Tree pruning methods address this problem of outfitting the data. Such methods typically use statistical measures to remove the least reliable branches</p>
11.	<p><b>Define frequent patterns. BTL1</b></p> <p>Frequent patterns are patterns (such as item sets, subsequences or substructures) that appear in a dataset frequently.</p>
12.	<p><b>Define Market basket analysis. BTL1</b></p> <p>A typical example of frequent item set mining is market basket analysis. This process analyzes customer buying habits by fining associations between the different items that customers place in their shopping baskets.</p>
13.	<p><b>What is STRONG? BTL1</b></p> <p>The rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong.</p>
14.	<p><b>Define Apriori property. BTL1</b></p> <p>The Apriori property states that all nonempty subsets of a frequent item set must also be frequent.</p>
15.	<p><b>What is Anti-monotone? BTL1</b></p> <p>If a set cannot pass a test, all of its subsets will fail the same test as well. This scenario is called as anti- monotone. Only if the set passes test only then all its subset will pass the test.</p>

16.	<b>Define Association rule.</b> BTL1 The Association rule works in such a way that, it searches for interesting relationships among items in a given data set.						
17.	<b>What are the constraints can include the constraint-based association mining?</b> BTL1 ➤ Knowledge type constraints ➤ Data constraints ➤ Dimension/level constraints ➤ Interestingness constraints ➤ Rule constraints						
18.	<b>What are the steps involved in data classification?</b> BTL1 Data classification is a 2 step process. They are ✓ Learning step ✓ Classification step.						
19.	<b>What are the preprocessing steps can be used in classification or prediction process?</b> BTL1 ❖ Data cleaning ❖ Relevance analysis ❖ Data transformation and reduction						
20.	<b>Define ID3?</b> BTL1 It is a decision tree algorithm used by decision tree induction.						
21.	<b>List some popular attribute selection measures.</b> BTL1 Some of the popular attribute selection measures are ➤ Information gain ➤ Gain ratio ➤ Gini index						
22.	<b>What are the classification tools for data mining? (May/June 2013)</b> BTL1 The classification tools for data mining are • Rapid Miner • WEKA						
23.	<b>Define predictive model. (May/June 2013)</b> BTL1 Predictive modeling is a process used in predictive analytics to create a statistical model of future behavior. Predictive analytics is the area of data mining concerned with forecasting probabilities and trends.						
24.	<b>How is prediction different from classification? (May/June 2014)</b> BTL2 <table border="1"> <thead> <tr> <th>Classification</th> <th>Prediction</th> </tr> </thead> <tbody> <tr> <td>The class label attribute is discrete-valued and unordered</td> <td>The class label attribute is continuous-valued and ordered</td> </tr> <tr> <td>The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier</td> <td>The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of <math>y</math> for each of the test tuples</td> </tr> </tbody> </table>	Classification	Prediction	The class label attribute is discrete-valued and unordered	The class label attribute is continuous-valued and ordered	The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier	The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of $y$ for each of the test tuples
Classification	Prediction						
The class label attribute is discrete-valued and unordered	The class label attribute is continuous-valued and ordered						
The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier	The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of $y$ for each of the test tuples						
25.	<b>How do you choose best split while constructing decision tree? (May/June 2014)</b> BTL3 The best split for decision tree is done by selecting the attribute. In decision tree induction algorithm calls <i>Attribute selection method</i> to determine the splitting criterion. The splitting criterion tells us which attribute to test at node $N$ by determining the “best” way to						

	separate or partition the tuples in $D$ into individual classes
26.	<p><b>List the phases involved in decision tree induction. (Nov/Dec 2016) BTL4</b></p> <p>The phases involved in decision tree induction are</p> <ul style="list-style-type: none"> <li>• Attribute – Node selection</li> <li>• Tree Pruning</li> </ul>
27.	<p><b>Methods to improve apriori efficiency. (Nov/Dec 2016) BTL4</b></p> <p>To improve the efficiency of Apriori algorithm the following techniques can be adapted:</p> <ul style="list-style-type: none"> <li>• Reducing the number of transactions</li> <li>• Partitioning the data to find candidate item sets</li> <li>• Sampling – mining on a subset of the given data</li> <li>• Dynamic itemset counting</li> </ul>
28	<p><b>Define Bayesian Belief networks. BTL1</b></p> <ul style="list-style-type: none"> <li>➢ It specifies joint conditional probability distributions they allow class conditional independencies to be defined between subsets of variables.</li> <li>➢ A belief network is defined by 2components:</li> </ul> <ul style="list-style-type: none"> <li>• Directed acyclic graph</li> <li>• Set of conditional probability tables.</li> </ul>
<b>PART * B</b>	
1	<p><b>Explain frequent item sets mining method using Apriori algorithm problem (13M) (May/June 2012) BTL3</b></p> <p><b>Answer: Page : 95 -106-Poonkuzhali</b></p> <p>Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining Frequent item sets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties.</p> <p>Join step – to find <math>L_k</math>, a set of candidate <math>k</math>- item sets is generate by joining <math>L_{k-1}</math>, with itself. This set of candidates is denoted <math>C_k</math> (7M)</p> <p>Prune step - <math>C_k</math> is a superset of <math>L_k</math>, that is, its members may or may not be frequent, but all of the frequent <math>k</math>- items set are included in <math>C_k</math> (6M)</p>
2	<p><b>Describe naïve Bayesian classifier.(May/June 2012) (NOV/DEC 2018) (13M) BTL1</b></p> <p><b>Answer: Page: 119- Poonkuzhali</b></p> <p>Bayes' Theorem (5M)</p> <p>Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Bayesian classification is based on Bayes' theorem. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called <i>class conditional independence</i>.</p> <p>Bayes' theorem is useful in that it provides a way of calculating the posterior probability, <math>P(H X)</math>, from <math>P(H)</math>, <math>P(X H)</math>, and <math>P(X)</math>.</p> <p>Naïve Bayesian Classification (5M)</p> <p>EXAMPLE table (3M)</p>
3	<p><b>Explain about the Classification by decision tree. (13M) (May/June 2013) BTL1</b></p> <p><b>Answer: Page:117- Poonkuzhali</b></p>

	<p>Decision tree induction- used to classify and decide to what to do next (5M) Algorithm (5M)</p> <pre> graph TD     A[age?] -- youth --&gt; B[student?]     A -- middle aged --&gt; C[credit_rating?]     A -- senior --&gt; D[ ]     B -- no --&gt; E(( ))     B -- yes --&gt; F(( ))     C -- fair --&gt; G(( ))     C -- excellent --&gt; H(( ))     </pre> <p>A decision tree for the concept <code>buys_computer</code>, indicating whether a customer at ALL Electronics is likely to purchase a computer. Example: (3M)</p>						
4	<p><b>Write and explain the algorithms for mining frequent item sets without candidate generation (13 M) (May/June 2014) BTL1</b></p> <p><b>Answer: Page: 95 -106- Poonkuzhali</b>  Join step – to find <math>L_k</math>, a set of candidate <math>k</math>- item sets is generate by joining <math>L_{k-1}</math>, with itself. This set of candidates is denoted <math>C_k</math> (7 M)  Prune step - <math>C_k</math> is a superset of <math>L_k</math>, that is, its members may or may not be frequent, but all of the frequent <math>k</math>- items set are included in <math>C_k</math>. (6 M)</p>						
5	<p><b>Compare : Classification and Prediction. (13 M) BTL2</b></p> <p><b>Answer: Page: 116-Poonkuzhali</b>  Classification (6M)  Prediction (7M)</p> <p>Databases are rich with hidden information that can be used for intelligent decision making. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large.</p> <p>Whereas <i>classification</i> predicts categorical (discrete, unordered) labels, <i>prediction</i> models continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.</p> <p><b>CLASSIFICATION:</b></p> <table border="1"> <thead> <tr> <th>Classification</th> <th>Prediction</th> </tr> </thead> <tbody> <tr> <td>The class label attribute is discrete-valued and unordered</td> <td>The class label attribute is continuous-valued and ordered</td> </tr> <tr> <td>The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier</td> <td>The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of <math>y</math> for each of the test tuples</td> </tr> </tbody> </table>	Classification	Prediction	The class label attribute is discrete-valued and unordered	The class label attribute is continuous-valued and ordered	The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier	The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of $y$ for each of the test tuples
Classification	Prediction						
The class label attribute is discrete-valued and unordered	The class label attribute is continuous-valued and ordered						
The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier	The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of $y$ for each of the test tuples						

6	<p><b>Describe Constraint based association mining. (13M) (May/June 2015) BTL1</b></p> <p><b>Answer:Page: 174 -Notes</b></p> <ul style="list-style-type: none"> <li>✓ Knowledge type constraints- specify the type of knowledge(3M)</li> <li>✓ Data constraints- set of task relevant data(3M)</li> <li>✓ Dimension/level constraints- desired dimensions of data(3M)</li> <li>✓ Interestingness constraints- threshold of measures(2M)</li> <li>✓ Rule constraints- metarule (2M)</li> </ul>
7	<p><b>Describe Data Discretization and Concept Hierarchy Generation. State why Concept Hierarchies are useful in data mining. (13 M) (Nov/Dec 2016) BTL2</b></p> <p><b>Answer:Page: 3.38-Poonkuzhali</b></p> <p>Concept hierarchies ( 2M)</p> <p>Constraint based association mining – association rule based on constraint ( 5M)</p> <p>How to generate the hierarchy- binning- histogram analysis- entropy based processing(4M)</p> <p>Rules- any rule for example ( 2M)</p>

**PART \* C**

1	<p><b>Apply the Apriori algorithm for discovering frequent item sets to the following data: Use 0.3 for the minimum support value. (15M) (May/June 2015) BTL6</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">TID</th><th style="text-align: left; padding: 2px;">Items Purchased</th></tr> </thead> <tbody> <tr> <td style="padding: 2px;">101</td><td style="padding: 2px;">Litchi, Hill Banana, Straw berry</td></tr> <tr> <td style="padding: 2px;">102</td><td style="padding: 2px;">Litchi, Passion fruit</td></tr> <tr> <td style="padding: 2px;">103</td><td style="padding: 2px;">Passion fruit, Tomato</td></tr> <tr> <td style="padding: 2px;">104</td><td style="padding: 2px;">Litchi, Hill Banana, Straw berry</td></tr> <tr> <td style="padding: 2px;">105</td><td style="padding: 2px;">Pears, Straw berry</td></tr> <tr> <td style="padding: 2px;">106</td><td style="padding: 2px;">Pears</td></tr> <tr> <td style="padding: 2px;">107</td><td style="padding: 2px;">Pears, Passion fruit</td></tr> <tr> <td style="padding: 2px;">108</td><td style="padding: 2px;">Litchi, Hill Banana, Water Melon, Straw berry</td></tr> <tr> <td style="padding: 2px;">109</td><td style="padding: 2px;">Water Melon, Tomato</td></tr> <tr> <td style="padding: 2px;">110</td><td style="padding: 2px;">Litchi, Hill Banana</td></tr> </tbody> </table> <p><b>Answer:Page:99 – 106-Poonkuzhali</b></p> <p>Apriori property – (1M)</p> <p>Apriori algorithm – (2M)</p> <p>Generating frequent item sets - (5M)</p> <p>Generating candidate sets- (5M)</p> <p>Apply the Apriori algorithm for the above table (2M)</p>	TID	Items Purchased	101	Litchi, Hill Banana, Straw berry	102	Litchi, Passion fruit	103	Passion fruit, Tomato	104	Litchi, Hill Banana, Straw berry	105	Pears, Straw berry	106	Pears	107	Pears, Passion fruit	108	Litchi, Hill Banana, Water Melon, Straw berry	109	Water Melon, Tomato	110	Litchi, Hill Banana
TID	Items Purchased																						
101	Litchi, Hill Banana, Straw berry																						
102	Litchi, Passion fruit																						
103	Passion fruit, Tomato																						
104	Litchi, Hill Banana, Straw berry																						
105	Pears, Straw berry																						
106	Pears																						
107	Pears, Passion fruit																						
108	Litchi, Hill Banana, Water Melon, Straw berry																						
109	Water Melon, Tomato																						
110	Litchi, Hill Banana																						
2	<p><b>Find all frequent item sets for the given training sets using Apriori And FP-growth, respectively. Compare the efficiency of the two mining processes. (15M) (Nov/Dec 2016) BTL6</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">TID</th><th style="text-align: left; padding: 2px;">items _bought</th></tr> </thead> </table>	TID	items _bought																				
TID	items _bought																						

	<p>T100 (M,O,N,K,E,Y)  T200 (D,O,N,K,E,Y)  T300 (M,A,K,E)  T400 (M,U,C,K,E,Y)  T500 (C,O,O,K,I,E)</p> <p><b>Answer: Page:99 – 106-Poonkuzhalai</b>  Apriori property – (1M)  Apriori algorithm - (2M)  Generating frequent item sets - (5M)  Generating candidate sets- (5M)  Apply the Apriori algorithm for the above table (2M)</p>																				
3	<p><b>A database has nine transactions let min-sup = 30%. (15M) (May/June 2014) BTL6</b></p> <table> <thead> <tr> <th>TID</th> <th>list of items-IDs</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>a, b, e</td> </tr> <tr> <td>2</td> <td>b, d</td> </tr> <tr> <td>3</td> <td>b, c</td> </tr> <tr> <td>4</td> <td>a, b, d</td> </tr> <tr> <td>5</td> <td>a, c</td> </tr> <tr> <td>6</td> <td>b, c</td> </tr> <tr> <td>7</td> <td>a, c</td> </tr> <tr> <td>8</td> <td>a, b, c, e</td> </tr> <tr> <td>9</td> <td>a, b, c</td> </tr> </tbody> </table> <p><b>Answer: Page:99 – 106-Poonkuzhalai</b>  Apriori property – (1M)  Apriori algorithm – (2M)  Generating frequent item sets - (5M)  Generating candidate sets-(5M)  Apply the Apriori algorithm for the above table (2M)</p>	TID	list of items-IDs	1	a, b, e	2	b, d	3	b, c	4	a, b, d	5	a, c	6	b, c	7	a, c	8	a, b, c, e	9	a, b, c
TID	list of items-IDs																				
1	a, b, e																				
2	b, d																				
3	b, c																				
4	a, b, d																				
5	a, c																				
6	b, c																				
7	a, c																				
8	a, b, c, e																				
9	a, b, c																				
4	<p><b>Consider a home finance loan to predict the housing loan payment. Design a general hierarchical a structure and analyze the factors using rule discovery techniques to accurately predict the number of loan payments in a given quarter/year. Loan is availed for a period of 20 to 25 years, but an average life span of the loan exists for only 7 to 10 years due topayment.</b></p> <p><b>Make necessary assumptions:</b> Maintenance record of the customer details and details of the prevailing interest rates, borrower characteristics, account dare, fine tune loan prepayment such as interest rates and fees in order to maximize the profits of the company. Elaborately discuss the association rule mining issues. Also Examine on the multi level association rules and find if you could relate any relation on from the aboveapplication. (15M) BTL6</p> <p><b>Answer:Page: 4.4-Gunasundari, Page:107 – 110 in notes</b>  Apriori property – (1M)  Apriori algorithm – (2M)  Generating frequent item sets - (5M)  Generating candidate sets-(5M)  Apply the Apriori algorithm for the above table (2M)</p>																				
5	<p><b>Explain and Apply the Apriori algorithm for discovering frequent item sets of the table. (15M) BTL6</b></p> <table border="1"> <thead> <tr> <th>Trans ID</th> <th>Items Purchased</th> </tr> </thead> <tbody> <tr> <td>101</td> <td>Milk,bread,eggs</td> </tr> </tbody> </table>	Trans ID	Items Purchased	101	Milk,bread,eggs																
Trans ID	Items Purchased																				
101	Milk,bread,eggs																				

102	Milk,juice
103	Juice,butter
104	Milk,bread,eggs
105	Coffee,eggs
106	Coffee
107	Coffee,Juice
108	Milk,bread,cookies,eggs
109	Cookies,butter
110	Milk,bread

Use 0.3 for the minimum support value. Illustrate each step of the Apriori Algorithm.

**Answer: Page: 4.4 –Gunasundari. Page:107 – 110 in notes**

Apriori property – (1M)

Apriori algorithm – (2M)

Generating frequent item sets - (5M)

Generating candidate sets-(5M)

Apply the Apriori algorithm for the above table (2M)

--	--

### UNIT 5- CLUSTERING AND TRENDS IN DATA MINING

**Cluster Analysis - Types of Data – Categorization of Major Clustering Methods – K-means–Partitioning Methods – Hierarchical Methods - Density-Based Methods –Grid Based Methods –Model-Based Clustering Methods – Clustering High Dimensional Data - Constraint – Based ClusterAnalysis – Outlier Analysis – Data Mining Applications.**

#### PART \* A

1.	<b>Why do you go for clustering analysis? (Nov/Dec 2011) BTL1</b> Clustering can be used to generate a concept hierarchy for A by following either a top down splitting strategy or a bottom-up merging strategy, where each cluster forms a node of the concept hierarchy. In the former, each initial cluster or partition may be further decomposed into several sub clusters, forming a lower level of the hierarchy. In the latter, clusters are formed by repeatedly grouping neighboring clusters in order to form higher-level concepts.
2.	<b>What are the requirements of cluster analysis? (Nov/Dec 2010) BTL1</b> <ul style="list-style-type: none"> <li>✓ Scalability</li> <li>✓ Ability to deal with different types of attributes</li> <li>✓ Discovery of clusters with arbitrary shape</li> <li>✓ Minimal requirements for domain knowledge to determine input parameters</li> <li>✓ Ability to deal with noisy data</li> <li>✓ Incremental clustering and insensitivity to the order of input records</li> <li>✓ High dimensionality</li> <li>✓ Constraint-based clustering</li> <li>✓ Interpretability and usability</li> </ul>
3.	<b>What is meant by cluster analysis? (April/May 2008) BTL1</b> A cluster analysis is the process of analyzing the various clusters to organize the different objects into meaningful and descriptive object.
4.	<b>Define: Outlier Analysis. ( Nov/Dec 2014) (NOV/DEC 2018) BTL1</b> A database may contain data objects that do not comply with the general behavior or model of the data. These data objects are outliers. The analysis of outliers data is referred to as outlier analysis.
5.	<b>Define CLARANS. BTL1</b> <ul style="list-style-type: none"> <li>• CLARANS (Cluster Large Applications based on Randomized Search) to improve the quality of CLARA we go for CLARANS.</li> <li>• It Draws sample with some randomness in each step of search.</li> <li>• It overcomes the problem of scalability that K-Medoids suffers from.</li> </ul>
6.	<b>Define BIRCH, ROCK and CURE. BTL1</b> BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies): Partitions objects hierarchically using tree structures and then refines the clusters using other clustering methods. It defines a clustering feature and an associated tree structure that summarizes a cluster. The tree is a height balanced tree that stores cluster information. BIRCH doesn't produce spherical Cluster and may produce unintended cluster. ROCK (RObust Clustering using links): Merges clusters based on their interconnectivity. Great for categorical data. Ignores information about the looseness of two clusters while emphasizing interconnectivity. CURE (Clustering Using Representatives): Creates clusters by sampling the database and shrinks them toward the center of the cluster by a specified fraction.

7.	<p><b>What is meant by web usage mining? (Nov/Dec 2007) (May/June 2010) BTL1</b></p> <p>Web usage mining is the process of extracting useful information from server logs i.e. users history. Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data.</p>
8.	<p><b>What is meant by audio data mining? (Nov/Dec 2007) BTL1</b></p> <p>Audio data mining uses audio signals to indicate the patterns of data or the features of data mining results. Although visual data mining may disclose interesting patterns using graphical displays, it requires users to concentrate on watching patterns and identifying interesting or novel features within them. This can sometimes be quite tiresome. If patterns can be transformed into sound and music, then instead of watching pictures, we can listen to pitches, rhythms, tune, and melody in order to identify anything interesting or unusual. This may relieve some of the burden of visual concentration and be more relaxing than visual mining. Therefore, audio data mining is an interesting complement to visual mining.</p>
9.	<p><b>Define visual data mining. (April/May 2008) BTL1</b></p> <p>Visual data mining discovers implicit and useful knowledge from large data sets using data and/or knowledge visualization techniques. The human visual system is controlled by the eyes and brain, the latter of which can be thought of as a powerful, highly parallel processing and reasoning engine containing a large knowledge base. Visual data mining essentially combines the power of these components, making it a highly attractive and effective tool for the comprehension of data distributions, patterns, clusters, and outliers in data.</p>
10.	<p><b>What is meant by the frequency item set property? (Nov/Dec 2008) BTL1</b></p> <p>A set of items is referred to as an item set. An item set that contains k items is a k-item set. The set {computer, antivirus software} is a 2-itemset. The occurrence frequency of an item set is the number of transactions that contain the item set. This is also known, simply, as the frequency, support count, or count of the item set.</p>
11.	<p><b>Mention the advantages of hierarchical clustering. (Nov/Dec 2008) BTL2</b></p> <p><i>Hierarchical clustering</i> (or <i>hierarchic clustering</i>) outputs a hierarchy, a structure that is more informative than the unstructured set of clusters returned by flat clustering. Hierarchical clustering does not require us to <b>prespecify</b> the number of clusters and most hierarchical algorithms that have been used in IR are deterministic. These advantages of hierarchical clustering come at the cost of lower efficiency.</p>
12.	<p><b>Define time series analysis. (May/June 2009) BTL1</b></p> <p>Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Time series are very frequently plotted via line charts.</p>
13.	<p><b>What is meant by web content mining? (May/June 2009) BTL1</b></p> <p>Web content mining, also known as text mining, is generally the second step in Web data mining. Content mining is the scanning and mining of text, pictures and graphs of a Web page to determine the relevance of the content to the search query. This scanning is completed after the clustering of web pages through structure mining and provides the results based upon the level of relevance to the suggested query. With the massive amount of information that is available on the World Wide Web, content mining provides the results lists to search engines in order of highest relevance to the keywords in the query.</p>
14.	<p><b>Write down some applications of data mining.(Nov/Dec 2009)BTL5</b></p> <p>✓ Financial Data Analysis</p>

	<ul style="list-style-type: none"> <li>✓ Retail Industry</li> <li>✓ Telecommunication Industry</li> <li>✓ Biological Data Analysis</li> <li>✓ Scientific Applications</li> <li>✓ Intrusion Detection</li> </ul>
15.	<p><b>List out the methods for information retrieval. (May/June 2010) BTL1</b></p> <p>They generally either view the retrieval problem as a document selection problem or as a document ranking problem. In document selection methods, the query is regarded as specifying constraints for selecting relevant documents. A typical method of this category is the Boolean retrieval model, in which a document is represented by a set of keywords and a user provides a Boolean expression of keywords, such as “car and repair shops,” “tea or coffee”.</p> <p>Document ranking methods use the query to rank all documents in the order of relevance. For ordinary users and exploratory queries, these methods are more appropriate than document selection methods.</p>
16.	<p><b>What is the categorical variable? (Nov/Dec 2010) BTL1</b></p> <p>A categorical variable is a generalization of the binary variable in that it can take on more than two states. For example, <i>map color</i> is a categorical variable that may have, say, five states: <i>red, yellow, green, pink, and blue</i>. Let the number of states of a categorical variable be <math>M</math>. The states can be denoted by letters, symbols, or a set of integers, such as <math>1, 2, \dots, M</math>. Notice that such integers are used just for data handling and do not represent any specific ordering</p>
17.	<p><b>What is the difference between row scalability and column scalability? (Nov/Dec 2010) BTL2</b></p> <p>Data mining has two kinds of scalability issues: row (or database size) scalability and column (or dimension) scalability.</p> <p>A data mining system is considered row scalable if, when the number of rows is enlarged 10 times, it takes no more than 10 times to execute the same data mining queries. A data mining system is considered column scalable if the mining query execution time increases linearly with the number of columns (or attributes or dimensions). Due to the curse of dimensionality, it is much more challenging to make a system column scalable than row scalable.</p>
18.	<p><b>What are the major challenges faced in bringing data mining research to market? (Nov/Dec 2010) BTL5</b></p> <p>The diversity of data, data mining tasks, and data mining approaches poses many challenging research issues in data mining. The development of efficient and effective data mining methods and systems, the construction of interactive and integrated data mining environments, the design of data mining languages, and the application of data mining techniques to solve large application problems are important tasks for data mining researchers and data mining system and application developers.</p>
19.	<p><b>What is meant by multimedia database? (Nov/Dec 2011) BTL1</b></p> <p>A multimedia database system stores and manages a large collection of <b>multimedia data</b>, such as audio, video, image, graphics, speech, text, document, and hypertext data, which contain text, text markups, and linkages. Multimedia database systems are increasingly common owing to the popular use of audio, video equipment, digital cameras, CD-ROMs, and the Internet.</p>
20.	<p><b>Define DB miner. (Nov/Dec 2011) BTL1</b></p> <p>DB Miner delivers business intelligence and performance management applications powered by data mining. With new and insightful business patterns and knowledge revealed by DB Miner. DB Miner Insight solutions are world's first server applications providing powerful and highly scalable association, sequence and differential mining capabilities for Microsoft SQL Server Analysis Services platform,</p>

	and they also provide market basket, sequence discovery and profit optimization for Microsoft Accelerator for Business Intelligence.
21.	<p><b>Define: Dendrogram.</b> BTL1</p> <ul style="list-style-type: none"> <li>✓ A tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering.</li> <li>✓ Decompose data objects into a several levels of nested partitioning (tree of clusters) called a dendrogram.</li> </ul>
22.	<p><b>Define cluster analysis (May/June 2013)</b> BTL1</p> <ul style="list-style-type: none"> <li>✓ The process of grouping a set of physical or abstract objects into classes of similar objects is called <b>clustering</b>. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters.</li> <li>✓ Cluster analysis has been widely used in numerous applications, including market research, pattern recognition, data analysis, and image processing.</li> </ul>
23.	<p><b>What are the different types of data used for cluster analysis? (May/June 2013)</b></p> <p>BTL1</p> <ul style="list-style-type: none"> <li>✓ Data matrix</li> <li>✓ Disimilarity matrix</li> </ul>
24.	<p><b>What is a STING? (May/June 2014)</b> BTL1</p> <p>STING : A Statistical Information Grid Approach to Spatial Data Mining</p> <p>It is a information grid approach that is statically given.</p>
25.	<p><b>Define WAVE Cluster. (May/June 2014)</b> BTL1</p> <p>Wave Cluster is a multi-resolution clustering algorithm that first summarizes the data by imposing a multidimensional grid structure onto the data space. It then uses a wavelet transformation to transform the original feature space, finding dense regions in the transformed space.</p>
26.	<p><b>Define cluster analysis. (May/June 2015)</b> BTL1</p> <ul style="list-style-type: none"> <li>✓ The process of grouping a set of physical or abstract objects into classes of similar objects is called <b>clustering</b>. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters.</li> <li>✓ Cluster analysis has been widely used in numerous applications, including market research, pattern recognition, data analysis, and image processing.</li> </ul>
27.	<p><b>Role of cluster analysis. (Nov/Dec 2016)</b> BTL1</p> <ul style="list-style-type: none"> <li>✓ Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.</li> <li>✓ Clustering can also help marketers discover distinct groups in their customer base. And they can characterize their customer groups based on the purchasing patterns.</li> </ul>
28.	<p><b>Write the Reason for clustering need in data mining. (Nov/Dec 2016)</b> BTL2</p> <p>Clustering may also help in the identification of areas of similar land use in an earth observation database and in the identification of groups of houses in a city according to house type, value, and geographic location, as well as the identification of groups of automobile insurance policy holders with a high average claim cost.</p>
	<b>PART * B</b>
1	<p><b>Explain about agglomerative clustering(May/June2012) (13M)</b> BTL1</p> <p><b>Answer: Page: 5.30- Poonkuzhalai</b></p> <p>Agglomerative hierarchical clustering:(5M)</p> <p>This bottom-up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all of the objects are in a single cluster or until certain termination conditions are satisfied. Most hierarchical clustering methods belong to this category. They differ only in their definition of inter cluster</p>

	similarity. Steps(4M) Diagram (4M)
2	<b>Explain divisive clustering (May/June 2012) (13M) BTL1</b> <b>Answer: Page: 5.30- Poonkuzhali</b> This top-down strategy does the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster. It subdivides the cluster into smaller and smaller pieces, until each object forms a cluster on its own or until it satisfies certain termination conditions, such as a desired number of clusters is obtained or the diameter of each cluster is within a certain threshold (5M) Steps(4M) Diagram (4M)
3	<b>Describe about Cluster analysis in detail (May/June 2013) (13M) BTL1</b> <b>Answer: Page: 5.1- Poonkuzhali</b> ✓ Partitioning methods:(2M) ✓ Hierarchical methods(2M) ✓ Density-based methods: (2M) ✓ Grid-based methods (2M) ✓ Model-based methods: (2M) ✓ Clustering high-dimensional data(1M) ✓ Constraint-based clustering (2M)
4	<b>Explain Partitioning methods – K-Means and K-Medoids Algorithm. (May/June 2013) (13 M) BTL1</b> <b>Answer: Page: 5.13- Poonkuzhali</b> <b>PARTITIONING METHODS</b> ✓ Centroid-Based Technique: The k-Means Method (7M)  The k-means algorithm proceeds as follows  ✓ First, it randomly selects k of the objects, each of which initially represents a cluster mean or center. ✓ For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. ✓ It then computes the new mean for each cluster. ✓ This process iterates until the criterion function converges. Typically, the square-error criterion is used, defined as E where E is the sum of the square error for all objects in the data set; p is the point in space representing a given object; and mi is the mean of cluster Ci (both p and mi are multidimensional). In other words, for each object in each cluster, the distance from the object to its cluster center is squared, and the distances are summed. This criterion tries to make the resulting k clusters as compact and as separate as possible. ✓ Representative Object-Based Technique: -The k-Medoids Method (6M) ✓ PAM (Partitioning Around Medoids) Partitioning Methods in Large Databases: From k-Medoids to CLARANS
5	<b>Explain the different types of data used in cluster analysis. (May/June 2014) (13M)</b> BTL1 <b>Answer: Page: 5.3- Poonkuzhali</b> ✓ Data matrix (or object-by-variable structure): (2M) ✓ Dissimilarity matrix (or object-by-object structure): (2M) ✓ Interval-scaled variables(2M) ✓ Binary Variables (2M) ✓ Categorical Variables (2M)

	<ul style="list-style-type: none"> <li>✓ Ordinal Variables(2M)</li> <li>✓ Ratio scaled variable (1M)</li> </ul>
6	<p><b>How would you discuss the outlier analysis? Discuss in detail about the various detection techniques in outlier. (13 M) (May/June 2014) (13M) BTL3</b></p> <p><b>Answer: Page: 5.46- Poonkuzhali</b></p> <p>Explanation (1M) The data objects that do not comply with the general behavior or model of the data and which are grossly different from or inconsistent with the remaining set of data, are called outliers. The outliers may be of particular interest, such as in the case of fraud detection, where outliers may indicate fraudulent activity. Thus, outlier detection and analysis is an interesting data mining task, referred to as outlier mining. Statistical Distribution based outliers detection (3M) Distance based outliers detection(3M) Density based local outliers detection (3M) Deviation based outliers detection (3M)</p>
7	<p><b>Write the difference between CLARA and CLARANS. (May/June 2014) (13M) BTL3</b></p> <p><b>Answer: Page: 5.28- Poonkuzhali</b></p> <p>Clara (6M) Clarans ( 7M)</p> <p>The idea behind CLARA is as follows:</p> <ul style="list-style-type: none"> <li>✓ Instead of taking the whole set of data into consideration, a small portion of the actual data is chosen as a representative of the data.</li> <li>✓ Medoids are then chosen from this sample using PAM.</li> <li>✓ If the sample is selected in a fairly random manner, it should closely represent the original data set.</li> <li>✓ The representative objects (medoids) chosen will likely be similar to those that would have been chosen from the whole data set.</li> <li>✓ CLARA draws multiple samples of the data set, applies PAM on each sample, and returns its best clustering as the output.</li> <li>✓ As expected, CLARA can deal with larger data sets than PAM. The complexity of each iteration now becomes <math>O(ks^2 + k(n-k))</math>, where s is the size of the sample, k is the number of clusters, and n is the total number of objects.</li> <li>✓ The effectiveness of CLARA depends on the sample size. Notice that PAM searches for the best k medoids among a given data set, whereas CLARA searches for the best k medoids among the selected sample of the data set.</li> <li>✓ CLARA cannot find the best clustering if any of the best sampled medoids is not among the best k medoids.</li> </ul> <p>A k-medoids type algorithm called CLARANS (Clustering Large Applications based upon RANdomized Search) was proposed, which combines the sampling technique with PAM.</p> <p>However, unlike CLARA, CLARANS does not confine itself to any sample at any given time. While CLARA has a fixed sample at each stage of the search, CLARANS draws a sample with some randomness in each step of the search. Conceptually, the clustering process can be viewed as a search through a graph, where each node is a potential solution (a set of k medoids).</p> <ul style="list-style-type: none"> <li>✓ CLARANS dynamically draws a random sample of neighbors in each step of a search. The number of neighbors to be randomly sampled is restricted by a user-specified parameter.</li> <li>✓ In this way, CLARANS does not confine the search to a localized area. If a better neighbor is found (i.e., having a lower error), CLARANS moves to the neighbor's node</li> </ul>

	<p>and the process starts again; otherwise, the current clustering produces a local minimum.</p> <ul style="list-style-type: none"> <li>✓ If a local minimum is found, CLARANS starts with new randomly selected nodes in search for a new local minimum. Once a user-specified number of local minima has been found, the algorithm outputs, as a solution, the best local minimum, that is, the local minimum having the lowest cost.</li> <li>✓ CLARANS has been experimentally shown to be more effective than both PAM and CLARA.</li> <li>✓ However, the computational complexity of CLARANS is about <math>O(n^2)</math>, where n is the number of objects.</li> <li>✓ Furthermore, its clustering quality is dependent on the sampling method used..</li> </ul>
8	<p><b>Explain how data mining is used for retail industry. (May/June 2014) (13M)</b> BTL3</p> <p><b>Answer: Page: 5.79- Poonkuzhali</b></p> <p>Introduction (2M)</p> <p>Design and construction of data warehouse based on benefits of data mining(2M)</p> <p>Multidimensional analysis of sales, customers, products, time and region (2M)</p> <p>Analysis of the effectiveness of sales (2M)</p> <p>Customer retention- analysis of customer loyalty (2M)</p> <p>Production Recommendation and cross referencing of items (3M)</p>
9	<p><b>Discuss the important requirements for Cluster Analysis. (May/June 2015) (13M)</b></p> <p>BTL2</p> <p><b>Answer: Page: 5.3 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Data matrix (or object-by-variable structure):(2M)</li> <li>✓ Dissimilarity matrix (or object-by-object structure): (2M)</li> <li>✓ Interval-scaled variables (2M)</li> <li>✓ Binary Variables (2M)</li> <li>✓ Categorical Variables (2M)</li> <li>✓ Ordinal Variables (1M)</li> <li>✓ Ratio scaled variable (1 M)</li> <li>✓ Need for these methods (1M)</li> </ul>
10	<p><b>Discuss the applications and trends in data mining in detail. (Nov/Dec 2016) (13M)</b></p> <p>BTL3</p> <p><b>Answer: Page: 5.78 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ Data Mining for Financial Data Analysis(2M)</li> <li>✓ Data Mining for the Retail Industry (2M)</li> <li>✓ Data Mining for the Telecommunication Industry (2M)</li> <li>✓ Data Mining for Biological Data Analysis (2M)</li> <li>✓ Data Mining in Other Scientific Applications (2M)</li> <li>✓ Data Mining for Intrusion Detection (3M)</li> </ul>
11	<p><b>Explain in detail about hierarchical based method. (13M)</b> BTL1</p> <p><b>Answer: Page: 5.30 - Poonkuzhali</b></p> <p>A hierarchical clustering method works by grouping data objects into a tree of clusters. Hierarchical clustering methods can be further classified as either agglomerative or divisive, depending on whether the hierarchical decomposition is formed in a bottom-up (merging) or top-down (splitting) fashion.</p> <ul style="list-style-type: none"> <li>✓ <u>Agglomerative hierarchical clustering:</u> (6M)</li> </ul> <p>This bottom-up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all of the objects are in a single cluster or until certain termination conditions are satisfied. Most hierarchical clustering methods belong to this category. They differ only in their definition of inter</p>

	<p>cluster similarity.</p> <ul style="list-style-type: none"> <li>✓ Divisive hierarchical clustering: (7M) This top-down strategy does the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster. It subdivides the cluster into smaller and smaller pieces, until each object forms a cluster on its own or until it satisfies certain termination conditions, such as a desired number of clusters is obtained or the diameter of each cluster is within a certain threshold</li> <li>✓ BIRCH: Balanced Iterative Reducing and Clustering Using Hierarchies</li> <li>✓ ROCK: A Hierarchical Clustering Algorithm for Categorical Attributes</li> <li>✓ Chameleon: A Hierarchical Clustering Algorithm Using Dynamic Modeling</li> </ul>
12	<p><b>Explain in detail about density based methods. (13M) BTL1</b></p> <p><b>Answer: Page: 5.35 - Poonkuzhali</b></p> <p>To discover clusters with arbitrary shape, density-based clustering methods have been developed. These typically regard clusters as dense regions of objects in the data space that are separated by regions of low density (representing noise). DBSCAN grows clusters according to a density-based connectivity analysis. OPTICS extends DBSCAN to produce a cluster ordering obtained from a wide range of parameter settings. DENCLUE clusters objects based on a set of density distribution functions. (4M)</p> <ul style="list-style-type: none"> <li>✓ DBSCAN: A Density-Based Clustering Method Based on Connected Regions with Sufficiently High Density (3M)</li> <li>✓ OPTICS: Ordering Points to Identify the Clustering Structure(3M)</li> <li>✓ DENCLUE: Clustering Based on Density Distribution Functions(3M)</li> </ul>
13	<p><b>What is grid based clustering? With an example explain an algorithm for grid based clustering. (13M) BTL1</b></p> <p><b>Answer: Page: 5.39 - Poonkuzhali</b></p> <p>The grid-based clustering approach uses a multi-resolution grid data structure. It quantizes the object space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed. The main advantage of the approach is its fast processing time, which is typically independent of the number of data objects, yet dependent on only the number of cells in each dimension in the quantized space. Some typical examples of the grid-based approach include STING, which explores statistical information stored in the grid cells; WaveCluster, which clusters objects using a wavelet transform method; and CLIQUE, which represents a grid-and density-based approach for clustering in high-dimensional data space. (3M)</p> <ul style="list-style-type: none"> <li>✓ STING: STatistical INformation Grid (5M)</li> <li>✓ WaveCluster: Clustering Using Wavelet Transformation (5M)</li> </ul>
14	<p><b>Demonstrate in detail about model based clustering methods. (13M) BTL1</b></p> <p><b>Answer: Page: 5.43 - Poonkuzhali</b></p> <p>Model-based clustering methods attempt to optimize the fit between the given data and some mathematical model. Such methods are often based on the assumption that the data are generated by a mixture of underlying probability distributions.</p> <ul style="list-style-type: none"> <li>✓ Expectation-Maximization (5M)</li> <li>✓ Conceptual Clustering(5M)</li> <li>✓ Neural Network Approach(3M)</li> </ul>
15	<p><b>Illustrate the topic on (i) CLIQUE (ii) DBSCAN (13M). BTL1</b></p> <p><b>Answer: Page: 5.38 - Poonkuzhali</b></p> <ul style="list-style-type: none"> <li>✓ CLIQUE (7M) Feature transformation methods, such as principal component analysis and singular</li> </ul>

	<p>value decomposition, transform the data onto a smaller space while generally preserving the original relative distance between objects.</p> <ul style="list-style-type: none"> <li>▪ CLIQUE: A Dimension-Growth Subspace Clustering Method</li> <li>✓ DBSCAN (6M)</li> </ul> <p>To discover clusters with arbitrary shape, density-based clustering methods have been developed. These typically regard clusters as dense regions of objects in the data space that are separated by regions of low density (representing noise). DBSCAN grows clusters according to a density-based connectivity analysis. OPTICS extends DBSCAN to produce a cluster ordering obtained from a wide range of parameter settings. DENCLUE clusters objects based on a set of density distribution functions.</p> <ul style="list-style-type: none"> <li>▪ DBSCAN: A Density-Based Clustering Method Based on Connected Regions with Sufficiently High Density</li> </ul>
--	---

**PART \* C**

1	<p><b>Illustrate the topic on 1. CLIQUE 2.DBSCAN. (15M) BTL1</b></p> <p><b>Answer: Page: 5.38 - Poonkuzhali</b></p> <p><b>CLIQUE (7M)</b> Feature transformation methods, such as principal component analysis and singular value decomposition, transform the data onto a smaller space while generally preserving the original relative distance between objects.</p> <p><b>CLIQUE: A Dimension-Growth Subspace Clustering Method</b></p> <p><b>DBSCAN (8M)</b> To discover clusters with arbitrary shape, density-based clustering methods have been developed. These typically regard clusters as dense regions of objects in the data space that are separated by regions of low density (representing noise). DBSCAN grows clusters according to a density-based connectivity analysis. OPTICS extends DBSCAN to produce a cluster ordering obtained from a wide range of parameter settings. DENCLUE clusters objects based on a set of density distribution functions.</p> <p><b>DBSCAN: A Density-Based Clustering Method Based on Connected Regions with Sufficiently High Density</b></p>
---	--

**CS6703****GRID AND CLOUD COMPUTING****L T P C****3 0 0 3****OBJECTIVES:**

- Understand how Grid computing helps in solving large scale scientific problems.
- Gain knowledge on the concept of virtualization 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**

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:**

1. 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.

**Subject Code: CS6703**  
**Subject Name: GRID AND CLOUD COMPUTING**

**Year/Semester: IV /07**  
**Subject Handler: S.NEELAKANDAN**

<b>UNIT -1- INTRODUCTION</b>	
<b>PART A</b>	
<b>Q.NO</b>	<b>QUESTIONS</b>
1.	<p><b>Illustrate the evolutionary trend towards parallel distributed and cloud computing.</b> BTL – 3</p> <p>The evolutionary trends emphasize the extension of the Internet to everyday objects. Instead of using a centralized computer to solve computational problems, a parallel and distributed computing system uses multiple computers to solve large-scale problems over the Internet.</p> <p>Some of the evolutionary modern computing techniques are</p> <ul style="list-style-type: none"> <li>➤ Cloud Computing</li> <li>➤ Ubiquitous Computing</li> <li>➤ High-performance Computing</li> <li>➤ High Throughput Computing</li> <li>➤ Internet of Things</li> </ul>
2.	<p><b>List and explain in brief the three new computing paradigms.</b> BTL – 1</p> <ul style="list-style-type: none"> <li>➤ Infrastructure as a Service(Iaas)</li> <li>➤ Platform as a Service(Paas)</li> <li>➤ Software as a Service(Saas)</li> </ul> <p>Iaas-instant computing infrastructure, provisioned over the internet      Paas-Third party providers delivers hardware and software tools      Saas-use cloud based application over the internet</p>
3.	<p><b>Describe the applications of high performance and high throughput systems.</b> April/May 2017 BTL – 1</p> <p>Applications of High-Performance Computing:</p> <ul style="list-style-type: none"> <li>➤ Weather predictors</li> <li>➤ Manufacturing process</li> <li>➤ Chemical reactors</li> <li>➤ Earth Observation</li> <li>➤ Military Sensors</li> </ul> <p>Applications of High Throughput Computing</p> <ul style="list-style-type: none"> <li>➤ Geostatistical simulations and analysis</li> <li>➤ 2D/3D hydrodynamic modeling</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Ecological modeling</li> </ul>
4.	<b>Define cyber physical systems. BTL – 1</b> <ul style="list-style-type: none"> <li>➤ A cyber-physical system (CPS) is the result of interaction between computational processes and the physical world.</li> <li>➤ A CPS integrates cyber (heterogeneous, asynchronous) with physical (concurrent and information-dense) objects.</li> <li>➤ A CPS merges the “3C” technologies of computation, communication, and control into an intelligent closed feedback system between the physical world and the information world</li> </ul>
5.	<b>Analyze the working of GPUs. BTL – 4</b> <ul style="list-style-type: none"> <li>➤ A GPU stands for graphics processing unit, a coprocessor or accelerator mounted on a computer’s graphics card or video card</li> <li>➤ Both multi-core CPU and many-core GPU processors can handle multiple instruction threads at different magnitudes.</li> <li>➤ GPU offloads the CPU from all data-intensive calculations and from tedious graphics tasks in video editing applications.</li> </ul>
6.	<b>Classify the primitive operations of virtual machines. BTL – 3</b> <ul style="list-style-type: none"> <li>➤ A virtual machine runs a dedicated operating system on shared physical hardware and referred as host</li> <li>➤ Each Virtual machine shares hardware resources of host machine(including CPU, RAM, Storage, Network) to run independent operating systems</li> </ul>
7.	<b>List out the cluster design. BTL – 1</b> <p>There are 3 types of clusters Designs are</p> <ul style="list-style-type: none"> <li>➤ Fail-over,</li> <li>➤ Load-balancing</li> <li>➤ HIGH Performance Computing</li> </ul>
8.	<b>Differentiate compute grid and data grid. BTL – 4</b> <ul style="list-style-type: none"> <li>➤ Compute Grid is a computation, optionally split it into multiple parts, and execute them on different grid nodes in parallel.</li> <li>➤ Computation will perform faster than data grid</li> <li>➤ Data Grids is distribution of data across the grid</li> <li>➤ Data grid performance is slow</li> </ul>
9.	<b>Examine the reasons to adapt the cloud for upgraded internet applications and web services. BTL - 3</b> <ul style="list-style-type: none"> <li>➤ Increases Performance</li> <li>➤ Security</li> <li>➤ Flexibility</li> </ul>

10.	<p><b>Discuss on SOA.</b> BTL – 2</p> <ul style="list-style-type: none"> <li>➤ SOA: A service-oriented architecture (SOA) is a software design where services are provided to the other components by application components, through a communication protocol over a network.</li> <li>➤ The basic principles of service-oriented architecture are independent of vendors, products and technologies</li> </ul> <p>Four properties of SOA</p> <ul style="list-style-type: none"> <li>➤ It logically represents a business activity with a specified outcome.</li> <li>➤ It is self-contained.</li> <li>➤ It is a black box for its consumers.</li> <li>➤ It may consist of other underlying services</li> </ul>										
11.	<p><b>Differentiate grid computing versus cloud computing.</b> BTL – 2</p> <table border="1" data-bbox="266 671 1127 956"> <tbody> <tr> <td>Grid Computing</td><td>Cloud Computing</td></tr> <tr> <td>Application Oriented</td><td>Service Oriented</td></tr> <tr> <td>Less Scalable</td><td>Very Scalable</td></tr> <tr> <td>Follow queuing pattern</td><td>Follow usage pattern</td></tr> <tr> <td>It is part of cloud computing</td><td>It is an independent concept</td></tr> </tbody> </table>	Grid Computing	Cloud Computing	Application Oriented	Service Oriented	Less Scalable	Very Scalable	Follow queuing pattern	Follow usage pattern	It is part of cloud computing	It is an independent concept
Grid Computing	Cloud Computing										
Application Oriented	Service Oriented										
Less Scalable	Very Scalable										
Follow queuing pattern	Follow usage pattern										
It is part of cloud computing	It is an independent concept										
12.	<p><b>Formulate the features of MPI ,Mapreduce and Hadoop.</b> Nov/Dec 2016 BTL- 6</p> <p>MPI Features:</p> <ul style="list-style-type: none"> <li>➤ High performance, scalability, and portability.</li> </ul> <p>Mapreduce Features:</p> <ul style="list-style-type: none"> <li>➤ Process large amounts of data –Flexibilty in data processing,Easily scalable,Fault tolerant</li> </ul>										
13.	<p><b>Summarize the technologies available in grid standards.</b> BTL – 5</p> <p>National Institute of Standards and Technology, Smart Grid Cloud Computing for Smart Grid applications are also introduced in terms of efficiency, security and usability</p>										
14.	<p><b>Discuss on OGSA.</b> BTL -2</p> <ul style="list-style-type: none"> <li>➤ 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</li> <li>➤ describes a service-oriented architecture for a grid computing environment for business and scientific use</li> </ul>										
15.	<p><b>Where OGSI and OGSA-DAI is utilized?</b> BTL – 1</p> <p>Reuse – The <b>OSGI</b> component model makes it very easy to use many third party components in an application</p> <p><b>OGSA-DAI</b> has contributed to projects and organisations around the world, in sectors including medical research, geographical information systems, meteorology, transport, computer-aided design, engineering and astronomy</p>										

16.	<b>Analyze the features of grid FTP. BTL – 4</b>  GridFTP is an extension of the File Transfer Protocol (FTP) for grid computing It is more reliable and high performance file transfer, GridFTP integrates with the Grid Security Infrastructure, which provides authentication and encryption to file transfers, with user-specified levels of confidentiality and data integrity, also for cross-server transfers
17.	<b>Define WSRF. BTL – 1</b> <ul style="list-style-type: none"> <li>➤ The Web Services Resource Framework (WSRF) defines a generic and open framework for modeling and accessing stateful resources using Web Services.</li> <li>➤ This describe views on the state, to support management of the state through properties associated with the Web Service,</li> <li>➤ To describe how these mechanisms are extensible to groups of Web Services.</li> </ul>
18.	<b>Describe the standards related to web service. BTL – 2</b>  XML SOAP WSDL UDDI
19.	<b>Summarize the elements of grid. BTL – 5</b>  Grid computing relies on complete computer systems, this means accessing devices such as desktop computers that have on-board CPUs, storage and power supplies. All of the elements of your grid require connectivity, usually in the form of internet connectivity, possibly through an Ethernet connection. Middleware: This software enables you to donate your idle computer time to projects like BOINC, UNICORE,
20.	<b>Generalize the layers in grid architecture. BTL- 6</b> <ul style="list-style-type: none"> <li>➤ Application Layer</li> <li>➤ Collective Layer</li> <li>➤ Resource Layer</li> <li>➤ Connectivity Layer</li> <li>➤ Fabric Layer</li> </ul>
21	<b>What is QOS? BTL1</b> <ul style="list-style-type: none"> <li>➤ Grid computing system is the ability to provide the quality of service requirements necessary for the end-user community.</li> <li>➤ QOS provided by the grid like performance, availability, management aspects.</li> </ul>

22	<b>Discuss business on demand.</b> BTL2
	<ul style="list-style-type: none"> <li>➤ 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.</li> <li>➤ 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.</li> </ul>
23	<b>List the properties of Cloud Computing. Nov/Dec 2016</b> BTL1 <p>There are six key properties of cloud computing:</p> <p>Cloud computing is</p> <ul style="list-style-type: none"> <li>➤ user-centric</li> <li>➤ task-centric</li> <li>➤ powerful</li> <li>➤ accessible</li> <li>➤ intelligent</li> <li>➤ programmable</li> </ul>
24	<b>What is meant by resource broker?</b> BTL1 <p>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.</p>
25	<b>Define – Distributed Computing.</b> BTL2 <p>Distributed computing is a field of computer science that studies distributed systems. A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal.</p>
<b>PART B</b>	
1	<b>i)Identify and explain in detail about evolutionary trend of computer technology.</b> (7M) BTL 1 <b>Answer: Page.1-2 Bhushan</b> <b>Evolution Trend: (2M)</b> Evolutionary trends emphasize - Internet extension. Parallel and distributed computing system - solve large-scale problems. <b>Modern Techniques(5M)</b> <ul style="list-style-type: none"> <li>➤ Cloud Computing</li> <li>➤ Ubiquitous Computing</li> <li>➤ High-performance Computing</li> </ul>

	<ul style="list-style-type: none"> <li>➤ High Throughput Computing</li> <li>➤ Internet of Things</li> </ul> <p><b>ii) Explain the three paradigms in detail. (6M) BTL 1</b></p> <p><b>Answer:page-3-7 Bhushan</b></p> <p><b>Service Models (4M)</b></p> <ul style="list-style-type: none"> <li>➤ Iaas(Infrastructure as a Service)</li> <li>➤ Paas(Platform as a Service)</li> <li>➤ Saas (Software as a Service)</li> </ul> <p>Iaas - Instant computing infrastructure, provisioned - internet</p> <p>Paas - Third party providers delivers hardware- software tools</p> <p>Saas - Use cloud based application – internet</p>
2	<p><b>i)Summarize in detail about the degrees of parallelism. (8M) BTL 2</b></p> <p><b>Answer:page-6 Bhushan</b></p> <p><b>Degree of Parallelism (2M)</b></p> <ul style="list-style-type: none"> <li>➤ DOP – metric - how many operations simultaneously executed by a computer.</li> </ul> <p><b>Parallelism(6M)</b></p> <ul style="list-style-type: none"> <li>➤ Program – running on parallel computer - utilize different numbers of processors - different times.</li> <li>➤ Number of processors - execute a program - degree of parallelism.</li> <li>➤ Plot DOP - function time for given program - parallelism profile.</li> </ul> <p><b>Discuss the application of high performance and high throughput system. (April/May 2017) (4M) BTL 2</b></p> <p><b>Answer :page - 1-3,1-4 Bhushan</b></p> <p><b>Application List (4M)</b></p> <p><b>Applications of High-Performance Computing:</b></p> <ul style="list-style-type: none"> <li>➤ Weather predictors</li> <li>➤ Manufacturing process</li> <li>➤ Chemical reactors</li> <li>➤ Earth Observation</li> <li>➤ Military Sensors</li> </ul> <p><b>Applications of High Throughput Computing</b></p> <ul style="list-style-type: none"> <li>➤ Geostatistical simulations and analysis</li> <li>➤ 2D/3D hydrodynamic modeling</li> <li>➤ Ecological modeling</li> </ul>
3	<p><b>i)Demonstrate in detail about internet of things and cyber physical systems. (7m)</b></p> <p><b>BTL 3</b></p> <p><b>Answer:page -10 Bhushan</b></p> <p><b>IOT(4M)</b></p> <p><b>Internet of Things :</b></p> <ul style="list-style-type: none"> <li>➤ Radio frequency Identification(RFID) - Global Positioning System(GPS)</li> <li>➤ Iot - network connection - computers,sendors - human-centric devices</li> <li>➤ Iot objects - devices interconnected- network interact intelligently</li> </ul> <p><b>CPS(3M)</b></p> <p><b>Cyber Physical System:</b></p>

	<ul style="list-style-type: none"> <li>➤ CPS- interaction between computational processes -physical world.</li> <li>➤ CPS integrates cyber (heterogeneous, asynchronous) - physical (concurrent and informationdense) objects.</li> <li>➤ CPS - merges “3C” technologies- computation, communication - control</li> </ul> <p><b>ii)Examine the memory ,storage and wide area networking technology in network based system. (6M) BTL 3</b></p> <p><b>Answer:</b>page- 8 Bhushan</p> <p><b>Network Based System:</b></p> <p><b>Memory Technology(2M)</b> Faster processor speed - larger memory capacity result - wider gap - processors memory.</p> <p><b>Disk and Storage Technology(2M)</b> Rapid growth of flash memory - solid-state drives (SSDs) - impacts future HPC , HTC systems.</p> <p>Power increases linearly - clock frequency - quadratically - voltage applied on chips</p> <p><b>System-Area Interconnects(2M)</b> SAN- connects servers - network storage -disk arrays. Network attached storage (NAS) - connects client hosts - disk arrays.</p> <p><b>Wide-Area Networking:</b> High-bandwidth networking increases - capability building –massively for distributed systems.</p>
4	<p><b>Define and examine in detail about the multi core CPUs and multithreading technologies. (13M) BTL 1</b></p> <p><b>Answer:</b> page - 3-5 Bhushan</p> <p><b>Explanation( 8M)</b></p> <p><b>Diagram(3M)</b></p> <p><b>CPU Technologies (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Multicore CPUs - increase tens cores - hundreds or more</li> <li>➤ DLP - forementioned memory wall problem.</li> </ul> <p><b>Multicore CPU and Many-Core GPU Architectures (4M)</b></p> <ul style="list-style-type: none"> <li>➤ Triggered development many-core GPUs - hundreds or more thin cores.</li> <li>➤ The GPU - applied large clusters - build supercomputers in MPPs.</li> </ul> <p><b>Multithreading Technology(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Only instructions</li> <li>➤ Fine-grain multithreading switches - execution of instructions - different threads per cycle.</li> <li>➤ Course-grain multithreading executes many instructions - same thread - few cycles before switching to another thread.</li> <li>➤ The multicore CMP executes instructions - different threads completely.</li> </ul>
5	<p><b>Analyze in detail about the GPU programming model. (13M) BTL 4</b></p> <p><b>Answer:</b> page – 22 Bhushan</p> <p><b>GPU(2M)</b></p> <ul style="list-style-type: none"> <li>➤ A GPU -graphics coprocessor - accelerator mounted on computer’s graphics card or video card.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ A GPU offloads - CPU from tedious graphics tasks, video editing applications.</li> </ul> <p><b>Programming Model (8M)</b></p> <ul style="list-style-type: none"> <li>➤ A modern GPU chip - built with hundreds of processing cores.</li> <li>➤ GPUs - throughput architecture that exploits massive parallelism</li> <li>➤ NVIDIA GPU - upgraded to 128 cores on single chip. GPU - handle eight threads of instructions.</li> <li>➤ GPUs - designed to handle large numbers of floating point operations in parallel.</li> </ul> <p><b>Diagram(3M)</b></p>
6	<p><b>i)Evaluate virtual machine and virtualization middleware in network based system? (8M) BTL 5</b></p> <p><b>Answer:</b>page - 3-1,3-20 <b>Bhushan Virtual Machine (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Virtual machines (VMs) - novel solutions to underutilized resources - application inflexibility - software manageability - security concerns in existing physical machines.</li> </ul> <p><b>Middleware(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Large clusters, grids, and clouds - large amounts of computing, storage, and networking resources - virtualized manner.</li> <li>➤ The VM - virtual resources managed by a guest - OS run a specific application.</li> <li>➤ VMs host platform - deploy a middleware layer called a virtual machine monitor (VMM).</li> <li>➤ The VMM called a hypervisor - privileged mode.</li> </ul> <p><b>Diagram (2M)</b></p> <p><b>ii)Explain the convergence of technologies in detail. (5M) BTL 5</b></p> <p><b>Answer:</b>page - 3-11 <b>Bhushan</b></p> <p><b>Convergence (5M)</b></p> <ul style="list-style-type: none"> <li>➤ Information technology and media - sectors originally operated largely independent of one another</li> <li>➤ Technical side - ability of any infrastructure - transport any type of data,</li> <li>➤ Functional side - consumers integrate in a seamless way - functions of computation, entertainment.</li> </ul>
	<b>PART C</b>
1	<p><b>Illustrate the architecture of virtual machine and brief about the operations. (15M) 'Nov/Dec 2016 BTL 2</b></p> <p><b>Answer:</b>page - 3-11 <b>Bhushan</b></p> <p><b>VM(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Virtual machines (VMs) -underutilized resources, application inflexibility, software manageability, and security concerns - existing physical machines.</li> </ul> <p><b>Operations(8M)</b></p>

	<ul style="list-style-type: none"> <li>➤ Build large clusters, grids, and clouds - need to access large amounts of computing, storage, and networking resources - virtualized manner.</li> <li>➤ Aggregate the resources - offer a single system image.</li> <li>➤ The VM - virtual resources managed by a guest OS - run a specific application.</li> <li>➤ VMs and host platform-deploy a middleware layer called a virtual machine monitor (VMM).</li> <li>➤ The VMM - hypervisor in privileged mode.</li> </ul> <p><b>VM Operations (5M)</b></p> <p><b>VM Primitive Operations:</b></p> <p>VMM provides the VM abstraction - guest OS.</p> <ul style="list-style-type: none"> <li>➤ First, the VMs - multiplexed between hardware machines</li> <li>➤ Second, a VM - suspended stored in stable storage</li> <li>➤ Third, a suspended VM - resumed or provisioned to new hardware platform</li> <li>➤ Finally, a VM - migrated from one hardware platform to another</li> </ul>
	<p>➤ Write a short note on: Clusters of Cooperative Computers. (7M) BTL1</p> <p><b>Answer:page - 3-28 Bhushan</b></p> <p><b>Clusters(2M)</b></p> <p><b>Cluster Architecture</b></p> <ul style="list-style-type: none"> <li>➤ Cluster – connected to internet via a virtual private network(VPN) gateway.</li> </ul> <p><b>Hardware, Software and Middleware Support</b></p> <p>Special cluster middleware supports – create SSI or high availability(HA) Both sequential and parallel application - clusters facilitate cluster resources.</p> <p><b>Major Cluster Design Issues:</b> Cluster wide OS for complete resource sharing</p> <p><b>Service Oriented Architecture (8M) ‘Nov/Dec 2016 BTL 1</b></p> <p><b>Answer:page -37 Bhushan</b></p> <p><b>SOA (2M)</b></p> <ul style="list-style-type: none"> <li>➤ In grids/web services, Java, and CORBA- an entity, a service, a Java object, and a CORBA - distributed object - variety of languages.</li> <li>➤ Architectures build on the traditional seven Open Systems Interconnection (OSI) layers - base networking abstractions</li> </ul> <p><b>Explanation (4M)</b></p> <ul style="list-style-type: none"> <li>➤ A service-oriented architecture - a collection of services.</li> <li>➤ Services communicate with each other.</li> <li>➤ The communication involve simple data passing - two or more services coordinating some activity.</li> </ul> <p><b>Diagram(2M)</b></p>
3	<p><b>Brief the interaction between the GPU and CPU in performing parallel execution of Operations (15M) ‘April/May 2017 BTL 4</b></p> <p><b>Answer:page – 32 Bhushan</b></p> <p><b>GPU (2M)</b></p>

	<ul style="list-style-type: none"> <li>➤ A GPU - graphics coprocessor or accelerator mounted on a computer's graphics card or video card.</li> <li>➤ A GPU offloads the CPU from tedious graphics - tasks in video editing applications.</li> </ul> <p><b>Comparison (8M)</b></p> <ul style="list-style-type: none"> <li>➤ A modern GPU chip - hundreds of processing cores. GPUs have a throughput architecture - exploits massive parallelism - executing many concurrent threads slowly</li> <li>➤ NVIDIA GPU - 128 cores on a single chip- each core on a GPU - handle eight threads of instructions.</li> </ul> <p><b>Diagram(5M)</b></p>
--	--

<b>UNIT II GRID SERVICES</b>	
Introduction to Open Grid Services Architecture (OGSA)-Motivation-Functionality Requirements-Practical & Detailed view of OGSA/OGSI-Data Intensive grid service models-OGSA services.	
<b>PART A</b>	
1	<b>Define OGSA. April/May 2017 BTL 1</b>  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 describes a service-oriented architecture for a grid computing environment for business and scientific use.
2	<b>Illustrate the relationship between resources and service. BTL 3</b>  A service can be defined as a self-contained, independently developed, deployed, managed, and maintained software implementation supporting specific business-relevant functionality for an enterprise as a whole and is “integratable” by design. A resource can be defined as a directly-accessible, independently-developed, deployed, managed and maintained software artifact supporting specific data.
3	<b>List the major goals of OGSA. BTL 1</b> <ul style="list-style-type: none"> <li>➤ To manage the resources across distributed heterogeneous platform</li> <li>➤ To support QOS oriented service level agreements</li> <li>➤ To define a standard to achieve interoperability</li> </ul>
4	<b>Summarize on the goals of GGF. BTL 2</b>  The Global Grid Forum has started an initiative to propose a standard architecture for grid computing to improve interoperability security, resource sharing, capability, policy management and grid manageability called OGSA and OGSI.
5.	<b>What is WSRF? BTL 1</b>  The Web Services Resource Framework (WSRF) defines a generic and open framework for modeling and accessing stateful resources using Web Services.T

	his includes mechanisms to describe views on the state, to support management of the state through properties associated with the Web Service.
6.	<p><b>Define grid infrastructure.</b> BTL1</p> <p>Grid infrastructure is a complex combination of a number of capabilities and resources identified for the specific problem and environment being addressed. It forms the core foundations for successful grid applications.</p>
7	<p><b>Summarize on grid service migration using GSH and GSR.</b> BTL 5</p> <p>Grid service migration is a mechanism for creating new services and specifying assertions regarding the lifetime of a service.</p> <ul style="list-style-type: none"> <li>➤ A GSH is a globally unique name that distinguishes a specific grid service instance from all others.</li> <li>➤ The OGSA employs a “handle-resolution” mechanism for mapping from a GSH to a GSR. The GSH must be globally defined for a particular.</li> </ul>
8	<p><b>Analyze the OGSA security model at various protection levels.</b> BTL 4</p> <ul style="list-style-type: none"> <li>➤ The OGSA supports security enforcement at various levels.</li> <li>➤ The grid works in a heterogeneous distributed environment, which is essentially open to the general public.</li> <li>➤ At the security policy and user levels, apply a service or endpoint policy, resource mapping rules, authorized access of critical resources, and privacy protection.</li> </ul>
9	<p><b>Discuss the strategies of data replication.</b> BTL 2</p> <p>Replication strategies determine when and where to create a replica of the data. The factors to consider include data demand, network conditions, and transfer cost.</p> <p>The strategies of replication can be classified into method types:</p> <ul style="list-style-type: none"> <li>➤ Dynamic and Static.</li> <li>➤ Dynamic strategies can adjust locations and number of data replicas according to changes in conditions.</li> </ul>
10	<p><b>List the model for organizing the data grid.</b> BTL 1</p> <ul style="list-style-type: none"> <li>➤ Monadic model</li> <li>➤ Hierarchical model</li> <li>➤ Federation model</li> <li>➤ Hybrid model</li> </ul>
11	<p><b>Differentiate parallel data transfer versus striped data transfer.</b> BTL 2</p> <p>Parallel data transfer opens multiple data streams for passing subdivided segments of a file simultaneously. Striped data transfer-a data object is partitioned into a number of sections.</p>
12	<p><b>Give the basic services of OGSA.</b> BTL 2</p> <ul style="list-style-type: none"> <li>➤ Infrastructure Service</li> <li>➤ Execution Management Services</li> <li>➤ Data Management Services</li> <li>➤ Resource Management Services</li> <li>➤ Security Services</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Information Services</li> <li>➤ Self-Management Services</li> </ul>
13	<b>Define WSRF.</b> BTL 1 <ul style="list-style-type: none"> <li>➤ The Web Services Resource Framework (WSRF) defines a generic and open framework for modeling and accessing stateful resources using Web Services.</li> <li>➤ This includes mechanisms to describe views on the state, to support management of the state through properties associated with the Web Service</li> </ul>
14	<b>Point out the objectives of OGSA.</b> BTL 4 <ul style="list-style-type: none"> <li>➤ To manage the resources across distributed heterogeneous platform</li> <li>➤ To support QOS oriented service level agreements</li> </ul>
15	<b>Deduce the fundamental requirements for describing Web services based on the OGSI.</b> BTL 5 <p>There are two fundamental requirements for describing Web services based on the OGSI:</p> <ul style="list-style-type: none"> <li>➤ The ability to describe interface inheritance—a basic concept with most of the distributed object systems.</li> <li>➤ The ability to describe additional information elements with the interface definitions.</li> </ul>
16	<b>Define grid service instance.</b> BTL 1 <p>A grid service instance is a service that conforms to a set of conventions, expressed as WSDL interfaces, extensions, and behaviours, for such purposes as lifetime management, discovery of characteristics, and notification.</p>
17	<b>Name the concepts involved in the components of OGSI.</b> BTL 1 <ul style="list-style-type: none"> <li>➤ Stateful Web Services</li> <li>➤ Extension of Web Service Interfaces</li> <li>➤ Asynchronous notification of state change</li> <li>➤ Reference to instances of services</li> </ul>
18	<b>Illustrate the Two approaches to the implementation of argument demarshaling functions in a grid service hosting environment.</b> BTL 3 <ul style="list-style-type: none"> <li>➤ OGSI does not dictate a particular service-provider-side implementation architecture.</li> <li>➤ A container implementation may provide a range of functionality beyond simple argument demarshaling.</li> </ul>
19	<b>Analyze the functional requirements of OGSA.</b> BTL 4 <p>Various functional requirements are:</p> <ul style="list-style-type: none"> <li>➤ Discovery and bokering</li> <li>➤ Metering and auditing</li> <li>➤ Data sharing and management</li> <li>➤ Deployment</li> <li>➤ Virtual organizations</li> <li>➤ Application Monitoring</li> </ul>

20	<b>Formulate the motivations that drive OGSA standards.</b> BTL 6 <ul style="list-style-type: none"> <li>➤ The OGSA developed within the OGSA Working Group of the Global Grid Forum, is a service-oriented architecture that aims to define a common, standard, and open architecture for grid-based applications.</li> <li>➤ “Open” refers to both the process to develop standards and the standards themselves.</li> </ul>
21	<b>What is meant by grid infrastructure?</b> BTL 1 <p>Grid infrastructure is a complex combination of a number of capabilities and resources identified for the specific problem and environment being addressed. It forms the core foundations for successful grid applications.</p>
22	<b>List the layers available in OGSA architectural organizations.</b> BTL 1 <ul style="list-style-type: none"> <li>➤ Native platform services and transport mechanisms.</li> <li>➤ OGSA hosting environment.</li> <li>➤ OGSA transport and security.</li> <li>➤ OGSA infrastructure (OGSI).</li> <li>➤ OGSA basic services (meta-OS and domain services).</li> </ul>
23	<b>Discuss the role of the grid computing organization.</b> BTL 2 <ul style="list-style-type: none"> <li>➤ Organizations developing grid standards and best practices guidelines.</li> <li>➤ Organizations developing grid computing toolkits, frameworks and middleware solutions.</li> <li>➤ Organizations building and using grid - based solutions to solve their computing, data, and network requirements.</li> <li>➤ Organizations working to adopt grid concepts into commercial products, via utility computing and business on demand computing.</li> </ul>
24	<b>Define WSDL.</b> BTL 1 <p>WSDL is an XML Info set based document, which provides a model and XML format for describe web services. This enables services to be described and enables the client to consume these services in a standard way without knowing much on the lower level protocol exchange binding including SOAP and HTTP.</p>
25	<b>Mention the fundamental components of SOAP specification.</b> BTL 2 <ul style="list-style-type: none"> <li>➤ An envelope that defines a framework for describing message structure.</li> <li>➤ A set of encoding rules for expressing instances of application defined data types.</li> <li>➤ A convention for representing remote procedure (RPC) and responses. A set of rules for using SOAP with HTTP. x Message exchange patterns (MEP) such as request-response, one-way and peer-to-peer conversations.</li> </ul>
<b>PART B</b>	
1	<b>i )Define OGSA and describe the grid service architecture in detail. (7M)</b> BTL 1 <b>Answer: Page - 2-2 Bhushan</b> <b>Definition(2M)</b> <ul style="list-style-type: none"> <li>➤ OGSA - open source grid service - standard jointly developed by academia - IT industry - working group in the Global Grid Forum (GGF).</li> <li>➤ Standard specifically developed for emerging grid - cloud service communities.</li> </ul> <b>Explanation (3M)</b>

	<p><b>OGSA Framework</b> Globus Toolkit - grid technology solution – scientific, technical computing - web services for business - network applications</p> <p><b>Diagram (2M)</b></p> <p><b>ii)Examine the grid service migration using GSH and GSR. (6M)</b> <b>(April/May 2017) BTL 1</b></p> <p><b>Answer: Page -2-4 Bhushan</b></p> <p><b>GSR(3M)</b></p> <ul style="list-style-type: none"> <li>➤ Creating new services - specifying assertions - lifetime of a service.</li> <li>➤ OGSA model - a standard interface - a factor to implement reference.</li> </ul> <p><b>GSH(3M)</b></p> <p><b>Grid Service Handle</b></p> <ul style="list-style-type: none"> <li>➤ GSH - globally unique name - distinguishes specific grid service instance</li> <li>➤ OGSA employs “handle-resolution” mechanism - mapping GSH to GSR.</li> <li>➤ GSH - globally defined for particular instance.</li> </ul>
2	<p><b>i)Summarize the OGSA security model implemented at various protection models. (7M) (April/May 2017) BTL 2</b></p> <p><b>Answer:Page - 2-6 Bhushan</b></p> <p><b>Defintion(2M)</b> OGSA - security enforcement levels - Grid works heterogeneously in distributed environment - open to general public.</p> <p><b>Explanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Security policy and user levels - apply service endpoint policy - resource mapping rules - authorized access critical resources - privacy protection.</li> <li>➤ Public Key Infrastructure (PKI) service level - OGSA demands security binding - security protocol stack and bridging of certificate authorities (CAs)</li> <li>➤ Trust models -secure logging practiced in grid platforms.</li> </ul> <p><b>i) Discuss how a GSH resolves to different GSR for migrated service instance. (6M) BTL 2</b></p> <p><b>Answer: Page 2-6 Bhushan</b></p> <p><b>GSR(3M)</b></p> <ul style="list-style-type: none"> <li>➤ Creating new services - specifying assertions regarding lifetime of service.</li> <li>➤ OGSA models standard interface -factor to implement reference.</li> <li>➤ Service address former services - reference of services.</li> <li>➤ Dynamically created grid service instance - associated with specified lifetime.</li> </ul> <p><b>GSH(3M)</b></p> <ul style="list-style-type: none"> <li>➤ GSH - globally unique name distinguishes specific grid service instance.</li> <li>➤ OGSA employs a “handle-resolution” mechanism mapping GSH to GSR.</li> <li>➤ GSH - globally defined for instance.</li> </ul>
3	<p><b>i)Demonstrate the service models of data intensive grid.(Nov/Dec 2016) (5M)</b> BTL 3</p> <p><b>Answer: Page - 1-12 Bhushan</b></p> <p><b>List and Expanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Applications grid normally grouped into two categories: computation-intensive and data-intensive.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Grid system - designed to discover, transfer, and manipulate - massive data sets.</li> <li>➤ Transferring massive data sets - time-consuming task.</li> <li>➤ Efficient data management demands low-cost storage - high-speed data movement.</li> </ul> <p><b>ii) Illustrate the architectural models for building a data grid. (8M) BTL 3</b></p> <p><b>Answer: Page - 1-15 Bhushan</b></p> <p><b>Types list(4M)</b></p> <p>There are four access models for organizing a data grid</p> <ul style="list-style-type: none"> <li>➤ Monadic model</li> <li>➤ Hierarchical model</li> <li>➤ Federation model</li> <li>➤ Hybrid model</li> </ul> <p><b>Diagram(4M)</b></p>
4	<p><b>i) Analyze the set of services for the building blocks of OGSA based grid. (5M)</b></p> <p>BTL 4</p> <p><b>Answer: Page - 2-2 Bhushan</b></p> <p><b>Explanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Deliver seamless QOS</li> <li>➤ Open published interfaces - provide interoperability diverse resources</li> <li>➤ Exploit industry standard integration technologies</li> <li>➤ Develop standards - achieve interoperability</li> </ul> <p><b>ii) Explain the services provided by OGSA architecture. (8M) (Nov/Dec 2016) BTL 4</b></p> <p><b>Answer: Page - 2-8 Bhushan</b></p> <p><b>List and Explanation(6M)</b></p> <ul style="list-style-type: none"> <li>➤ Infrastructure Services</li> <li>➤ Execution Management Services</li> <li>➤ Data Management Services</li> <li>➤ Resource Management Services</li> <li>➤ Security Services</li> <li>➤ Information Services</li> <li>➤ Self-Management Services</li> </ul> <p><b>Diagram(2M)</b></p>
5	<p><b>Describe in detail about the practical view of OGSA and OGSI. (13M) BTL 1</b></p> <p><b>Answer: Page - 2-10 Bhushan</b></p> <p><b>Definition (2M)</b></p> <p>Defined set of interfaces - systems built on open standards as WSDL.</p> <p><b>Explanation (8M)</b></p> <p>Manage resources - distributed heterogeneous platforms.</p> <p>QoS –Oriented Service Level Agreements(SLAs) - Topology of grid complex - interaction between grid resources - invariably dynamic –authorization,access control,delegation</p> <p><b>Diagram(2M)</b></p>

6	<p><b>i)Examine the client side programming patterns for grid services. (7M) BTL 3</b></p> <p><b>Answer: Page - 2-14 Bhushan</b></p> <p><b>Explanation(6M)</b></p> <p>OGSI exploits component of the web services framework: use of WSDL multiple protocol bindings - encoding styles - messaging styles web services.</p> <p>Web services Invocation Framework(WSIF) and Java API for XML RPC (JAX – RPC) - infrastructure software that provide capability</p> <p><b>Diagram(1M)</b></p> <p><b>ii)Demonstrate in detail about the conceptual hosting environment for grid service. (6M) BTL – 3</b></p> <p><b>Answer: Page - 2-14 Bhushan</b></p> <p><b>Explanation(6M)</b></p> <ul style="list-style-type: none"> <li>➤ OGSI - service-provider-side implementation architecture.</li> <li>➤ Container implementation - range of functionality- simple argument demarshaling.</li> </ul>
<b>PART C</b>	
1	<p><b>Explain the data intensive grid service models with suitable diagrams. (15M)</b></p> <p><b>'Nov/Dec 2016 BTL 1</b></p> <p><b>Answer: Page:2-14 in Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Applications grid - grouped into two categories: computation-intensive and data-intensive.</li> <li>➤ Grid system designed to discover, transfer, and manipulate - massive data sets.</li> </ul> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Transferring massive data sets - time-consuming task.</li> <li>➤ Efficient data management demands low-cost storage - high-speed data movement.</li> </ul> <p><b>Diagram(3M)</b></p>
2	<p><b>Write a neat sketch, Discuss the OGSA framework. (15M) 'Nov/Dec 2016 BTL2</b></p> <p><b>Answer: Page - 2-4 in Bhushan</b></p> <p><b>Definition(2M)</b></p> <p><b>OGSA Framework</b></p> <p><b>Explantion(10M)</b></p> <ul style="list-style-type: none"> <li>➤ OGSA - creation, termination, management - invocation of stateful transient grid services - standard interfaces conventions</li> <li>➤ OGSA framework specifies physical environment - security infrastructure - profile resource provisioning -virtual domains and execution environments - API access tools</li> </ul> <p><b>Diagram(3M)</b></p>
3	<p><b>Write a detailed note on OGSA security models. (15M) 'April/May 2017 BTL 1</b></p> <p><b>Answer: Page: 2-6 Bhushan</b></p> <p><b>Definition(2M)</b></p> <p>OGSA supports security enforcement levels. Grid works in heterogeneous distributed environment - open to general public</p> <p><b>Diagram(3M)</b></p>

**Explanation(10M )****OGSA Service Models**

- Security policy, user levels -apply a service or endpoint policy, resource mapping rules, authorized access - critical resources, and privacy protection.
- Public Key Infrastructure (PKI) service level - OGSA demands security binding - security protocol stack - bridging certificate authorities (CAs) - use of multiple trusted intermediaries
- Trust models and secure logging - practiced in grid platforms.

JIT 2106

**UNIT 3-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 centre Automation.

**PART A**

1	<b>Define public private and hybrid clouds.</b> Nov/Dec 2016 BTL 1 <ul style="list-style-type: none"> <li>➤ Public: Accessible, via the internet to anyone who pays E.g., Google App Engine, Amazon Web Service.</li> <li>➤ Private: Accessible via an internet to the members of owing organization E.g., NASA cloud for climate modeling.</li> <li>➤ Hybrid: A private cloud might buy computing resources from a public cloud.</li> </ul>
2	<b>Differentiate centralized and distributed computing.</b> BTL 2 <p>Centralized computing: This is a computing paradigm by which all computer resources are centralized in one physical system.</p> <p>Distributed computing: a distributed system consists of multiple autonomous computers, each having its own private memory, communicating through a computer network.</p>
3	<b>List the design objective of cloud.</b> BTL 1 <p>Make cloud systems scalable by design so that they can exploit the elasticity of the cloud, as well as maintaining and also improving scalability during system evolution.</p>
4	<b>Define IaaS.</b> BTL 1 <ul style="list-style-type: none"> <li>➤ This model allows users to use virtualized IT resources for computing, storage, and networking. In short, the service is performed by rented cloud infrastructure.</li> <li>➤ The user can deploy and run his applications over his chosen OS environment.</li> <li>➤ The user does not manage or control the underlying cloud infrastructure, but has control over the OS, storage, deployed applications, and possibly select networking components.</li> </ul>
5	<b>Generalize on PaaS and SaaS.</b> April/May 2017 BTL 6 <p>Platform as a Service(PaaS)</p> <p>To be able to develop,deploy and manage the execution of applications using provisioned resources demands a cloud platform with proper software environment.</p> <p>Software as a Service(SaaS):</p> <p>This refers to browser –initiated application software over thousands of cloud customers. Services and tools offered by Paas are utilized in construction of applications and management.</p>
6	<b>Show the levels of virtualization implementation.</b> BTL 3 <ul style="list-style-type: none"> <li>➤ After virtualization, different user applications managed by their own operating systems (guest OS) can run on the same hardware, independent of the host OS.</li> <li>➤ This is often done by adding additional software, called a virtualization layer. This virtualization layer is known as hypervisor or virtual machine monitor (VMM).</li> <li>➤ The VMs are shown in the upper boxes, where applications run with their own guest OS over the virtualized CPU, memory, and I/O resources.</li> </ul>
7	<b>Discuss the design requirements of VMM.</b> BTL 2 <p>There are three requirements for a VMM.</p> <ul style="list-style-type: none"> <li>➤ First, a VMM should provide an environment for programs which is essentially identical to the original machine.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Second, programs run in this environment should show, at worst, only minor decreases in speed.</li> <li>➤ Third, a VMM should be in complete control of the system resources.</li> </ul>
8	<p><b>Analyze the advantages and disadvantages of Grid Computing.</b> BTL 4</p> <p>Advantages</p> <ul style="list-style-type: none"> <li>➤ Can solve larger, more complex problems in a shorter time.</li> <li>➤ Easier to collaborate with other organizations.</li> <li>➤ Make better use of existing hardware.</li> </ul> <p>Disadvantages</p> <ul style="list-style-type: none"> <li>➤ Grid software and standards are still evolving.</li> <li>➤ Learning curve to get started.</li> <li>➤ Non-interactive job submission.</li> </ul>
9	<p><b>How does the virtualization support the Linux platform?</b> BTL 5</p> <p>The Xen hypervisor has been applied to virtualize x86-based machines to run Linux or other guest OS.</p>
10	<p><b>Compare binary translation with full virtualization.</b> BTL 4</p> <p>Full virtualization: With full virtualization, noncritical instructions run on the hardware directly while critical instructions are discovered and replaced with traps into the VMM to be emulated by software.</p> <p>Binary translation: VMware puts the VMM at Ring 0 and the guest OS at Ring 1. The VMM scans the instruction stream and identifies the privileged, control- and behavior-sensitive instructions.</p>
11	<p><b>Demonstrate the need of virtualization need of multi- core processor.</b> BTL 3</p> <ul style="list-style-type: none"> <li>➤ Virtualizing a multi-core processor is relatively more complicated than virtualizing a unicore processor.</li> <li>➤ Though multicore processors are claimed to have higher performance by integrating multiple processor cores in a single chip, muti-core virtualization has raised some new challenges to computer architects, compiler constructors, system designers, and application programmers.</li> </ul>
12	<p><b>Discuss the design issues of virtual clusters.</b> BTL 2</p> <p>Three critical design issues of virtual clusters:</p> <ul style="list-style-type: none"> <li>➤ Live migration of VMs,</li> <li>➤ Memory and file migrations,</li> <li>➤ Dynamic deployment of virtual clusters.</li> </ul>
13	<p><b>List the properties of Virtual clusters when virtual machines are dynamically allocated.</b> BTL 1</p> <p>The virtual cluster nodes can be either physical or virtual machines. Multiple VMs running with different OSes can be deployed on the same physical node</p>

14	<b>Define GSR. BTL 1</b> Grid service instances are made accessible to client applications through the use of a grid service handle and a grid service references(GSR)
15	<b>Describe the resource managers of eucalyptus for virtual network. BTL 2</b> CM(Cloud Manager) GM (Group Manager) IM (Instance Manager) Works like AWS APIs
16	<b>How the data storage is classified in virtual environment? BTL 3</b> Four main Layers are: ➤ Storage Device ➤ Block aggregation Layer ➤ File/Record Layer ➤ Application Layer
17	<b>Formulate the side effects of server virtualization. BTL 6</b> ➤ Creation of more high-density areas and hot spots ➤ Potentially detrimental effect on power usage effectiveness (PUE) ➤ Dynamic IT load swings ➤ Lower redundancy requirements
18	<b>Where OS level virtualization is needed? BTL 1</b> ➤ An abstraction layer between traditional OS and user applications ➤ OS-level virtualization is commonly used in creating virtual hosting environments to allocate hardware resources among a large number of mutually distrusting users
19	<b>Discuss on the support of middleware for virtualization. BTL 5</b> ➤ Library-level virtualization is also known as user-level Application Binary Interface (ABI) or API emulation. ➤ This type of virtualization can create execution environments for running alien programs on a platform rather than creating a VM to run the entire operating system.
20	<b>Compare Full virtualization and Para virtualization. BTL 4</b> Para virtualization: Virtualization in which the guest operating system 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. Full virtualization: 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.
21	<b>Define Cloud services with example. BTL2</b> ➤ Any web-based application or service offered via cloud computing is called a cloud. ➤ Cloud services can include anything from calendar and contact applications to word processing and presentations.
22	<b>Explain cloud provider and cloud broker. BTL1</b> Cloud Provider: Is a company that offers some component of cloud computing typically infrastructure as a service, software as a Service or Platform as a Service. It

	<p>is something referred as CSP.</p> <p><b>Cloud Broker:</b> It is a third party individual or business that act as an intermediary between the purchase of cloud computing service and sellers of that service.</p>
23	<p><b>Define anything-as-a-service.</b> BTL1</p> <p>Providing services to the client on the basis on meeting their demands at some pay per use cost such as data storage as a service, network as a service, communication as a service etc. It is generally denoted as anything as a service (XaaS).</p>
24	<p><b>List the types of hypervisor.</b> BTL2</p> <p>There are two types of hypervisors:</p> <ul style="list-style-type: none"> <li>➤ Type 1 (bare-metal)</li> <li>➤ Type 2 (hosted)</li> </ul> <p>Type 1: Hypervisors run directly on the system hardware. They are often referred to as a "native" or "bare metal" or "embedded" hypervisors in vendor literature.</p> <p>Type 2: Hypervisors run on a host operating system. When the virtualization movement first began to take off, Type 2 hypervisors were most popular.</p> <p>Administrators could buy the software and install it on a server they already had.</p>
25	<p><b>What is the working principle of Cloud Computing? April/May 2017</b> BTL1</p> <ul style="list-style-type: none"> <li>➤ The cloud is a collection of computers and servers that are publicly accessible via the This hardware is typically owned and operated by a third party on a consolidated basis in one or more data center locations.</li> <li>➤ The machines can run any combination of operating systems.</li> </ul>
<b>PART B</b>	
1	<p><b>i)Examine in detail about public private and hybrid cloud. (8M) (Nov/Dec 2016)</b> BTL 1</p> <p><b>Answer:Page:3-30 - Bhushan</b></p> <p><b>ii)Examine in detail about data center networking Structure (5M) (April/May 2017)</b> BTL 1</p> <p><b>Answer:Page: 3-5 Bhushan</b></p> <p><b>i) Definition (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Public: Accessible via internet - anyone who pays E.g., Google App Engine,Amazon Web Service.</li> <li>➤ Private: Accessible via internet – members owing organization E.g., NASA cloud climate modeling.</li> </ul> <p>Hybrid: Private cloud buy computing resources.</p> <p><b>Explanation(4M)</b></p> <p>Public cloud - publicly accessible cloud environment - owned by a third party cloud provider.</p> <p>IT resources on public clouds - provisioned via cloud delivery models -offered to cloud consumers - cost – commercialized.</p> <p><b>Diagram(2M)</b></p> <p><b>ii)Explanation (3M)</b></p> <ul style="list-style-type: none"> <li>➤ Physical cluster - collection of servers (physical machines) - interconnected -physical network ( LAN)</li> <li>➤ Virtual clusters - VMs installed at distributed servers - physical clusters.</li> <li>➤ VMs in virtual cluster - interconnected logically - virtual network physical networks.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Virtual cluster - physical machines - VM hosted multiple physical clusters.</li> <li>➤ Virtual cluster boundaries - distinct boundaries.</li> </ul> <p><b>Diagram(2M)</b></p>
2	<p><b>Analyze the uses of</b></p> <p>i) <b>Infrastructure as a service (4M)</b></p> <p>ii) <b>Platform as a service (4M)</b></p> <p>iii) <b>Software as a service(5M) April/May 2017 BTL 4</b></p> <p><b>Answer:Page:3-7 Bhushan</b></p> <p><b>Infrastructure as a Service(IaaS)</b> Allows users - virtualized IT resources for computing, storage, and networking. Service - rented cloud infrastructure.</p> <p><b>Platform as a Service(PaaS)</b> Able to develop, deploy and manage - execution of applications - provisioned resources demands a cloud platform - software environment</p> <p><b>Software as a Service(SaaS):</b> Browser –initiated application software. Services and tools - offered by Paas - construction of applications and management.</p>
3	<p>i) <b>Discuss the various levels of virtualization implementation. (7M)</b> <b>April/May 2017 BTL 2</b></p> <p><b>Answer:Page:3-7 Bhushan</b></p> <p>ii) <b>Summarize the design requirements and providers of VMM. (6M) BTL 2</b></p> <p><b>Answer:Page:3-11 Bhushan</b></p> <p>i) <b>Explanation (7M)</b></p> <ul style="list-style-type: none"> <li>➤ Instruction set architecture(ISA) level</li> <li>➤ Hardware level</li> <li>➤ Operating system level</li> <li>➤ Library support level</li> <li>➤ Application level</li> </ul> <p>ii) <b>Explanation(6M)</b></p> <ul style="list-style-type: none"> <li>➤ First, VMM - provide environment for programs - identical to original machine.</li> <li>➤ Second, programs run - environment minor decreases in speed.</li> <li>➤ Third, VMM - complete control of system resources</li> </ul>
4	<p>i) <b>List the advantages and disadvantages of OS extension in virtualization. (6M)</b> <b>BTL 1</b></p> <p><b>Answer:Page:3-11 Bhushan</b></p> <p>ii) <b>Identify the support of virtualization Linux platform. (7M) BTL 1</b></p> <p><b>Answer: Page :71 Bhushan</b></p> <p>i) <b>Advantages(3M)</b></p> <ul style="list-style-type: none"> <li>➤ advantage - failover flexibility</li> </ul> <p><b>Two difficulties: (3M)</b></p>

	<ul style="list-style-type: none"> <li>➤ Application programs - parallelized use all cores - software explicitly assign tasks - very complex problem.</li> </ul> <p><b>ii) Virtualization Linux(7M)</b></p> <ul style="list-style-type: none"> <li>➤ Allows complete client control - virtualized system hardware.</li> <li>➤ Executes operations directly - hardware resources including CPUs.</li> <li>➤ Two classes of Hypervisor: Type1 and Type 2.</li> <li>➤ Hardware Compatibility.</li> </ul>
5	<p><b>i) Summarize the support of middleware and library for virtualization. (7M) BTL 2</b>  <b>Answer: Page:3-23 – Bhushan</b></p> <p><b>ii) Describe the vCUDA architecture for virtualization of general purpose GPUs (7M) BTL 2</b>  <b>Answer: page:3-11 in Bhushan</b></p> <p><b>i) Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Applications use APIs - exported user-level libraries - use lengthy system calls - systems provide well-documented APIs - interface becomes another candidate - virtualization.</li> </ul> <p><b>Diagram(3M)</b></p> <p><b>ii) Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ CUDA - programming model - library for general-purpose GPUs- high performance of GPUs - compute-intensive applications - host operating systems.</li> </ul> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Difficult to run CUDA applications - hardware-level VMs directly.</li> <li>➤ vCUDA virtualizes the CUDA library - installed on guest OSes.</li> <li>➤ CUDA applications run - guest OS - issue a call to CUDA API,</li> <li>➤ vCUDA intercepts the call - redirects - CUDA API running host OS.</li> </ul> <p><b>Diagram(1M)</b></p>
6	<p><b>i) Compose in detail about the classes of VM architecture based on the position of virtualization layer Hypervisor and Xen architecture. (8M) BTL 6</b>  <b>Answer: Page:3-16 - Bhushan</b></p> <p><b>ii) Design the implementation technology of hardware virtualization. (5M) BTL 6</b>  <b>Answer: Page:3-16 Bhushan</b></p> <p><b>i) Explanation(6M)</b></p> <ul style="list-style-type: none"> <li>➤ <b>Hypervisor</b> supports hardware-level virtualization - bare metal devices - CPU, memory, disk - network interfaces.</li> <li>➤ Hypervisor software - directly between physical hardware - OS.</li> <li>➤ Virtualization layer - VMM or hypervisor.</li> <li>➤ The hypervisor provides hyper calls - guest OSes - applications.</li> </ul> <p><b>Diagram(2M)</b></p> <p><b>ii) Hardware-level virtualization(5M) - top of bare hardware.</b></p> <ul style="list-style-type: none"> <li>➤ Generates virtual hardware environment</li> <li>➤ Process manages - hardware through virtualization.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Virtualizes a computer's resources - processors, memory, I/O devices.</li> <li>➤ Intention to upgrade - hardware utilization rate – concurrently.</li> </ul>
<b>PART C</b>	
1	<p><b>What do you mean by data center automation using virtualization? (15M)</b></p> <p><b>April/May 2017 BTL 2</b></p> <p><b>Answer:Page:3-31 - Bhushan</b></p> <p><b>List(4M)</b></p> <p><b>Virtualization for Data-Center Automation</b></p> <ul style="list-style-type: none"> <li>➤ Server Consolidation in Data Centers</li> <li>➤ Virtual Storage Management</li> <li>➤ Cloud OS for Virtualized Data Centers</li> <li>➤ Trust Management in Virtualized Data Centers</li> </ul> <p><b>Explanation(8M)</b></p> <p>Data center automation - managing and automating - workflow and processes - data center facility. Automating bulk of data center operations - management - monitoring - maintenance tasks.</p> <p><b>Diagram (3M)</b></p>
2	<p><b>Discuss how virtualization is implemented in different layers. (15M) April/May 2017 BTL 4</b></p> <p><b>Answer:Page:3-16 - Bhushan</b></p> <p><b>Explanation(8M)</b></p> <p>Virtualize - portion of a computing environment - organization seeking performance - reliability/availability – scalability – consolidation – agility - a unified management.</p> <p>Implementation Levels of Virtualization</p> <p><b>List(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Instruction level architecture</li> <li>➤ Hardware level</li> <li>➤ Operating system level</li> <li>➤ Library support level</li> <li>➤ Application level</li> </ul> <p><b>Diagram (3M)</b></p>
3	<p><b>List the cloud deployment models and give a detailed note on them. (15M) Nov/Dec 2016 BTL 1</b></p> <p><b>Answer:Page: 3-4 Bhushan</b></p> <p><b>Definition &amp;List(4M)</b></p> <p><b>Infrastructure as a Service(IaaS)</b></p> <p>Allows users - virtualized IT resources for computing, storage, and networking. Service - rented cloud infrastructure.</p> <p><b>Platform as a Service(PaaS)</b></p> <p>Able to develop, deploy and manage - execution of applications - provisioned resources demands a cloud platform - software environment</p> <p><b>Software as a Service(SaaS):</b></p>

	<p>Browser –initiated application software - Services and tools - offered by Paas - construction of applications and management.</p> <p><b>Explanation(8M)</b></p> <p>Capability provided to consumer - provider's applications running - cloud infrastructure. Applications - accessible from various client devices - thin client interface - web browser (e.g., web-based email) - program interface.</p> <p><b>Diagram (3m)</b></p>
<b>UNIT-4- PROGRAMMING MODEL</b>	
<p>Open source grid middleware packages- Globus Toolkit(GT4) Architecture, Configuration-Usage of Globus - Main Components and Programming Model- Introduction to Hadoop Framework- Map reduce, 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 interfaces, dataflow of File read &amp; File write</p>	
<b>PART A</b>	
1	<p><b>Analyze on grid software support and middleware packages.</b> BTL 4</p> <ul style="list-style-type: none"> <li>➤ The software on the Grid includes programs such as ROOT, a set of object-oriented core libraries used by all the LHC experiments; POOL, a framework that provides storage for event data.</li> <li>➤ The Grid depends on the computer and communications networks of the underlying internet, novel software allows users to access computers distributed across the network. This software is called "middleware".</li> </ul>
2	<p><b>Define condor.</b> BTL 1</p> <p>Condor is an open-source high-throughput computing software framework for coarse-grained distributed parallelization of computationally intensive tasks.</p> <p>Example: Condor-G</p>
3	<p><b>Examine the sequences of events of SGE workflow.</b> BTL 3</p> <ul style="list-style-type: none"> <li>➤ The user delegates his credentials to a delegation service.</li> <li>➤ The user submits a job request to GRAM with the delegation identifier as a parameter.</li> <li>➤ GRAM parses the request, retrieves the user proxy certificate from the delegation service, and then acts on behalf of the user.</li> <li>➤ GRAM sends a transfer request to the RFT, which applies Grid FTP to bring in the necessary files.</li> </ul>
4	<p><b>Summarize on Globus toolkit architecture.</b> BTL 2</p> <ul style="list-style-type: none"> <li>➤ The Globus Toolkit, started in 1995 with funding from DARPA, is an open middleware library for the grid computing communities.</li> <li>➤ The toolkit addresses common problems and issues related to grid resource discovery, management, communication, security, fault detection, and portability. The library includes a rich set of service implementations.</li> </ul>
5	<p><b>List the functional modules in GT4 library.</b> Nov/Dec 2016 BTL 1</p> <ul style="list-style-type: none"> <li>➤ Global Resource Allocation Manager</li> <li>➤ Communication</li> <li>➤ Grid Security Infrastructure</li> <li>➤ Monitor and Discovery Service</li> <li>➤ Health and Status</li> <li>➤ Global Access of Secondary Storage</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Grid File Transfer</li> </ul>
6	<p><b>Evaluate how data's are managed using GT4.</b> BTL 5</p> <p>Data management tools are concerned with the location, transfer, and management of distributed data.</p> <ul style="list-style-type: none"> <li>➤ GridFTP</li> <li>➤ RFT</li> <li>➤ RLS</li> </ul>
7	<p><b>Define Globus client Interaction.</b> BTL 1</p> <p>There are strong interactions between provider programs and user code. GT4 makes heavy use of industry-standard web service protocols and mechanisms in service Description, discovery, access, authentication, authorization, and the like. GT4 makes extensive use of Java, C, and Python to write user code.</p>
8	<p><b>Analyze the need of MDS services in distributed system.</b> BTL 4</p> <ul style="list-style-type: none"> <li>➤ The Monitoring and Discovery System (MDS) is a suite of web services to monitor and determine resources and services on Grids.</li> <li>➤ Allows users to discover what resources are considered part of a Virtual Organization.</li> <li>➤ It offers trigger and indexing services.</li> </ul>
9	<p><b>Illustrate the building blocks in GT4 library.</b> BTL 3</p> <ul style="list-style-type: none"> <li>➤ The GT4 Library GT4 offers the middle-level core services in grid applications.</li> <li>➤ The high-level services and tools, such as MPI, Condor-G, and Nirod/G, are developed by third parties for general purpose distributed computing applications.</li> <li>➤ The local services, such as LSF, TCP, Linux, and Condor, are at the bottom level and are fundamental tools supplied by other developers.</li> </ul>
10	<p><b>List the security measures in grid.</b> BTL 1</p> <p>The necessary security measures are in terms of authentication, authorization, resource protection, secure communication, data confidentiality, data integrity, policy management and network security.</p>
11	<p><b>Evaluate why is a Block in HDFS so large?</b> BTL 5</p> <ul style="list-style-type: none"> <li>➤ HDFS blocks are large compared to disk blocks, and the reason is to minimize the cost of seeks.</li> <li>➤ If the block is large enough, the time it takes to transfer the data from the disk can be significantly longer than the time to seek to the start of the block.</li> </ul>
12	<p><b>Differentiate name node with data node in hadoop file system.</b> BTL 2</p> <ul style="list-style-type: none"> <li>➤ An HDFS cluster has two types of node operating in a master-worker pattern: a <b>name node</b> (the master) and a number of <b>data nodes</b> (workers).</li> <li>➤ The name node manages the file system namespace. It maintains the file system tree and the metadata for all the files and directories in the tree.</li> <li>➤ This information is stored persistently on the local disk in the form of two files:</li> <li>➤ the namespace image and the edit log.</li> <li>➤ The name node also knows the data nodes on which all the blocks for a given file are located.</li> </ul>
13	<p><b>Interpret how file permission is achieved in HDFS.</b> BTL 2</p>

	<p><b>For each file system, Hadoop uses a different URI scheme for the file system instance in order to connect with it. For example, you list the files in the local system by using the file URI scheme, as shown here:</b></p> <pre>\$ hdfs dfs -ls file:///</pre>
14	<p><b>Generalize how a name node is not able to serve a request.</b> BTL 6</p> <p>The namenode is still a single point of failure (SPOF), since if it did fail, all clients—including MapReduce jobs—would be unable to read, write, or list files, because the namenode is the sole repository of the metadata and the file-to-block mapping.</p>
15	<p><b>Analyze how a standby takes over when an active name node is failed.</b> BTL 4</p> <p>When a standby name node comes up it reads up to the end of the shared edit log to synchronize its state with the active name node, and then continues to read new entries as they are written by the active name node.</p>
16	<p><b>Define failover and fencing.</b> BTL 1</p> <ul style="list-style-type: none"> <li>➤ The transition from the active name node to the standby is managed by a new entity in the system called the failover controller</li> <li>➤ The HA implementation goes to great lengths to ensure that the previously active name node is prevented from doing any damage and causing corruption—a method known as fencing</li> </ul>
17	<p><b>Generalize how an anatomy of File read is done.</b> BTL 6</p> <ul style="list-style-type: none"> <li>➤ The client opens the file it wishes to read by calling open () on the File System object, which for HDFS is an instance of Distributed File System.</li> <li>➤ Distributed File System calls the name node, using RPC, to determine the locations of the blocks for the first few blocks in the file.</li> <li>➤ The name node returns the addresses of the data nodes that have a copy of that block.</li> </ul>
18	<p><b>Discuss how a data is read from hadoop URL.</b> BTL 2</p> <p>One of the simplest ways to read a file from a Hadoop files system is by using a java.net.URL object to open a stream to read the data from.</p> <p>The general idiom is:</p> <pre>InputStream in = null; try {     in = new URL("hdfs://host/path").openStream();     // process in } finally {     IOUtils.closeStream(in); }</pre>
19	<p><b>Name the details of file querying system.</b> BTL 1</p> <p>Files and directories are like standard SQL tables</p> <p>The following example shows a query on a file system database in a Hadoop distributed file system.</p>

	SELECT * FROM hdfs.logs.`AppServerLogs/20104/Jan/01/part0001.txt`;
20	<b>Demonstrate how does the name node choose which data nodes to store replicas on?</b> <b>BTL 3</b> <ul style="list-style-type: none"> <li>➤ The replica placement strategy is that if the writer is on a data node,</li> <li>➤ the 1st replica is placed on the local machine, otherwise a random data node.</li> <li>➤ The 2nd replica is placed on a data node that is on a different rack.</li> <li>➤ The 3rd replica is placed on a data node which is on the same rack as the first replica.</li> </ul>
21	<b>What are the available input formats?</b> <b>BTL 1</b> <ul style="list-style-type: none"> <li>➤ Key Value Text Input Format</li> <li>➤ Text Input Formant</li> <li>➤ NLine Input Format</li> <li>➤ Multi File Input Format</li> <li>➤ Sequence FILE Input Format</li> </ul>
22	<b>Define Block.</b> <b>BTL 1</b> <ul style="list-style-type: none"> <li>➤ A disk has a block size, which is the minimum amount of data that it can read or write.</li> <li>➤ File systems for a single disk build on this by dealing with data in blocks, which are an integral multiple of the disk block size. File system blocks are typically a few kilobytes in size.</li> </ul>
23	<b>List two types of nodes that control the job execution process.</b> <b>BTL 2</b> A job tracker and a number of task trackers controls the job execution process.
24	<b>What is meant by FUSE?</b> <b>BTL 1</b> <ul style="list-style-type: none"> <li>➤ File system in User space (FUSE) allows file systems that are implemented in user space to be integrated as a Unix file system.</li> <li>➤ Hadoop's Fuse-DFS contribute module allows any Hadoop file system (but typically HDFS) to be mounted as a standard file system.</li> </ul>
25	<b>Define the term Globus Container.</b> <b>BTL 1</b> The Globus Container provides a basic runtime environment for hosting the web services needed to execute grid jobs.
<b>PART B</b>	
1	<b>Describe the relative strength and limitation of open source grid middleware packages. (13M)</b> <b>BTL 1</b> <b>Answer:</b> Page:4-2 - Bhushan <b>Advantages(7M)</b> <ul style="list-style-type: none"> <li>➤ Computational resources</li> <li>➤ Storage resources</li> <li>➤ Network resources</li> <li>➤ Scientific instruments</li> </ul> <b>Disadvantages(6M)</b> <ul style="list-style-type: none"> <li>➤ Access to resources</li> <li>➤ Computing ability</li> </ul>
2	<b>i)List the features in condor kernel and condor G for grid computing. (7M)</b> <b>BTL 1</b> <b>Answer:</b> Page:4-7 - Bhushan

	<p><b>Definition (2M)</b></p> <p><b>Explanation (2M)</b></p> <p>High-throughput computing –</p> <ul style="list-style-type: none"> <li>➤ Large amounts of fault-tolerant computational power</li> <li>➤ Effective utilization of resource</li> </ul> <p>Opportunistic computing</p> <ul style="list-style-type: none"> <li>➤ Use resource whenever available</li> </ul> <p><b>Condor G (3M)</b></p> <p>Preserve local execution environment -Condor can transfer files</p> <ul style="list-style-type: none"> <li>➤ Automatically send back changed files</li> <li>➤ Atomic transfer multiple files</li> <li>➤ Can encrypted over the wire</li> <li>➤ Remote I/O Socket</li> <li>➤ Standard Universe - use remote system calls</li> </ul> <p><b>ii)Describe sun grid engine middleware package in detail. (6M) BTL 1</b></p> <p><b>Answer:Page: 71 -Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ System - offers centralized management -resources allocated to individual jobs.</li> <li>➤ Enhance efficiency – performance – suspend and resume tools - allow users to halt job - restart without losing work.</li> </ul> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Accepts jobs from users.</li> <li>➤ Places jobs in computer area.</li> <li>➤ Sends jobs holding area- host executed.</li> <li>➤ Manages jobs during execution.</li> <li>➤ Logs record of execution.</li> </ul>
3	<p><b>i)Summarize the grid standards and APIs. (8M) BTL 2</b></p> <p><b>Answer: Page:4-6 - Bhushan</b></p> <p><b>Explanation(4M)</b></p> <p>Open Grid Forum (formally Global Grid Forum) - Object Management Group - well-formed organizations behind standards.</p> <p><b>List(2M)</b></p> <ul style="list-style-type: none"> <li>➤ SAGA (Simple API for Grid Applications),</li> <li>➤ GSI (Grid Security Infrastructure),</li> <li>➤ OGSI (Open Grid Service Infrastructure),</li> <li>➤ WSRE (Web Service Resource Framework).</li> </ul> <p><b>Diagram (2M)</b></p> <p><b>ii)Discuss on grid software support and middleware package. (5M) BTL 2</b></p> <p><b>Answer:Page:4-2 - Bhushan</b></p> <p><b>Explanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Grid middleware - layer between hardware and software.</li> <li>➤ Middleware products - enable sharing - heterogeneous resources - managing virtual organizations - created around the grid.</li> <li>➤ Middleware - glues allocated resources - specific user applications.</li> </ul>

4	<p><b>i) Illustrate Globus tool kit architecture in detail. (8M) (Nov/Dec 2016) BTL 3</b></p> <p><b>Answer:Page:4-6 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Open source software libraries - support operational grids - applications on international basis.</li> <li>➤ Toolkit addresses – issues like grid resource discovery – management – communication – security - fault detection – portability - software provides variety of components - capabilities.</li> </ul> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Library - rich set of service implementations.</li> <li>➤ Implemented software - supports grid infrastructure management - provides tools - building new web services in Java, C, and Python.</li> </ul> <p><b>Diagram (2M)</b></p> <p><b>ii) Classify the functional modules in GT4 library. (5M) BTL 3</b></p> <p><b>Answer:Page:4 - 7 - Bhushan</b></p> <p><b>List(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Global Resource Allocation Manager (GRAM)</li> <li>➤ Grid Resource Access and Management (HTTP)</li> <li>➤ Communication (Nexus) - Unicast - multicast communication</li> <li>➤ Grid Security Infrastructure (GSI) - Authentication - security services</li> <li>➤ Monitoring and Discovery Service(MDS) – Distributed - access structure and state information.</li> <li>➤ Health and Status (HBM) - Heartbeat monitoring of system components.</li> <li>➤ Global Access of Secondary Storage (GASS) - Grid access data - remote secondary storage.</li> <li>➤ Grid File Transfer Grid (FTP) - Inter-node fast file transfer.</li> </ul>
5	<p><b>i) Explain the concepts involved in resource management using GRAM. (7M) BTL 4</b></p> <p><b>Answer:Page:80 - Bhushan</b></p> <p><b>Explanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Global Resource Allocation Manager (GRAM)</li> <li>➤ Grid Resource Access and Management (HTTP)</li> <li>➤ GRAM module supports web services – initiating – monitoring - managing execution - computational jobs on remote computers.</li> <li>➤ GRAM - built local resource allocation services.</li> </ul> <p><b>Diagram(2M)</b></p> <p><b>ii) Classify the GT4 tools used by data management. (6M) BTL 4</b></p> <p><b>Answer:Page:4-6 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ GT4 tools - used individually - conjunction with other tools - develop interesting solutions - efficient data access.</li> <li>➤ <b>GridFTP</b> supports reliable – secure - fast memory-to-memory - disk-to-disk data movement - over high-bandwidth WANs.</li> </ul> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ <b>RFT</b> - provides reliable management - multiple Grid FTP transfers</li> <li>➤ <b>RLS</b> (Replica Location Service) - scalable system for maintaining - providing access to information - location of replicated files - data sets.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ <b>OGSA-DAI</b> (Globus Data Access and Integration) - tools developed by UK e-Science program - provide access - relational XML databases.</li> </ul>
6	<p><b>i)Evaluate the interaction in the functional module client globus job work flow. (8M)</b> <b>BTL 5</b></p> <p><b>Answer:Page:4-12 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ GRAM parses request - retrieves user proxy certificate - delegation service.</li> <li>➤ GRAM sends transfer request - RFT (Reliable File Transfer) - applies Grid FTP - bring necessary files.</li> </ul> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Typical job execution sequence - user delegates his credentials - delegation service. User submits job request - GRAM - delegation identifier as parameter.</li> </ul> <p><b>Diagram(2M)</b></p> <p><b>ii)Summarize the functional components in CGSP library. (5M) BTL 5</b></p> <p><b>Answer: Page:4-8 - Bhushan</b></p> <p><b>Listing Components(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Service container</li> <li>➤ Security manager</li> <li>➤ Information center</li> <li>➤ Data manager</li> <li>➤ Execution manager</li> <li>➤ Domain manager</li> <li>➤ Grid monitor</li> <li>➤ Portal</li> </ul>
7	<p><b>i)Generalize the functional components of china grid support platform library. (7M)</b> <b>BTL 6</b></p> <p><b>Answer: Page:4-16 - Bhushan</b></p> <p><b>List (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Service container</li> <li>➤ Security manager</li> <li>➤ Information center</li> <li>➤ Data manager</li> </ul> <p><b>Explanation(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Execution manager</li> <li>➤ Domain manager</li> <li>➤ Grid monitor</li> <li>➤ Portal</li> </ul> <p><b>ii)Design the functional building blocks in the CGSP library that represents the job executional flow. (6M) BTL 6</b></p> <p><b>Answer: Page:4-16 - Bhushan</b></p> <p><b>Explanation(4M)</b></p> <ul style="list-style-type: none"> <li>➤ ChinaGrid - constructing public grid service system - research and higher education in China.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ CGSP - integrates sorts of heterogeneous resources - educational and research resources - distributed over CERNET in China.</li> </ul> <p><b>Diagram(2M)</b></p>
	<b>PART C</b>
1	<p><b>Draw and Explain the Global Toolkit architecture (15M) (Nov/Dec 2016) BTL 2</b></p> <p><b>Answer: Page:4-6 - Bhushan</b></p> <p><b>Definition (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Globus Toolkit - open middleware library - grid computing communities.</li> <li>➤ Open source software libraries - support many operational grids - applications on international basis.</li> </ul> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Toolkit addresses common problems - issues related to grid resource discovery-management – communication – security - fault detection - portability.</li> <li>➤ Software provides variety of components – capabilities - rich set of service implementations.</li> </ul> <p><b>Diagram(3M)</b></p>
2	<p><b>Give a detailed note on Hadoop framework. (15M) (Nov/Dec 2016) BTL 4</b></p> <p><b>Answer: Page:4-19 - Bhushan</b></p> <p><b>Definition (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Hadoop - Apache Software Foundation top-level project - holds various Hadoop subprojects - graduated from Apache Incubator.</li> </ul> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Hadoop handles processing details - leaving developers free focus on application logic.</li> <li>➤ Hadoop project - provides supports development - open source software - supplies a framework - development of highly scalable - distributed computing applications.</li> </ul> <p><b>Diagram(3M)</b></p>
3	<p><b>Discuss MAPREDUCE with suitable diagrams. (15M) (April/May 2017) BTL 3</b></p> <p><b>Answer: Page:4-22 - Bhushan</b></p> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Web programming model - scalable data processing on large clusters - over large data sets.</li> <li>➤ Model - applied web-scale search - cloud computing applications.</li> </ul> <p><b>Diagram(3M)</b></p>
4	<p><b>Elaborate HDFS concepts with suitable Illustrations. (15M) (April/May2017) BTL 4</b></p> <p><b>Answer: Page:4-29 - Bhushan</b></p> <p><b>Definition (2M)</b></p> <p>Hadoop Distributed File System(HDFS) – Map Reduce environment – provides user with sophisticated framework – manage execution of map – reduce tasks – across cluster of machines.</p> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Location(s) - distributed file system of job input</li> <li>➤ Location(s) - distributed file system for job output</li> <li>➤ Input format</li> <li>➤ Output format</li> <li>➤ Class containing map function</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Class containing reduce function</li> <li>➤ JAR file(s) containing map - reduce functions - support classes</li> </ul> <p><b>Diagram(3M)</b></p>
--	---



2	<b>Define IDS.</b> BTL 1 An intrusion detection system (IDS) is a type of security software designed to automatically alert administrators when someone or something is trying to compromise information system through malicious activities or through security policy violations.
3	<b>Summarize on reputation trust model.</b> BTL 2 Reputation-based trust model and techniques are used for securing P2P and social networks could be merged to defend data centers and cloud platforms against attacks from the open network.
4	<b>List various trust models.</b> BTL 1 <ul style="list-style-type: none"> <li>➤ A Generalized Trust Model</li> <li>➤ Reputation-Based Trust Model</li> <li>➤ A Fuzzy-Trust Model</li> </ul>
5	<b>Relate Active and Passive attacks.</b> BTL 3 <ul style="list-style-type: none"> <li>➤ Passive attacks steal sensitive data or passwords.</li> <li>➤ Active attacks manipulate kernel data structures which will cause major damage to cloud servers.</li> </ul>
6	<b>Evaluate the authorization model of grid security.</b> BTL 5 <ul style="list-style-type: none"> <li>➤ The subject-push model</li> <li>➤ The resource-pulling model</li> <li>➤ The authorization agent model</li> </ul>
7	<b>Define Authentication.</b> BTL 1 The process of identifying an individual, usually based on a username and password. In security systems, authentication is distinct from authorization , which is the process of giving individuals access to system objects based on their identity.
8	<b>Formulate the categories of authorization for access control.</b> BTL 6 <ul style="list-style-type: none"> <li>➤ attribute authorities</li> <li>➤ policy authorities</li> <li>➤ identity authorities</li> </ul>
9	<b>Discuss on GSI.</b> BTL 2 <ul style="list-style-type: none"> <li>➤ The major authentication methods in the grid include passwords, PKI, and Kerberos.</li> <li>➤ The password is the simplest method to identify users, but the most vulnerable one to use.</li> <li>➤ The PKI is the most popular method supported by GSI.</li> </ul>
10	<b>Differentiate transport level security and message level security Nov/Dec 2016</b> BTL 4 Transport Level security: means providing security at the transport layer itself. When dealing with security at Transport level, we are concerned about integrity, privacy and authentication of message as it travels along the physical wire. Message Level Security: For Transport level security, we actually ensure the transport that is being used should be secured but in message level security, we actually secure the message. We encrypt the message before transporting it.
11	<b>Compose the primary pieces of information of a certificate in GSI authentication.</b> BTL 6

	<ul style="list-style-type: none"> <li>➤ A subject name, which identifies the person or object that the certificate represents</li> <li>➤ The public key belonging to the subject</li> <li>➤ The identity of a CA that has signed the certificate to certify that the public key and the identity both belong to the subject</li> <li>➤ The digital signature of the named CA. X.509 provides each entity with a unique identifier.</li> </ul>
12	<p><b>How will you measure the mutual authentication between two parties? BTL 5</b></p> <p>Mutual authentication is machine-to-machine, when the remote authentication fails Non-technical mutual-authentication also exists to mitigate this problem, requiring the user to complete a challenge, effectively forcing them to notice, and blocking them from authenticating with a false endpoint. Mutual authentication is two types:</p> <ul style="list-style-type: none"> <li>➤ Certificate based</li> <li>➤ User name-password based</li> </ul>
13	<p><b>Define aspects of Data Security. BTL 3</b></p> <p>Security for</p> <ul style="list-style-type: none"> <li>➤ Data in transit</li> <li>➤ Data at rest</li> <li>➤ Processing of data including multitenancy</li> <li>➤ Data Lineage</li> <li>➤ Data Provenance</li> </ul>
14	<p><b>Discuss the risk factors of network level of cloud infrastructure. BTL 2</b></p> <ul style="list-style-type: none"> <li>➤ Ensuring data confidentiality and integrity of the organizations data in transit to and from the public cloud provider</li> <li>➤ Ensuring proper access control (Authentication, Authorization, Auditing) to resources in the public cloud</li> <li>➤ Ensuring availability of the Internet facing resources of the public cloud used by the organization.</li> <li>➤ Replacing the established network zones and tiers with domains</li> </ul>
15	<p><b>Explain the security levels at Host Level. BTL 1</b></p> <p>Host security at PaaS and SaaS Level</p> <ul style="list-style-type: none"> <li>➤ Both the PaaS and SaaS hide the host operating system from end users</li> <li>➤ Host security responsibilities in SaaS and PaaS are transferred to CSP.</li> </ul>
16	<p><b>Compare SaaS and PaaS Application security. BTL 4</b></p> <p>Application security at the SaaS level</p> <ul style="list-style-type: none"> <li>➤ SaaS Providers are responsible for providing application security</li> <li>➤ Application security at the PaaS level</li> <li>➤ Security of the PaaS Platform</li> <li>➤ Security of the customer applications deployed on a PaaS platform.</li> </ul>
17	<p><b>Show how will you categorize host security in IaaS? BTL 3</b></p> <p>Host security at IaaS Level</p> <ul style="list-style-type: none"> <li>➤ Virtualization software security</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Hypervisor security</li> <li>➤ Threats: Blue Pill attack on the hypervisor.</li> </ul>
18	<p><b>Identify the security Challenges in VMs.</b> BTL 1</p> <p>Buffer overflows, DoS attacks, spyware, malware, rootkits, Trojan horses, and worms. In a cloud environment, newer attacks may result from hypervisor malware, guest hopping and hijacking, or VM rootkits, the man-in-the-middle attack for VM migrations.</p>
19	<p><b>List out the categories in PaaS application security.</b> BTL 1</p> <p>Application security at the PaaS level</p> <ul style="list-style-type: none"> <li>➤ Security of the PaaS Platform</li> <li>➤ Security of the customer applications deployed on a PaaS platform</li> </ul>
20	<p><b>Point out security issues in cloud.</b> BTL 4</p> <ul style="list-style-type: none"> <li>➤ Loss of Control</li> <li>➤ Lack of trust</li> <li>➤ Multi-tenancy</li> </ul>
21	<p><b>What is password based authentication?</b> BTL 2</p> <p>The authentication is a process of checking authenticity of entities using different methods like password, public key infrastructure(PKI) or Kerberos</p>
22	<p><b>Explain the common attacks happen at host level.</b> BTL 1</p> <ul style="list-style-type: none"> <li>➤ Hijacking of accounts</li> <li>➤ Stealing the keys like SSH private keys</li> <li>➤ Attacking unpatched and vulnerable services</li> <li>➤ Attcking systems that are not secured</li> <li>➤ Deploying Trojans embedded viruses in the software</li> </ul>
23	<p><b>What do you mean by Identity and access management in cloud?</b> Nov/Dec 2016</p> <p>BTL 2</p> <p>It is the security framework composed of policy and governance components used for creation ,maintanence and termination of digital indentities with controlled access of shared resources</p>
24	<p><b>Define GSI.</b> BTL1</p> <p>The Grid Security Infrastructure is a part of Globus toolkit that provides fundamental security services and standards to support grid including authentication,authorization and delegation</p>
25	<p><b>List two types of Network level attacks.</b> BTL1</p> <ul style="list-style-type: none"> <li>➤ Eavesdropping</li> <li>➤ Port Scanning</li> <li>➤</li> </ul>
	<b>PART B</b>
1	<p><b>Examine in detail about trust model for grid security enforcement (13M) (April/May 2017)</b> BTL 3</p> <p><b>Answer: Page - 5-2 - Bhushan</b></p> <p><b>Explanation(8M)</b></p> <ul style="list-style-type: none"> <li>➤ Potential security issues- occur in grid environment- qualified security mechanisms.</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Issues include- network sniffers- out-of-control access-faulty operation-malicious operation-integration- local security mechanisms-delegation-dynamic resources and services-attack provenance</li> <li>➤ A Generalized Trust Model</li> <li>➤ A Reputation-Based Trust Model</li> <li>➤ A Fuzzy-Trust Model</li> </ul> <p><b>Diagram(5M)</b></p>
2	<p><b>i) Define Authentication and Summarize on three authorization models of GSI. (8M)</b> BTL1</p> <p><b>Answer: Page - 5-4,5-5 - Bhushan</b></p> <p><b>Definition (2M)</b></p> <ul style="list-style-type: none"> <li>➤ Authentication methods- grid include passwords-PKI- Kerberos.</li> <li>➤ Password - method to identify users – but vulnerable to use.</li> </ul> <p><b>Explanation models(4M)</b></p> <ul style="list-style-type: none"> <li>➤ PKI - most popular- method supported by GSI.</li> <li>➤ Authorization Access Control</li> <li>➤ Three Authorization Models</li> </ul> <p><b>Diagram(2M)</b></p> <p><b>ii) Discuss on the trust delegation operations using proxy credentials in GSI. (5M)</b> BTL 1</p> <p><b>Answer: Page - 5-7 - Bhushan</b></p> <p><b>Credentials(5M)</b></p> <ul style="list-style-type: none"> <li>➤ To reduce - avoid number of times - user enter his passphrase -several grids are used - have agents (local or remote)- request services –on behalf of user- GSI provides - delegation capability -delegation service - provides an interface - allow clients to delegate (and renew) - X.509 proxy certificates service.</li> <li>➤ Proxy - a new certificate and private key- Key pair used for proxy- public key embedded in certificate - private key- either be regenerated -for each proxy</li> </ul>
3	<p><b>i) Define GSI and describe in detail about GSI functional layers.GT4 provides distinct WS and pre-WS authentication and authorization capabilities. (8M)</b> BTL 1</p> <p><b>Answer: Page - 5-7 - Bhushan</b></p> <p><b>Explanation(4M)</b></p> <p><b>Transport-Level Security(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Transport-level security entails - SOAP messages -conveyed over a network - connection protected by TLS.</li> </ul> <p><b>Message-Level Security(2M)</b></p> <ul style="list-style-type: none"> <li>➤ GSI provides- message-level security - message protection for SOAP messages - implements WS-Security standard -WS-Secure Conversation specification</li> </ul> <p><b>Diagram(4M)</b></p> <p><b>ii) Examine in detail about multiple handshaking in mutual authentication scheme. (5M)</b> BTL 1</p> <p><b>Answer: Page - 5-9 - Bhushan</b></p> <p><b>Authentication Scheme(5M)</b></p> <ul style="list-style-type: none"> <li>➤ Mutual authentication - process by two parties - certificates signed by CA -prove to each other -based on the certificate -trust of the CAs - signed each other's certificates</li> </ul>

	<ul style="list-style-type: none"> <li>➤ GSI uses - Secure Sockets Layer (SSL) - mutual authentication protocol</li> <li>➤ Mutually authenticate- the first person (Alice) - establishes a connection-second person (Bob) - start authentication process.</li> </ul>
4	<p><b>i) Demonstrate the infrastructure security: Network level in cloud. (7M) (Nov/Dec 2016) BTL 3</b></p> <p><b>Answer: Page - 5-9 - Bhushan</b></p> <p><b>Network level in cloud(7M)</b></p> <ul style="list-style-type: none"> <li>➤ Ensuring data confidentiality - integrity of organizations data - transit to and from - public cloud provider</li> <li>➤ Ensuring proper access control (Authentication, Authorization, Auditing)- resources in public cloud</li> <li>➤ Ensuring availability-Internet facing resource - public cloud used by organization</li> <li>➤ Replacing - established network zones - tiers with domains</li> </ul> <p><b>ii) Classify the Key privacy issues in the cloud. (6M) BTL 3</b></p> <p><b>Answer: Page - 5-9 - Bhushan</b></p> <ul style="list-style-type: none"> <li>➤ Loss of Control (2M)</li> <li>➤ Lack of trust(2M)</li> <li>➤ Multi-tenancy(2M)</li> </ul>
5	<p><b>i) Analyze the infrastructure security of cloud at host level (7M) (Nov/Dec 2016)</b></p> <p><b>BTL 4</b></p> <p><b>Answer: Page - 5-9- Bhushan</b></p> <p><b>Host security at PaaS and SaaS Level (4M)</b></p> <ul style="list-style-type: none"> <li>➤ PaaS and SaaS - hides host operating system - end users</li> <li>➤ Host security responsibilities- SaaS and PaaS transferred to CSP</li> </ul> <p><b>Host security at IaaS Level(3M)</b></p> <ul style="list-style-type: none"> <li>➤ Virtualization software security</li> <li>➤ Hypervisor security</li> <li>➤ Threats: Blue Pill attack on hypervisor</li> </ul> <p><b>ii) Explain in detail about virtual server security of cloud. (6M) BTL 4</b></p> <p><b>Answer: Page - 5-10 - Bhushan</b></p> <p><b>Cloud Security(6M)</b></p> <p>Customer guest OS - virtual server security</p> <ul style="list-style-type: none"> <li>➤ Attacks the guest OS: e.g., stealing keys used - access and manage hosts</li> </ul>
6	<p><b>Explain in detail about application level security in i)SaaS ii)PaaS iii)IaaS (13M)</b></p> <p><b>BTL 4</b></p> <p><b>Answer: Page - 5-12 - Bhushan</b></p> <p><b>Explanation(12M)</b></p> <p><b>Application security at the SaaS level(4M)</b></p> <ul style="list-style-type: none"> <li>➤ SaaS Providers - responsible for providing - application security</li> </ul> <p><b>Application security at the PaaS level(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Security of PaaS Platform</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Security of customer applications - deployed on PaaS platform</li> </ul> <p><b>Application security at the IaaS Level(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Customer applications –treated as black box</li> <li>➤ IaaS - not responsible for application level security</li> </ul> <p><b>Diagram(1M)</b></p>
7	<p>i) Compose in detail about the aspects of data security (13M) BTL 6</p> <p><b>Answer: Page - 5-14 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Data security - weakest link- cloud models- new cloud security standards- apply common API tools - cope with data lock-in problem - network attacks or abuses.</li> </ul> <p><b>Explanation(8M)</b></p> <ul style="list-style-type: none"> <li>➤ IaaS model - Amazon - sensitive to external attacks.</li> <li>➤ Role-based interface tools - alleviate complexity of provisioning system.</li> <li>➤ Distributed Defense - against DDoS Flooding Attacks</li> <li>➤ Data and Software Protection Techniques</li> <li>➤ Data Integrity and Privacy Protection</li> <li>➤ Data Coloring and Cloud Watermarking</li> <li>➤ Data Lock-in Problem and Proactive Solutions</li> <li>➤ Reputation-Guided Protection of Data Centers</li> </ul> <p><b>Diagram(3M)</b></p>
	<b>PART C</b>
1	<p><b>Explain Trust models for Grid security Environment (15M) (Nov/Dec 2016) BTL 2</b></p> <p><b>Answer: Page - 5-2 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <p><b>Diagram(3M)</b></p> <ul style="list-style-type: none"> <li>➤ Potential security issues -occur in grid environment - qualified security mechanisms.</li> </ul> <p><b>Explanation(10M)</b></p> <ul style="list-style-type: none"> <li>➤ Issues include- network sniffers- out-of-control access-faulty operation-malicious operation-integration of local security mechanisms-delegation-dynamic resources and services-attack provenance</li> <li>➤ A Generalized Trust Model</li> <li>➤ Reputation-Based Trust Model</li> <li>➤ A Fuzzy-Trust Model</li> </ul>
2	<p><b>Write in detail about cloud security infrastructure (15M) (Nov/Dec 2016) BTL 1</b></p> <p><b>Answer: Page- 5-9 - Bhushan</b></p> <p><b>Definition(2M)</b></p> <p>Lacking trust - service providers and cloud users - hindered universal acceptance - cloud computing - service on demand-trust models -developed to protect - e-commerce and online shopping - provided by eBay and Amazon</p> <p><b>Cloud Security Defense Strategies(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Basic Cloud Security</li> <li>➤ Security Challenges in VMs</li> <li>➤ Cloud Defense Methods</li> <li>➤ Defense with Virtualization</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Privacy and Copyright Protection</li> </ul> <p><b>Distributed Intrusion/Anomaly Detection</b></p> <ul style="list-style-type: none"> <li>➤ Distributed Defense -DDoS Flooding Attacks</li> </ul> <p><b>Data and Software Protection Techniques(4M)</b></p> <ul style="list-style-type: none"> <li>➤ Data Integrity and Privacy Protection</li> <li>➤ Data Coloring and Cloud Watermarking</li> <li>➤ Data Lock-in Problem and Proactive Solutions</li> </ul> <p><b>Reputation-Guided Protection of Data Centers(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Reputation System Design Options</li> <li>➤ Reputation Systems for Clouds</li> <li>➤ Trust Overlay Networks</li> </ul> <p><b>Diagram(3M)</b></p>
3	<p><b>Write a detailed note on Identity and access management architecture (15M) (April/May 2017) BTL 1</b></p> <p><b>Answer: Page – 140 -Notes</b></p> <p><b>Definition(2M)</b></p> <ul style="list-style-type: none"> <li>➤ Improves operational efficiency - regulatory compliance management</li> <li>➤ IAM - organizations to achieve access control- operational security</li> </ul> <p><b>Explanation(10m)</b></p> <ul style="list-style-type: none"> <li>➤ Cloud use cases - need IAM</li> <li>➤ Organization employees accessing- SaaS provide identity federation</li> <li>➤ IT admin access CSP management console - provision resources - access users - use corporate identity</li> <li>➤ Developers creating accounts - partner users in PaaS</li> <li>➤ End users access storage service in cloud</li> <li>➤ Applications - residing in cloud service provider - access storage from another cloud service</li> </ul> <p><b>Diagram(3M)</b></p>

**IT6004****SOFTWARE TESTING****LTPC  
3003****UNITI-INTRODUCTION**

Testing as an Engineering Activity – Testing as a Process – Testing axioms – Basic definitions – Software Testing Principles – The Tester’s Role in a Software Development Organization – Origins of Defects – Cost of defects – Defect Classes – The Defect Repository and Test Design – Defect Examples – Developer/Tester Support of Developing a Defect Repository – Defect Prevention strategies.

**UNITII-TESTCASEDESIGN**

Test case Design Strategies – Using Black Box Approach to Test Case Design – Random Testing – Requirements based testing – Boundary Value Analysis – Equivalence Class Partitioning – State-based testing – Cause-effect graphing – Compatibility testing – user documentation testing – domain testing – Using White Box Approach to Test design – Test Adequacy Criteria – static testing vs. structural testing – code functional testing – Coverage and Control Flow Graphs – Covering Code Logic – Paths – code complexity testing – Evaluating Test Adequacy Criteria.

**UNITIII-LEVELSOFTESTING**

The need for Levers of Testing – Unit Test – Unit Test Planning – Designing the Unit Tests – The Test Harness – Running the Unit tests and Recording results – Integration tests – Designing Integration Tests – Integration Test Planning – Scenario testing – Defect bash elimination System Testing – Acceptance testing – Performance testing – Regression Testing – Internationalization testing – Ad-hoc testing – Alpha, Beta Tests – Testing OO systems – Usability and Accessibility testing – Configuration testing – Compatibility testing – Testing the documentation – Website testing.

**UNITIV-TESTAMANAGEMENT**

People and organizational issues in testing – Organization structures for testing teams – testing services – Test Planning – Test Plan Components – Test Plan Attachments – Locating Test Items – test management – test process – Reporting Test Results – The role of three groups in Test Planning and Policy Development – Introducing the test specialist – Skills needed by a test specialist –Building-TestingGroup.

**UNIT V-TESTAUTOMATION**

Software test automation – skill needed for automation – scope of automation – design and architecture for automation – requirements for a test tool – challenges in automation – Test metrics and measurements – project, progress and productivity metrics.

**Subject Code: IT6004****Year/Semester :IV/07****Subject Name : Software Testing****Subject Handler: Ms.J.Aruna Jasmine****UNIT I INTRODUCTION**

Testing as an Engineering Activity – Testing as a Process – Testing axioms – Basic definitions – Software Testing Principles – The Testers Role in a Software Development Organization – Origins of Defects – Cost of defects – Defect Classes – The Defect Repository and Test Design – Defect Examples – Developer/Tester Support of Developing a Defect Repository – Defect Prevention strategies.

**PART\* A**

<b>Q.NO</b>	<b>QUESTIONS</b>
1.	<p><b>Define Software Engineering.(BTL1)</b> Software Engineering is a discipline that produces error free software with in a time and budget.</p>
2.	<p><b>Define software Testing.(AU April/May 2016)(BTL1)</b> Testing can be described as a process used for revealing defects in software, and for establishing that the software has attained a specified degree of quality with respect to selected attributes.</p>
3.	<p><b>List the elements of the engineering disciplines.(BTL1)</b></p> <ul style="list-style-type: none"> <li>• Basic principles</li> <li>• Processes</li> <li>• Standards</li> <li>• Measurements</li> <li>• Tools</li> <li>• Methods</li> <li>• Best practices</li> <li>• Code of ethics</li> <li>• Body of knowledge</li> </ul>
4.	<p><b>Define process in the context of software quality.(BTL1)</b> “Process” in the software engineering domain, is a set of methods, practices, Standards, documents, activities, polices, and procedures that software engineers use to develop and maintain a software system and its associated artifacts, such as project and test plans, design documents, code, and manuals</p>
5.	<p><b>Define the term Testing(BTL1)</b></p> <ul style="list-style-type: none"> <li>• Testing is generally described as a group of procedures carried out to evaluate some aspect of a piece of software.</li> <li>• Testing can be described as a process used for revealing defects in software, and for</li> </ul>

	establishing that the software has attained a specified degree of quality with respect to selected attributes.
6.	<p><b>Interpret the term Debugging or fault localization.(BTL1)</b></p> <p>Debugging or fault localization is the process of</p> <ul style="list-style-type: none"> <li>• Locating the fault or defect</li> <li>• Repairing the code, and</li> <li>• Retesting the code.</li> </ul>
7.	<p><b>List the levels of TMM.(AU Nov/Dec2016)(BTL1)</b></p> <p>The testing maturity model or TMM contains five levels.</p> <p>They are</p> <ul style="list-style-type: none"> <li>• Level1: Initial</li> <li>• Level2: Phase definition</li> <li>• Level3: Integration</li> <li>• Level4: Management and Measurement</li> <li>• Level5: Optimization /Defect prevention and Quality Control.</li> </ul>
8.	<p><b>List the members of the critical groups in a testing process.</b> (BTL1)</p> <ul style="list-style-type: none"> <li>• Manager</li> <li>• Developer/Tester</li> <li>• User/Client</li> </ul>
9.	<p><b>Define Error.(BTL1)</b></p> <p>An error is mistake or misconception or misunderstanding on the part of a software developer.</p>
10.	<p><b>Define Faults (Defects).(AU Nov/Dec2016) ( BTL1)</b></p> <p>A fault is introduced into the software as the result of an error. It is an anomaly in the software that may cause it to behave incorrectly, and not according to its specification.</p>
11.	<p><b>Define failures.(AU NOV/DEC 2016) (BTL5)</b></p> <p>A failure is the inability of a software or component to perform its required functions within specified performance requirements.</p>
12	<p><b>What is the need of TMM?</b></p> <p>Test maturity model gives the level at which an organization stands in meeting Testing criterias.</p>
13	<p><b>Define Validation.(BTL1)</b></p> <p>Validation is the process of evaluating a software system or component during, or at the end of, the development cycle in order to determine whether it satisfies specified requirements.</p>
14	<p><b>Explain in short about Verification.( BTL1)</b></p> <p>Verification is the process of evaluating a software system or component to determine whether the product of a given development phase satisfy the conditions imposed at the start of that phase.</p>
15	<p><b>Programmer A and Programmer B are working on a group of interfacing modules. Programmer A tends to be a poor communicator and does not get along well with Programmer B. Due to this situation, what types of defects are likely to surface in these interfacing modules?( BTL3)</b></p> <ul style="list-style-type: none"> <li>• Communication defects.</li> </ul>

16	<b>List the stages in Software Engineering (BTL1)</b> <ul style="list-style-type: none"> <li>• Requirement Analysis</li> <li>• Design</li> <li>• Coding</li> <li>• Deployment</li> <li>• Delivery</li> </ul>
17	<b>Define Test Cases.</b> ( BTL1) A test case in a practical sense is a test related item which contains the following information. <b>A set of test inputs.</b> These are data items received from an external source by the code under test. The external source can be hardware, software, or human. <b>Execution conditions.</b> These are conditions required for running the test, for example, a certain state of a database, or a configuration of a hardware device. <b>Expected outputs.</b> These are the specified results to be produced by the code under test.
18.	<b>Define Test Oracle</b> ( BTL1) Test Oracle is a document, or a piece of software that allows tester to determine whether a test has been passed or failed.
19.	<b>Define Test Bed.(AU Nov/Dec 2017)( BTL1)</b> A test bed is an environment that contains all the hardware and software needed to test a software component or a software system.
20.	<b>Define Software Quality.</b> ( BTL1) Quality relates to the degree to which a system, system component, or process meets specified requirements. Quality relates to the degree to which a system, system component, or process meets Customer or user needs, or expectations.
21.	<b>List the Quality Attributes.( BTL1)</b> <ul style="list-style-type: none"> <li>• Correctness</li> <li>• Reliability</li> <li>• Usability</li> <li>• Integrity</li> <li>• Portability</li> <li>• Maintainability</li> <li>• Interoperability</li> </ul>
22.	<b>Define SQA group.(BTL1)</b> The software quality assurance (SQA) group is a team of people with the necessary training and skills to ensure that all necessary actions are taken during the development process so that the resulting software confirms to established technical requirements.
23.	<b>Explain the work of SQA group.( BTL2)</b> Testers to develop quality related policies and quality assurance plans for each project. The group is also involved in measurement collection and analysis, record keeping, and Reporting. The SQA team members participate in reviews and audits, record and track Problems, and verify that corrections have been made.
24.	<b>Define reviews.</b> ( BTL1) A review is a group meeting whose purpose is to evaluate a software artifact or a set of Software artifacts. Review and audit is usually conducted by a SQA group.
25.	<b>List the sources of Defects or Origins of defects.</b> (AU April/May 2017)( BTL1) <ul style="list-style-type: none"> <li>• Education</li> <li>• Communication</li> </ul>

	<ul style="list-style-type: none"> <li>• Oversight</li> <li>• Transcription</li> <li>• Process.</li> </ul>
<b>PART *B</b>	
1	<p><b>Discuss about the role of process in software quality (Testing). (13M) BTL2</b></p> <p><b>Answer:</b> page : 1 - Notes</p> <p><b>Process</b> – Creates an impact in the system.(2M)</p> <p><b>Testing as a process:</b> Set of activities well planned in advance.(2M)</p> <p><b>Figure</b> - Components of an engineered process.(2M) (Pg no: 2 in notes)</p> <p><b>Explanation:</b></p> <p><b>Testing Maturity model TMM (4M)</b></p> <ul style="list-style-type: none"> <li>• Level 1: Initial</li> <li>• Level 2: Phase Definition</li> <li>• Level 3: Integration</li> <li>• Level 4: Management and Measurement</li> <li>• Level 5: Optimization/Defect Prevention/Quality control</li> </ul> <p><b>Verification and Validation Model(3M)</b></p> <ul style="list-style-type: none"> <li>• <b>Diagram :</b> Pg no:2 in notes</li> <li>• <b>Verification:</b> Checks if software confirm to Functional and Non – Functional requirements</li> <li>• <b>Validation:</b> Confirms if Software meets user requirements.</li> </ul>
Q6	
2	<p><b>Draw the 5-level structure of the testing maturity model ,discuss about it.(13M) BTL2</b></p> <p><b>Answer :</b> page : 9 - Notes</p> <p><b>Test maturity model :</b> Gives an overview of the activities done in each level of testing.</p> <p><b>Explanation: Testing Maturity model TMM (4M)</b></p> <ul style="list-style-type: none"> <li>• Level 1: Initial</li> <li>• Level 2: Phase Definition</li> <li>• Level 3: Integration</li> <li>• Level 4: Management and Measurement</li> <li>• Level 5: Optimization/Defect Prevention/Quality control</li> </ul> <p><b>Verification and Validation Model(3M)</b></p> <ul style="list-style-type: none"> <li>• <b>Diagram :</b> Pg no:2 in notes (4M)</li> <li>• <b>Verification:</b> Checks if software confirm to Functional and Non – Functional requirements</li> <li>• <b>Validation:</b> Confirms if Software meets user requirements.</li> </ul>
3	<p><b>Explain in detail about the software testing principles.(13M)</b></p> <p><b>April/may 2017BTL2</b></p> <p><b>Answer :</b> Page: 3-23 - Srinivasan &amp; Ramaswamy</p> <p><b>Definition:</b> Principle is any rule that governs the system.</p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Principle 1 :Revealing defects and evaluating quality (2M)</li> <li>• Principle 2 :Effectiveness of testing effort (1M)</li> <li>• Principle 3 : Test results should be inspected (1M)</li> <li>• Principle 4 : Test case must contain the expected output (1M)</li> <li>• Principle 5:Test case developed for both valid and invalid input conditions (1M)</li> </ul>

	<ul style="list-style-type: none"> <li>• Principle 6 :Defects ratio (1M)</li> <li>• Principle7 : Testing should be carried out by a group (1M)</li> <li>• Principle8 : Tests must be repeatable and reusable (1M)</li> <li>• Principle9 :Testing should be planned (1M)</li> <li>• Principle 10: Testing activities should be integrated into software lifecycle (1M)</li> <li>• Principle 11: Testing is a creative and challenging task (2)</li> </ul>
4	<p><b>Give an example for defect classes and discuss them in detail.(13M)</b>  <b>(Nov/Dec 2016) BTL4</b></p> <p><b>Answer : page : 10 - Notes</b></p> <p><b>Definition:</b>Any abnormal condition that affects the execution of a program is called defect. (2M)</p> <p><b>Figure:</b>Defect classes and a defect repository.(2M)</p> <p><b>Explanation: (9M)</b></p> <ul style="list-style-type: none"> <li>• <b>Requirements and specification defects</b> <ol style="list-style-type: none"> <li>a. Functional Description defects</li> <li>b. Feature defects</li> <li>c. Feature interaction defects</li> <li>d. Interface description defects,</li> </ol> </li> <li>• <b>Design defects</b> <ol style="list-style-type: none"> <li>a. Algorithmic and processing defects</li> <li>b. Control ,logic, and sequence defects</li> <li>c. Data defects.</li> <li>d. Module interface description defects</li> <li>e. External Interface description defects.</li> </ol> </li> <li>• <b>Coding defects</b> <ol style="list-style-type: none"> <li>a. Algorithmic and processing defects.</li> <li>b. Control ,logic, and sequence defects</li> <li>c. Typographical defects</li> <li>d. Initialization defects.</li> <li>e. Dataflow defects</li> <li>f. Data defects</li> <li>g. Module interface defects,</li> <li>h. Code document defects.</li> <li>i. External hardware and software interface defects,</li> <li>j. Testing defects.</li> <li>k. Test harness defects</li> <li>l. Test case design and test procedure Defects.</li> </ol> </li> </ul>
5	<p><b>Explain in detail about Testing as a Process.</b> <span style="float: right;">(13M )BTL2</span></p> <p><b>Anwer : Page :29 - 31 - Srinivasan &amp; Ramaswamy book</b></p> <p><b>Process –</b> Creates an impact in the system.(2M)</p> <p><b>Testing as a process:</b> Set of activities well planned in advance.(2M)</p> <p><b>Figure</b> - Components of an engineered process.(2M) (Pg no: 2 in notes)</p> <p><b>Explanation:</b></p> <p><b>Testing Maturity model TMM (4M)</b></p> <ul style="list-style-type: none"> <li>• Level 1: Initial</li> <li>• Level 2: Phase Definition</li> <li>• Level 3: Integration</li> <li>• Level 4: Management and Measurement</li> </ul>

	<ul style="list-style-type: none"> <li>• Level 5: Optimization/Defect Prevention/Quality control</li> </ul> <p><b>Verification and Validation Model(3M)</b></p> <ul style="list-style-type: none"> <li>• <b>Diagram :</b> Pg no:2 in notes</li> <li>• <b>Verification:</b> Checks if software confirm to Functional and Non – Functional requirements</li> </ul> <p><b>Validation:</b> Confirms if Software meets user requirements.</p>
6	<p><b>Give a detailed account on the origins of defects.</b> (13M)  <b>(Nov/Dec 2016) BTL2</b></p> <p><b>Answer : page: 10 - Notes</b></p> <p><b>Definition:</b> Any abnormal condition that affects the execution of a program is called defect. (2M)</p> <p><b>Figure:</b> Defect classes and a defect repository.(4M)</p> <p><b>Explanation:</b> (3M)</p> <p><b>Types of Defect:</b></p> <ul style="list-style-type: none"> <li>• Variance from product specification</li> <li>• Variance from customer or user specification</li> <li>• Wrong requirement</li> <li>• Missing Requirement</li> <li>• Extra Requirement</li> </ul> <p><b>Sources of Error:</b></p> <ul style="list-style-type: none"> <li>• Education</li> <li>• Communication</li> <li>• Oversight</li> <li>• Transcription</li> <li>• Process</li> </ul> <p><b>Hypothesis:</b></p> <ul style="list-style-type: none"> <li>• Design Test</li> </ul> <p><b>Fault Model:</b> Employed to prepare fault list.</p> <p><b>Physical Defects:</b></p> <ul style="list-style-type: none"> <li>• Manufacturing Errors</li> <li>• Component Wear out.</li> <li>• Environmental Effects</li> </ul> <p><b>Figure : Origins of defects.(4M)</b></p>
7	<p><b>Discuss the Coin Problem along with the defect categories associated withit.(13M) BTL6</b></p> <p><b>Answer : page : 17 - Notes</b></p> <p><b>Figure :</b> Sample specification with defects.(4M)</p> <p><b>Figure :</b> a sample design specification with defects(4M)</p> <p><b>Explanation:</b>(5M)</p> <ul style="list-style-type: none"> <li>• Algorithmic and processing defects.</li> <li>• Precondition</li> <li>• Post-condition</li> <li>• Control,logic, and sequence defects.</li> <li>• Typographical defects.</li> </ul>

	<ul style="list-style-type: none"> <li>Initialization defects.</li> <li>Dataflow defects.</li> </ul>
8	<p><b>Analyse the role of tester in software development Organization.(13M) BTL4 (Nov/Dec 2017)</b></p> <p><b>Answer : page:9 - Notes</b></p> <p><b>Tester:</b> Objective of testing to get high quality software which should satisfy all requirements of software. Role of tester to ensure whether all requirements of software are satisfied. (2M)</p> <p><b>Explanation:</b></p> <p><b>Tester's job:(4M)</b></p> <ul style="list-style-type: none"> <li>Reveal defects</li> <li>Find weak points</li> <li>Inconsistent behavior</li> <li>Circumstances where the software does not work as expected.</li> </ul> <p><b>Tester's Need:</b></p> <ul style="list-style-type: none"> <li>Communication Skills</li> <li>Team working skills</li> <li>Decision Making skills</li> </ul> <p><b>Testers are said to be specialist:</b></p> <ul style="list-style-type: none"> <li>Provide plan</li> <li>Do Execution</li> <li>Recording result</li> <li>Analysing the test result</li> </ul>
<b>PART – C</b>	
1	<p><b>Given 6 different denominations of coins ,The program finds total dollars &amp; cents values for a set of coins and outputs the number of dollars. Find the possible defects in the above scenario.(15M) BTL6 (April/May 2017)</b></p> <p><b>Answer : Appendix - Srinivasan,Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>Requirements or functional Defects(4M)</li> <li>Functional description defects</li> <li>Interface description defects</li> <li>Pre conditions(5M)</li> <li>Post conditions</li> <li>Control,Logic and sequence defects(3M)</li> <li>Algorithmic and processing defects</li> <li>Data flow defects(3M)</li> <li>Data Defects</li> <li>External</li> </ul>
2	<p><b>If you were testing a feature of your software on Monday and finding a new bug every hour, at what rate would you expect to find bugs on Tuesday? (15M) BTL4</b></p> <p><b>Answer : Appendix - Srinivasan,Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>Number of bugs remaining is proportional to the number of bugs you have already found.(8M)</li> <li>Pesticide paradox – Tells you that if you continue to run the same tests over and over that you eventually won't find new and different bugs until you add more tests.(7M)</li> </ul>
3	<p><b>Visiting all the states that the program has assures that you have also traverses all the</b></p>

**transitions among them. The statement is true or false? Justify your answer.(15M) BTL4**

**Answer : Appendix - Srinivasan Ramaswamy**

**STATEMENT :False(8M)**

**Explanation(7M)**

- Think of visiting 50 different cities spread out across the entire United states.
- You could plan a trip that would take you to each city.
- But it would be impossible for you to travel all the roads that connects all the cities.

JIT - 2106

## UNIT II TEST CASE DESIGN

Test case Design Strategies – Using Black Box Approach to Test Case Design – Random Testing – Requirements based testing – Boundary Value Analysis – Equivalence Class Partitioning – State-based testing – Cause-effect graphing – Compatibility testing – user documentation testing – domain testing – Using White Box Approach to Test design – Test Adequacy Criteria – static testing vs. structural testing – code functional testing – Coverage and Control Flow Graphs – Covering Code Logic – Paths – code complexity testing – Evaluating Test Adequacy Criteria.

### PART \*A

1	<b>Define Smart Tester.(BTL1)</b> Software must be tested before it is delivered to users. It is responsibility of the testers to Design tests that (i) reveal defects (ii) can be used to evaluate software performance, usability and reliability. To achieve these goals, tester must select a finite no. of test cases (i/p, o/p, & conditions).
2	<b>Define responsibility.(AU Nov/Dec 2016)(BTL1)</b> A straightforward definition for object-responsibility is this: An object must contain the data (attributes) and code (methods) necessary to perform any and all services that are required by the object.
3	<b>Write short notes on Random testing and Equivalence class portioning.(BTL1) (Nov/Dec 2015)</b> Each software module or system has an input domain from which test input data is selected. If a tester randomly selects inputs from the domain, this is called random testing. In equivalence class partitioning the input and output is divided in to equal classes or partitions.
4	<b>Define State.(BTL1)</b> A state is an internal configuration of a system or component. It is defined in terms of the values assumed at a particular time for the variables that characterize the system or component.
5	<b>Define Finite-State machine. (BTL1)</b> A finite-state machine is an abstract machine that can be represented by a state graph having a finite number of states and a finite number of transitions between states.
6	<b>Define Error Guessing.(BTL1)</b> The tester/developer is sometimes able to make an educated “guess” as to which type of defects may be present and design test cases to reveal them. Error Guessing is an ad-hoc approach to test design in most cases.
7	<b>Define COTS Components.(BTL1)</b> The reusable component may come from a code reuse library within their org or, as is most likely, from an outside vendor who specializes in the development of specific types of software components. Components produced by vendor org are known as commercial off-the shelf, or COTS, components.
8	<b>Express the benefits of low coupling.(BTL2)</b> <b>Maintainability</b> – changes are confined in a single module <b>Testability</b> – modules involved in unit testing can be limited to a minimum <b>Readability</b> – classes that need to be analysed are kept at a minimum.
9	<b>Define usage profiles and Certification.(BTL1)</b> Usage profiles are characterizations of the population of intended uses of the software in its intended environment. Certification refers to third party assurance that a product,process, or service meets a specific set of requirements.

10	<b>Write the application scope of adequacy criteria?(BTL4)</b> <ul style="list-style-type: none"> <li>Helping testers to select properties of a program to focus on during test.</li> <li>Helping testers to select a test data set for a program based on the selected properties.</li> <li>Supporting testers with the development of quantitative objectives for testing</li> <li>Indicating to testers whether or not testing can be stopped for that program.</li> </ul>
11	<b>Define path.(BTL1)</b> A path is a sequence of control flow nodes usually beginning from the entry node of a graph through to the exit node.
12	<b>Write the formula for cyclomatic complexity?(AU Nov/Dec 2016)(BTL1)</b> The complexity value is usually calculated from control flow graph(G) by the formula. $V(G) = E - N + 2$ Where The value E is the number of edges in the control flow graph The value N is the number of nodes.
13	<b>List the various iterations of Loop testing. ?(BTL1)</b> <ul style="list-style-type: none"> <li>Zero iteration of the loop</li> <li>One iteration of the loop</li> <li>Two iterations of the loop</li> <li>K iterations of the loop where <math>k &lt; n</math></li> <li><math>n-1</math> iterations of the loop</li> <li><math>n+1</math> iterations of the loop</li> </ul>
15	<b>What are the errors uncovered by black box testing?(BTL1)</b> <ul style="list-style-type: none"> <li>Incorrect or missing functions</li> <li>Interface errors</li> <li>Errors in data structures</li> <li>Performance errors</li> <li>Initialization or termination error.</li> </ul>
16	<b>Define Equivalence class partitioning?(BTL1)</b> If a tester is viewing the software-under-test as a black box with well defined inputs and outputs, a good approach to selecting test inputs is to use a method called Equivalence class partitioning.
17	<b>Define Cause effect graphing?(BTL1)</b> Cause Effect Graph is a black box testing technique that graphically illustrates the relationship between a given outcome and all the factors that influence the outcome.
18	<b>What is Certification?(BTL1)</b> Certification refers to third-party assurance that a product, process, or service meets a specific set of requirements.
19	<b>What is the goal of smart tester?(BTL1)</b> The goal of the smart tester is to understand the functionality, input/output domain, and the environment of use for the code being tested.
20	<b>List the two major assumptions in Mutation testing.(BTL1)</b> <ul style="list-style-type: none"> <li>The component programmer hypothesis</li> <li>The coupling effects</li> </ul>
21	<b>List the two basic Testing strategies.(BTL1)</b> <ul style="list-style-type: none"> <li>Black box testing.</li> <li>White box testing.</li> </ul>
22	<b>What are the knowledge sources for Black box testing?(BTL2)</b> <ul style="list-style-type: none"> <li>Requirement</li> <li>Document specification</li> <li>Domain knowledge</li> </ul>

	<ul style="list-style-type: none"> <li>Defect analysis data</li> </ul>
23	<p><b>What are the knowledge sources for White box testing? (AU Nov/Dec 2015)(BTL2)</b></p> <ul style="list-style-type: none"> <li>High level design</li> <li>Detailed design</li> <li>Control flow graphs</li> <li>Cyclomatic complexity</li> </ul>
24	<p><b>List the methods of Black box testing?(AU Nov/Dec 2017)(BTL1)</b></p> <ul style="list-style-type: none"> <li>Equivalence class partitioning</li> <li>Boundary value analysis</li> <li>State transition testing</li> <li>Cause and effect graphing</li> <li>Error guessing</li> </ul>
25	<p><b>List the methods of White box testing?(AU Nov/Dec 2017)(BTL1)</b></p> <ul style="list-style-type: none"> <li>Statement testing</li> <li>Branch testing</li> <li>Path testing</li> <li>Data flow testing</li> <li>Mutation testing</li> <li>Loop testing</li> </ul>
<b>PART* B</b>	
1	<p><b>Elaborate the qualities of a smart Tester.BTL2(13M)</b></p> <p><b>Answer : page : 18 - Notes</b></p> <ul style="list-style-type: none"> <li>Reveal defects(4M) <ul style="list-style-type: none"> <li>Find the bugs before the software becomes operational</li> <li>Find errors at the early stage (Requirement Analysis)</li> <li>Find the weak points</li> <li>Situations at which error may occur</li> </ul> </li> <li>Evaluate quality(4M) <ul style="list-style-type: none"> <li>Ensures if software meets user requirements</li> <li>Ensures if software meets requirement specification</li> <li>Ensures if software meets performance criteria such as reliability,usability, portability</li> </ul> </li> <li>Finite no of test case(5M) <ul style="list-style-type: none"> <li>Number of test cases.</li> <li>A test case that makes the tester to makesure that software meets all user requirements.</li> <li>Test cases that are capable enough to make the system to crash.</li> </ul> </li> </ul>
2	<p><b>Discuss the test case design strategies.BTL2 (13M)</b></p> <p><b>Answer: page : 18 - Notes</b></p> <ul style="list-style-type: none"> <li>Two strategies</li> <li>Whitebox (clear or glass box) (2M)</li> <li>Testing the software with X – Ray glasses</li> <li>Black box(Functional or specification)(2M)</li> <li>Testing the software blind folded.</li> </ul>

	<ul style="list-style-type: none"> <li>Table: The two basic testing strategies.(9M)Page 18 in notes</li> </ul>
3	<p><b>List and explain the types of black box testing.AU April/May 2016 BTL2 (13M)</b></p> <p><b>Answer:Page:73 - 105 - Srinivasan &amp; Ramaswamy</b></p> <ul style="list-style-type: none"> <li>Random testing (1M) <ul style="list-style-type: none"> <li>Randomly select the input.</li> <li>Three conditions.</li> </ul> </li> <li>Equivalence class partitioning(2M) <ul style="list-style-type: none"> <li>Adv of Equivalence class partitioning</li> <li>List of conditions.</li> <li>Figure: A specification of a square root function</li> <li>Example of equivalence class reporting table</li> </ul> </li> <li>Boundary value analysis(1M) <ul style="list-style-type: none"> <li>List the conditions</li> <li>Figure: Boundaries of on Equivalence partition</li> <li>Example of Boundary value analysis.</li> </ul> </li> <li>State Transition Testing(1M) <ul style="list-style-type: none"> <li>Abstract Machine</li> <li>State graph having a finite number of states and transitions between</li> <li>Internal configuration of system or component</li> </ul> </li> <li>Error guessing(1M) <ul style="list-style-type: none"> <li>Tester/Developer's past experience</li> </ul> </li> <li>Cause and Effect Graphing(2M) <ul style="list-style-type: none"> <li>Nodes in the graph are causes and effects</li> <li>Tester need to identify causes and effects</li> <li>Graph must be annotated with constraints</li> <li>Graph is then converted into decision table</li> <li>Columns in the decision table are converted into test cases</li> </ul> </li> <li>Requirement Based Testing(1M) <ul style="list-style-type: none"> <li>Test Requirement Specification</li> <li>Explicit Requirement</li> <li>Implicit Requirement</li> <li>Requirement traceability Matrix</li> </ul> </li> <li>Compatibility Testing(1M) <ul style="list-style-type: none"> <li>Confirms working of product with different infrastructure components</li> <li>Forward Compatibility Testing</li> <li>Backward Compatibility testing</li> </ul> </li> <li>User documentation Testing(2M) <ul style="list-style-type: none"> <li>Manuals, User guidelines</li> <li>Installation guidelines</li> <li>Setup guidelines, Readme files</li> <li>Software Release notes,Online help</li> </ul> </li> <li>Domain Testing(1M) <ul style="list-style-type: none"> <li>Needs business domain knowledge than software knowledge</li> <li>They get trained in software ,instead of training the software professional in business domain.</li> </ul> </li> </ul>

4	<p><b>Discuss the various approaches in White Box test design.(13M)</b></p> <p><b>AU Nov/Dec 2016BTL2</b></p> <p><b>Answer:Page :29 - 31 - Srinivasan &amp; Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• Coverage and control flow graph(3M)             <ol style="list-style-type: none"> <li>1. Three basic primes</li> <li>2. Sequential</li> <li>3. Condition</li> <li>4. Iteration</li> </ol> </li> <li>• Coverage code logic(3M)             <ol style="list-style-type: none"> <li>1. Figure: Code sample with branch and loop.</li> <li>2. Figure: A control flow graph representation for the code.</li> <li>3. Table: A test case for the code ,that satisfies the decision</li> </ol> </li> <li>• coverage criterion.(3M)             <ol style="list-style-type: none"> <li>1. Table: Test cases for simple decision coverage</li> <li>2. Table: Test cases for condition coverage</li> <li>3. Table: Test cases for decision condition coverage.</li> </ol> </li> <li>• Path Testing (4M)             <ol style="list-style-type: none"> <li>1. Path</li> <li>2. Cyclomatic complexity formula.</li> </ol> </li> </ul>
5	<p><b>Evaluate test adequacy Criteria with necessary properties.(13M)BTL3</b></p> <p><b>Answer: page:27 - Srinivasa &amp; Ramamurty</b></p> <ul style="list-style-type: none"> <li>• Axioms –Set of assumptions(1M)</li> <li>• Applicability Property(1M)</li> <li>• Non exhaustive applicability property(1M)</li> <li>• Monotonicity Property(2M)</li> <li>• Inadequate Empty set(1M)</li> <li>• General multiple change Property(1M)</li> <li>• Anti decomposition Property(2M)</li> <li>• Renaming Property(1M)</li> <li>• Complexity Property(1M)</li> <li>• Statement Coverage Property(2M)</li> </ul>
6	<p><b>Demonstrate the various black box testing approaches using Equivalence partitioning and boundary value Analysis.(13M)</b></p> <p><b>Nov/Dec 2016BTL5</b></p> <p><b>AnswerPage : 84,90 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Equivalence Partition: (8M)</b></p> <p>Software testing technique – divides input data of software unit into partitions of equivalent data – test cases can be derived – main principal of test cases are deigned to coverpartition at least once.</p> <p><b>Boundary value analysis: (5M)</b></p> <p>Test case design technique to test boundary value between partitions-boundary value is an input or output value on the border of an equivalence partition.</p>
7	<p><b>Compare static testing with that of dynamic testing and list the major difference between both.(13M) BTL4</b></p>

	<b>STATIC TESTING</b>	<b>DYNAMIC TESTING</b>
Prevention.	Cure.	
More cost-effective.	Less cost – effective.	
Greater marginal benefits.	Lesser marginal benefits.	
Comprehensive diagnostics for code.	More diagnostics for code.	
Finds more bugs.	Finds fewer bugs.	
Takes lesser time.	Takes longer time.	
Testing covers more areas.	Testing covers less areas.	
Done in verification stage.	done in validation stage.	
<b>PART *C</b>		
1	<b>Explain What a tester should worry about with this line from a spec. The software will allow up to 100 million simultaneous connections, although no more than 1 million will normally be used.BTL4(15 M)</b>	
	<b>Answer : Appendix - Srinivasan Ramaswamy</b>	
	<ul style="list-style-type: none"> <li>• Testability(4M)</li> <li>• It doesn't matter that typical usage is only 1 million connections.(4M)</li> <li>• If the specification states that 100 million are possible(4M)</li> <li>• The 100 million must be tested.(3M)</li> </ul>	
2	<b>Assume that you are assigned to test the windows calculator, Is it possible to test all the test cases. How do you test it systematically and explain the principle involved.BTL6(15M)</b>	
	<b>Answer: Appendix - Srinivasan Ramaswamy</b>	
	<ul style="list-style-type: none"> <li>• Equivalence Partitioning(5M)</li> <li>• Grouping similar input(4M)</li> <li>• Grouping similar Output(4M)</li> <li>• Grouping similar operations of software(2M)</li> </ul>	
3	<b>Visiting all the states that the program has assures that you have also traverses all the transitions among them. The statement is true or false? Justify your answer.BTL4(15M)</b>	
	<b>Answer : Appendix - Srinivasan Ramaswamy</b>	
	False(8M)	
	<b>Explanation(7M)</b>	
	<ul style="list-style-type: none"> <li>• Think of visiting 50 different cities spread out across the entire United States.</li> <li>• You could plan a trip that would take you to each city.</li> <li>• But it would be impossible for you to travel all the roads that connect all the cities.</li> </ul>	

### UNIT III LEVELS OF TESTING

The need for Levels of Testing – Unit Test – Unit Test Planning – Designing the Unit Tests – The Test Harness – Running the Unit tests and Recording results – Integration tests – Designing Integration Tests – Integration Test Planning – Scenario testing – Defect bash elimination System Testing – Acceptance testing – Performance testing – Regression Testing – Internationalization testing – Ad-hoc testing – Alpha, Beta Tests – Testing OO systems – Usability and Accessibility testing – Configuration testing – Compatibility testing – Testing the documentation – Website testing.

#### PART A

1	<b>Define Unit Testing (Nov/Dec 2017)BTL1</b> A unit is the smallest possible testable software component that can be characterized in several ways.
2	<b>Write the different levels of testing.BTL1</b> <ul style="list-style-type: none"> <li>• Unit test</li> <li>• Integration test</li> <li>• System test</li> <li>• Acceptance test.</li> </ul>
3	<b>List the components suitable for unit test.BTL1</b> <ul style="list-style-type: none"> <li>• Procedures and functions</li> <li>• Classes/objects and methods</li> </ul> Procedure-sized reusable components.
4	<b>List the phases in the unit test planning. (April/May 2015)BTL1</b> <ul style="list-style-type: none"> <li>• Phase 1: Describe unit test approach and risks.</li> <li>• Phase 2: Identify unit features to be tested.</li> <li>• Phase 3: Add levels of detailed to the plan.</li> </ul>
5	<b>Write the issues in the unit test.BTL1</b> <ul style="list-style-type: none"> <li>• Issue 1: Adequately testing classes.</li> <li>• Issue 2: Observation of objects states and state changes.</li> <li>• Issue 3: The retesting of classes-I</li> <li>• Issue 4: The retesting of classes-II</li> </ul>
6	<b>What is Test harness?(Nov/Dec 2016)BTL1</b> The auxiliary code developed to support to testing of units and components is called a test harness. The harness consists of drivers that call the target code and stubs that represent modules it calls.
7	<b>List the major goals of Integration test.BTL1</b> <ul style="list-style-type: none"> <li>• To detect defects that occurs on the interfaces of units.</li> <li>• To assemble the individual units into working subsystems and the finally a complete system that is ready for system test</li> </ul>
8	<b>What is the advantage of Bottom up integration?BTL1</b> Bottom-up integration has the advantage that the lower-level modules are usually well tested early in the integration process. This is important if these modules are candidates for reuse.
9	<b>What is a cluster?BTL1</b> A cluster consists of classes that are related, for example, they may work together to support a required functionality for the complete system.

10	<b>List the several types of system tests.(Nov/Dec 2016)BTL1</b> <ul style="list-style-type: none"> <li>• Functional testing</li> <li>• Performance testing</li> <li>• Stress testing</li> <li>• Configuration testing</li> <li>• Security testing</li> <li>• Recovery testing</li> </ul>
11	<b>Define Load.</b> BTL1 A load is a series of inputs that simulates a group of transactions.
12	<b>List the two major requirements of Performance testing.</b> BTL1 <ul style="list-style-type: none"> <li>• Functional requirements</li> <li>• Quality requirements.</li> </ul>
13	<b>What is meant by Stress testing?</b> BTL1 When a system is tested with a load that causes it to allocate its resources in maximum amounts, this is called stress testing.
14	<b>Define Recovery testing.</b> BTL1 Recovery testing subjects a system to losses of resources in order to determine if it can recover properly from these losses.
15	<b>Define Use case.</b> BTL1 A use case is a pattern, scenario, or exemplar of usage. It describes a typical interaction between the software system under development and a user.
16	<b>Define Regression testing .</b> BTL1 Regression testing is not a level of testing, but it is the retesting of the software that occurs when the changes are made to ensure that the new version of the software has retained the capabilities of the old version and that no defect have been introduced due to the changes.
17	<b>Write the objectives of configuration testing.</b> BTL1 <ul style="list-style-type: none"> <li>• Show that all the configuration changing commands and menus work properly</li> <li>• Show that all interchangeable devices are really interchangeable, and that they each enter the proper states for the specified conditions</li> <li>• Show that the system's performance level is maintained when devices are interchanged, or when they fail.</li> </ul>
18	<b>List the effect of security breaches.</b> BTL1 <ul style="list-style-type: none"> <li>• Loss of information</li> <li>• Corruption of information</li> <li>• Misinformation</li> <li>• Privacy violations</li> <li>• Denial of service</li> </ul>
19	<b>Define functional Testing.</b> BTL1 Functional tests at the system level are used ensure that the behavior of the system adheres to the requirement specifications.
20	<b>What is load generator and Load?</b> BTL1 An important tool for implementing system tests is a load generator. A load generator is essential for testing quality requirements such as performance and stress.  A load is a series of inputs that simulates a group of transactions.
21	<b>What are the approaches used to develop the software?</b> BTL1 There are two major approaches to software development

	<ul style="list-style-type: none"> <li>• Bottom-Up</li> <li>• Top-Down</li> </ul>
22	<p><b>List the objectives of configuration testing.</b> BTLL</p> <ul style="list-style-type: none"> <li>• Show that all the configuration changing commands and menus work properly</li> <li>• Show that all interchangeable devices are really interchangeable, and that they each enter the proper states for the specified conditions</li> <li>• Show that the system's performance level is maintained when devices are interchanged, or when they fail.</li> </ul>
23	<p><b>List the effect of security breaches.</b> BTLL</p> <ul style="list-style-type: none"> <li>• Loss of information</li> <li>• Corruption of information</li> <li>• Misinformation</li> <li>• Privacy violations</li> <li>• Denial of service.</li> </ul>
24	<p><b>Give the examples of security testing.</b> BTLL2</p> <ul style="list-style-type: none"> <li>• Password checking.</li> <li>• Legal and illegal entry with password.</li> <li>• Password Expiration.</li> <li>• Encryption.</li> <li>• Browsing.</li> <li>• Trap doors.</li> <li>• Viruses.</li> </ul>
25	<p><b>List the areas covered during recovery testing.</b> BTLL</p> <ul style="list-style-type: none"> <li>• Restart.</li> <li>• Switchover.</li> </ul>

### PART B

1	<p><b>How would you define a software unit? In terms of your definition, what constitutes a unit for procedural code; for object-oriented code?(13M)BTLL4</b></p> <p><b>Answer : Page : 38,261-264 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Functions, procedures, classes and methods as units</li> <li>• <b>Fig:</b> Some components suitable for unit test(1M)</li> <li>• Unit Test: Need for preparation <ul style="list-style-type: none"> <li>• Planning</li> <li>• Both black box and White box</li> <li>• Reviewe</li> <li>• Several Tasks</li> </ul> </li> </ul> <p>1. Unit Test Planning(4M)</p> <p>Phase I: Describe unit test approach and Risks</p> <p>Phase II: Identify unit features to be tested</p> <p>Phase III: Add levels of detail to the planning</p> <p>2. Designing the Unit Test(3M)</p> <ul style="list-style-type: none"> <li>• Test Cases</li> <li>• Test Procedure</li> </ul> <p>3. Running and recording the results(3M)</p>
---	---

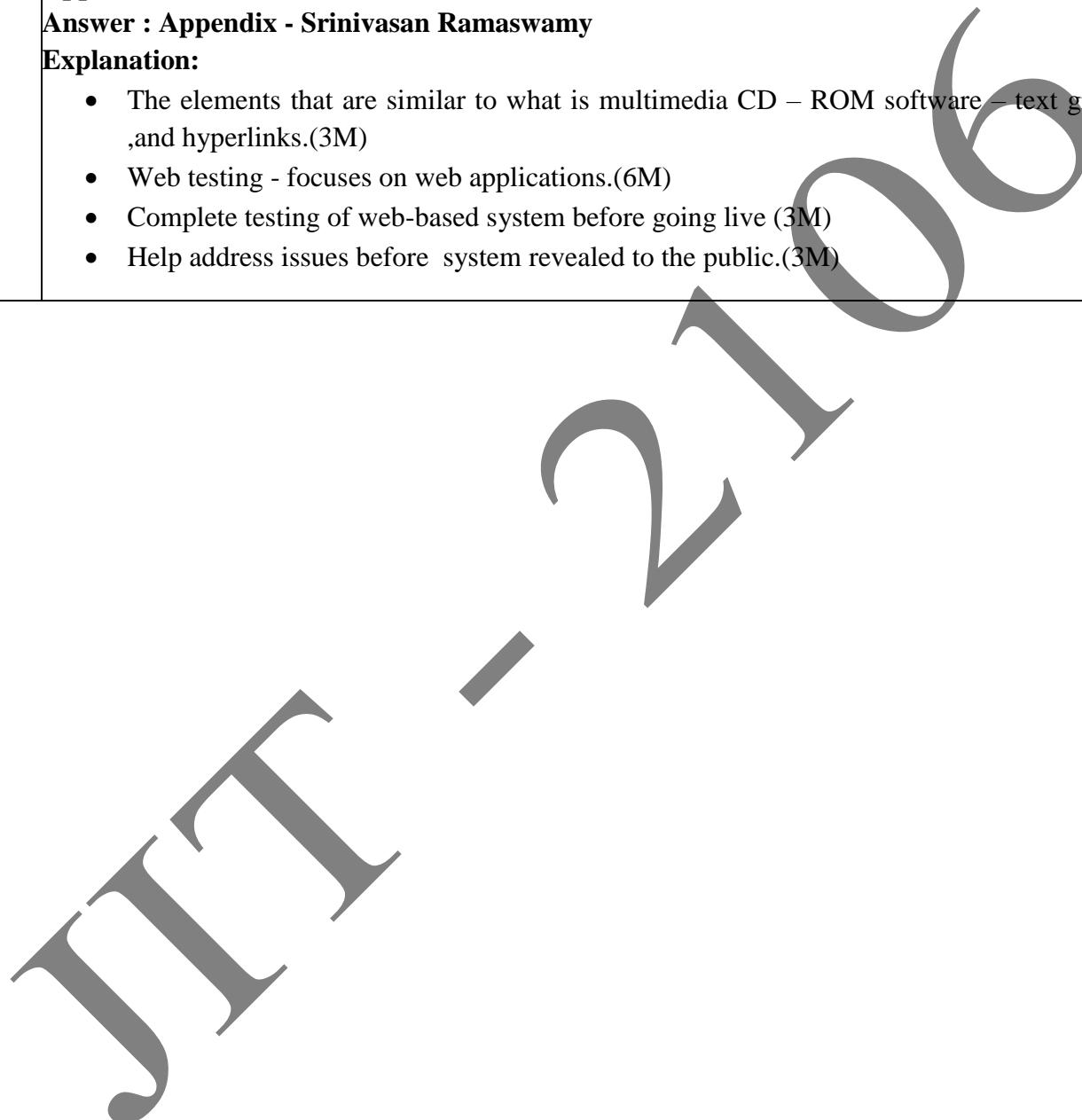
	<ul style="list-style-type: none"> <li>• Perform the unit test in all the units of system</li> <li>• Record the results.</li> </ul> <p>4. Test Harness(2M)</p> <ul style="list-style-type: none"> <li>• Additional code included to perform testing.</li> </ul>
2	<p><b>Why is it so important to design a test harness for reusability?(13M)BTL2</b></p> <p><b>Answer:</b> Page : 35 - Notes</p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Auxiliary code developed</li> <li>• Support testing of units , components (3M)</li> <li>• Harness consists of drivers that call the target code</li> <li>• Stubs that represent modules it calls.(3M)</li> </ul> <p><b>Fig:</b> The test Harness(4M)</p> <p>Driver(2M)</p> <p>Stub(1M)</p>
3	<p><b>What are the key differences in integrating procedural-oriented systems as compared to object-oriented systems?(13M)BTL3</b></p> <p><b>Answer:</b> Page : 35 - Notes</p> <ul style="list-style-type: none"> <li>• Goals(2M)</li> <li>• Integration Strategies: <ul style="list-style-type: none"> <li>i. Top – Down</li> <li>ii. Bottom – Up</li> <li>iii. Bi – Directional</li> </ul> </li> <li>• Designing Integration Test: <ul style="list-style-type: none"> <li>i.Black Box Approach</li> <li>ii.White Box Approach</li> </ul> </li> <li>• Integration test strategy for procedures(5M)</li> <li>• Integration test strategy for classes(6M)</li> <li>• Critical Module characteristics</li> </ul>
4	<p><b>Describe the activities/Tasks and responsibilities for developer/testers in support of multilevel testing.(13M)BTL2</b></p> <p><b>Answer :</b> Page :261 - Srinivasan &amp; Ramaswamy</p> <ul style="list-style-type: none"> <li>• <b>Fig:</b> Levels of testing</li> </ul> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Levels of Testing(4M) <ul style="list-style-type: none"> <li>i. Unit Test</li> <li>ii. Integration test</li> <li>iii. System Test</li> <li>iv. Acceptance Test</li> </ul> </li> <li>• Two Approaches(4M) <ul style="list-style-type: none"> <li>Bottom_Up</li> <li>Top_Down</li> </ul> </li> <li>• Two types of Language(5M) <ul style="list-style-type: none"> <li>Procedure Oriented</li> <li>Object Oriented</li> </ul> </li> </ul>
5	<p><b>Explain Integration Test with example.(13M) (Nov/dec 2016)BTL3</b></p> <p><b>Answer:</b>Page : 107 - Srinivasan &amp; Ramaswamy</p>

	<p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Goals(2M)</li> <li>• Integration Strategies:           <ul style="list-style-type: none"> <li>i. Top – Down</li> <li>ii. Bottom – Up</li> <li>iii. Bi – Directional</li> </ul> </li> <li>• Designing Integration Test:           <ul style="list-style-type: none"> <li>1. Black Box Approach</li> </ul> </li> <li>• White Box Approach</li> <li>• Integration test strategy for procedures(5M)</li> <li>• Integration test strategy for classes(6M)</li> <li>• Critical Module characteristics</li> <li>• Example : Sandwich Testing</li> </ul>
6	<p><b>Explain the different types of system testing with example.(13M)BTL2</b></p> <p><b>Answer:Page : 130 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Functional testing(1M)</li> <li>• Performance testing(1M)</li> <li>• Stress testing(1M)</li> <li>• Configuration testing(1M)</li> <li>• Security testing(1M)</li> <li>• Recovery testing(1M)</li> <li>• <b>Fig:</b> Types of System Test(4M)</li> <li>• <b>Fig:</b> Example of special resources needed for a performance test(3M)</li> </ul>
7	<p><b>Explain in detail about scenario Testing.(13M)BTL2</b></p> <p><b>Answer:Page : 130 - Srinivasan &amp; Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• Two Methods(4M)           <ul style="list-style-type: none"> <li>i. System Scenarios</li> <li>ii. Use Case Scenarios</li> </ul> </li> <li>• Why Scenario test?(4M)           <ul style="list-style-type: none"> <li>i. Learn product</li> <li>ii. Connect Testing to documented requirement</li> <li>iii. Expose failure to deliver described benefits</li> <li>iv. Expose expert use of program</li> <li>v. Bring requirement related issues</li> </ul> </li> <li>• Twelve ways to create good scenarios(5M)</li> </ul>
7	<p><b>How would you identify hardware and software for configuration testing and how would you apply website testing?(13M)(Nov/dec 2016)BTL5</b></p> <p><b>Answer:Page : 195,198,369 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Configuration testing - testing application with multiple combinations (7M)</li> <li>• To find out the optimal <b>configurations</b></li> <li>• Web testing - focuses on web applications.(6M)</li> <li>• Complete testing of web-based system before going live</li> <li>• Help address issues before system revealed to the public.</li> </ul>

8	<p><b>i) Explain about Defect Bash Elimination.(7M)BTL2</b></p> <p><b>Answer : Page : 39 – Notes</b></p> <ul style="list-style-type: none"> <li>• Ad-hoc Testing(2M)</li> <li>• Not based on written test cases(2M)</li> <li>• Brings together plenty of good practices(1M)</li> <li>• Steps in defect bash(2M)</li> </ul> <p><b>ii) Explain about Ad-hoc Testing in detail.(6M)BTL2</b></p> <p><b>Answer : Page : 39 – Notes</b></p> <ul style="list-style-type: none"> <li>• Discovers unfound errors in software(2M)</li> <li>• Impacted due to(2M) <ul style="list-style-type: none"> <li>i. Intuition</li> <li>ii. Previous Experience</li> <li>iii. Expert knowledge of the platform</li> <li>iv. Experience in Testing</li> </ul> </li> <li>• Drawback</li> <li>• Figure : Ad - hoc Testing(2M)</li> </ul>
9	<p><b>i) Explain about usability and accessibility Testing.(7M)BTL2</b></p> <p><b>Answer : Page : 49 - Notes</b></p> <p><b>Usability testing:(4M)</b></p> <ul style="list-style-type: none"> <li>• Characteristics</li> <li>• Quality Factors</li> <li>• Approach to usability</li> <li>• Aesthetic testing</li> </ul> <p><b>Accessibility Testing:(3M)</b></p> <ul style="list-style-type: none"> <li>• Basic accessibility</li> <li>• Product accessibility</li> </ul> <p><b>ii) Explain Testing OO Model in detail.(6M)(BTL2)</b></p> <ul style="list-style-type: none"> <li>• Unit Testing</li> <li>• Integration testing</li> <li>• Validate and system testing</li> <li>• Regression testing</li> </ul>

10	<p><b>i) Differentiate Alpha and Beta Testing and discuss the phases in which alpha and beta testing are done?(7M)</b></p> <p><b>ii) Explain about documentation testing in detail.(6M)(Nov/Dec 2017)BTL3</b></p> <p><b>Answer:Page : 137-140 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Alpha Testing: (4M)</b></p> <ul style="list-style-type: none"> <li>• Type of acceptance testing</li> <li>• Performed to identify all possible issues/bugs</li> <li>• Before releasing the product to everyday users or public.</li> <li>• Aim to carry out the tasks that a typical user might perform.</li> </ul> <p><b>Beta Testing:(3M)</b></p> <ul style="list-style-type: none"> <li>• Second phase of <b>Software Testing</b></li> <li>• Sampling of the intended audience tries the product out.</li> <li>• Beta Testing of a product is performed by <b>real users</b> of the software application in a <b>real environment</b>.</li> </ul> <p><b>ii)Explanation:</b></p> <ul style="list-style-type: none"> <li>• Importance of documentation testing</li> <li>• Main things to look for in reviewing the document</li> <li>• Packaging and text graphics</li> <li>• Marketing materials,ads and other inserts</li> <li>• Warranty/Registration</li> <li>• EULA</li> <li>• Label and stickers</li> <li>• Installation setup &amp; Instructions</li> <li>• Users Manual</li> <li>• Online help</li> </ul>
1	<p><b>PART – C</b></p> <p><b>If you are assigned to test compatibility of your product's data file formats, How would you approach the task?(15M)BTL6</b></p> <p><b>Answer:Appendix - Srinivasan Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Research whether your program follows existing standards for its files.(6M)</li> <li>• If so, test that it meets its standards.(1M)</li> <li>• Equivalence partition the possible programs that would read and write your program's files.(6M)</li> <li>• Design test documents with representative sample of the types of data.(2M)</li> </ul> <p><b>2</b></p> <p><b>Explain the significance of control flow graph and cyclomatic complexity with the pseudo code for the sum of n numbers(13M).(Nov/Dec 2017)BTL6</b></p> <p><b>Answer : Appendix - Srinivasan Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Cyclomatic complexity is a software metric used to measure the complexity of a program.(5M)</li> </ul>

	<ul style="list-style-type: none"> <li>• This metric measures independent paths through the program's source code. An independent path is defined as a path that has at least one edge which has not been traversed before in any other paths.(6M)</li> <li>• Cyclomatic complexity can be calculated with respect to functions, modules, methods or classes within a program.(4M)</li> </ul>
3	<p><b>What basic elements of a web page can easily be tested with black box Approach?(15M)BTL6</b></p> <p><b>Answer : Appendix - Srinivasan Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• The elements that are similar to what is multimedia CD – ROM software – text graphics ,and hyperlinks.(3M)</li> <li>• Web testing - focuses on web applications.(6M)</li> <li>• Complete testing of web-based system before going live (3M)</li> <li>• Help address issues before system revealed to the public.(3M)</li> </ul>



JIT - 2106

<b>UNIT IV TEST MANAGEMENT</b>	
People and organizational issues in testing – Organization structures for testing teams – testing services – Test Planning – Test Plan Components – Test Plan Attachments – Locating Test Items – test management – test process – Reporting Test Results – The role of three groups in Test Planning and Policy Development – Introducing the test specialist – Skills needed by a test specialist – Building a Testing Group.	
<b>PART* A</b>	
1	<b>Define Goal in testing.</b> BTL1 A Goal can be described as a statement of intent or a statement of accomplishment of an individual test person.
2	<b>What are the three types of goals in testing?</b> BTL1 <ul style="list-style-type: none"> <li>• Business Goal</li> <li>• Technical Goal</li> <li>• Political Goal</li> </ul>
3	<b>Define the term policy.(Nov/Dec2016)</b> BTL1 A policy can be defined as a high-level statement of principle or course of action that is used to govern a set of activities in an organization.
4	<b>Define Test Plan.( Nov/Dec 2015)</b> BTL1 A Plan is a document that provides a frame work or approach for achieving a set of goals.
5	<b>List the various Test Plan components.(Nov/Dec2016)</b> BTL1 <ul style="list-style-type: none"> <li>• Test Plan identifier</li> <li>• Introduction</li> <li>• Items to be tested</li> <li>• Features to be tested</li> <li>• Pass/Fail criteria</li> <li>• Suspension &amp; Resumption criteria</li> <li>• Testing tasks Test environment</li> <li>• Risks &amp; Contingencies</li> <li>• Testing costs</li> <li>• Approvals</li> </ul>
6	<b>Define Features.</b> BTL1 Features may be described as distinguishing characteristics of a software component or system.
8	<b>What is the meaning of the term Pass / Fail Criteria?</b> BTL1 Given a test item and a test case, the tester must have a set of criteria to decide on whether the test has been passed or failed upon execution.
9	<b>What is Suspension &amp; Resumption criteria?</b> BTL1 The criteria to suspend and resume testing are described in the simplest of cases testing is suspended at the end of a working day and resumed the following morning.
10	<b>Define Work Breakdown Structure (WBS).</b> BTL1 A Work Break Down structure is a hierarchical or tree like representation of all the tasks that are required to complete a project.
11	<b>Define Risks &amp; Contingencies.</b> BTL1 Every testing effort has risks associated with it. Testing software with a high degree of critically, complexity, or a tight delivery deadline all impose risks that may have negative impacts on project goals.
12	<b>What is Cost Drive?</b> BTL1

	A Cost Driver can be described as a process or product factor that has an impact on overall project costs.
13	<b>What are the various components of the test plan?</b> AU Nov/Dec 2016 BTL1 <ul style="list-style-type: none"> <li>• Test Design Specification</li> <li>• Test Case Specification</li> <li>• Test Procedures specifications</li> </ul>
14	<b>Define Test Summary Report.</b> BTL1 This report is prepared when testing is complete. It is summary of the results of the testing efforts. It also becomes a part of the projects historical database and provides a basis for lessons learned as applied to future projects.
15	<b>List the skills needed by a Test specialist.</b> BTL1 <ul style="list-style-type: none"> <li>• Organizational and planning skills</li> <li>• The ability to keep track of and pay attention to details</li> <li>• The determination to discover and solve problems</li> <li>• The ability to mentor and train others</li> <li>• The ability to work with users and clients</li> <li>• The ability to think creatively</li> </ul>
16	<b>What is the use of V-model in testing?</b> BTL1 The V-model is model that illustrates how testing activities can be integrated in to each phase of the standard software life cycle.
17	<b>Write the WBS elements for testing.</b> BTL1 <ul style="list-style-type: none"> <li>• Project start-up</li> <li>• Management coordination</li> <li>• Tool selection</li> <li>• Test planning</li> <li>• Test design</li> <li>• Test development</li> <li>• Test execution</li> <li>• Test measurement, and monitoring</li> <li>• Test analysis and reporting</li> <li>• 10. Test process improvement</li> </ul>
18	<b>What is the function of Test Item Transmittal Report or Locating Test Items?</b> BTL2 Suppose a tester is ready to run tests on the data described in the test plan. We need to be able to locate the item and have knowledge of its current status. This is the function of the Test Item Transmittal Report. Each Test Item Transmittal Report has a unique identifier.
19	<b>Define Test Log.</b> BTL1 The Test log should be prepared by the person executing the tests. It is a diary of the events that take place during the test. It supports the concept of a test as a repeatable experiment.
20	<b>What are the Three critical groups in testing planning and test plan policy? ( April/May 2015)</b> BTL1 <ul style="list-style-type: none"> <li>• Managers:</li> <li>• Developers/Testers</li> <li>• Users/Clients</li> </ul>
21	<b>What is scenario Testing?</b> The process of giving the usage scenario of the system in the client's point of view and checking how the system reacts to it is called as scenario Testing.

22	<b>What are the information present in the Test Item Transmittal Report or Locating Test Items?BTL1</b> <ul style="list-style-type: none"> <li>• Version/revision number of the item</li> <li>• Location of the item</li> <li>• Person responsible for the item (the developer)</li> <li>• References to item documentation and test plan it is related to.</li> <li>• Status of the item</li> <li>• Approvals – space for signatures of staff who approve the transmittal.</li> </ul>
23	<b>What are the skills needed by a test specialist?BTL1</b> <ul style="list-style-type: none"> <li>• Personal and managerial Skills <ul style="list-style-type: none"> <li>• Organizational, and planning skills, work with others, resolve conflicts, mentor and train others, written /oral communication skills, think creatively.</li> </ul> </li> <li>• Technical Skills <ul style="list-style-type: none"> <li>• General software engineering principles and practices, understanding of testing principles and practices, ability to plan, design, and execute test cases, knowledge of networks, database, and operating System.</li> </ul> </li> </ul>
24	<b>Write the test term hierarchy?BTL2</b> <ul style="list-style-type: none"> <li>• Test Manager</li> <li>• Test leader</li> <li>• Test Engineer</li> <li>• Junior Test Engineer</li> </ul>
25	<b>Write the approaches to test cost Estimation?BTL2</b> <ul style="list-style-type: none"> <li>• The COCOMO model and heuristics</li> <li>• Use of test cost drivers</li> <li>• Test tasks</li> <li>• Tester/developer ratios</li> <li>• Expert judgment</li> </ul>
	<b>PART* B</b>
1	<b>Explain the role of the 3 critical groups in software testing. (13M)BTL2</b> <b>Answer:Page: 321 - Srinivasan &amp; Ramaswamy</b> <ol style="list-style-type: none"> <li>1. <b>Managers(4M)</b> <ul style="list-style-type: none"> <li>• Task forces,policies,standards</li> <li>• Planning</li> <li>• Resource allocation</li> <li>• Support for education and training</li> <li>• Interact with users</li> </ul> </li> <li>2. <b>Developers/ testers(5M)</b> <ul style="list-style-type: none"> <li>• Apply black and white box methods</li> <li>• Assist with test planning</li> <li>• Test at all levels</li> <li>• Train and mentor</li> <li>• Participate in task forces</li> <li>• Interact with users</li> </ul> </li> <li>3. <b>Users/clients(4M)</b> <ul style="list-style-type: none"> <li>• Specify requirements clearly</li> <li>• Participate in usability test</li> </ul> </li> </ol>

2	<p><b>Explain the various documents involved in reporting Test Results. (13M)BTL2</b></p> <p><b>Answer:</b> Page : 59 - Notes</p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Test log(1M)</li> <li>• Test log identifier(2M)</li> <li>• Description(1M)</li> <li>• Activity and event entities(1M)</li> <li>• Test incident report(3M)</li> <li>• Test incident report identifier(1M)</li> <li>• Summary(1M)</li> <li>• Impact(1M)</li> <li>• Test summary report(2M)</li> </ul>
3	<p><b>Explain the various Test Plan attachments? (13M)BTL2</b></p> <p><b>Answer:</b> Page : 381 - Srinivasan &amp; Ramaswamy</p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Test design specifications(4M)</li> <li>• Test case specifications(5M)</li> <li>• Test procedure specifications(4M)</li> </ul>
4	<p><b>Discuss in detail about the test plan components.(13M)(Nov/Dec 2016,Nov/Dec 2017)</b></p> <p><b>BTL2 Answer:</b> Page : 59 – Notes</p> <p><b>Test Plan Components(13M)</b></p> <ul style="list-style-type: none"> <li>• Test plan identifier(5M)</li> <li>• Introduction</li> <li>• Items to be tested</li> <li>• Features to be tested</li> <li>• Approach</li> <li>• Pass/fail criteria(4M)</li> <li>• Suspension and resumption criteria</li> <li>• Test deliverables</li> <li>• Testing tasks</li> <li>• Test environment</li> <li>• Responsibilities(4M)</li> <li>• Staffing and training needs</li> <li>• Scheduling</li> <li>• Risks and contingencies</li> <li>• Testing costs</li> <li>• Approvals</li> </ul>
5	<p><b>Evaluate the testing and debugging goals and policies in detail.(13M)(April/May 2017)BTL5</b></p> <p><b>Answer:</b> Page :62 - Notes</p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Debugging goal (4M)</li> <li>• Debugging policy(4M)</li> <li>• Testing Policy: Organization X(3M)</li> <li>• Debugging policy: Organization X(2M)</li> </ul>
6	<p><b>Describe Test planning in detail. (13M) BTL2</b></p> <p><b>Answer:</b> Page : 352 - Srinivasan &amp; Ramaswamy</p>

	<p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Planning(1M)</li> <li>• Milestone (1M)</li> <li>• Overall test objectives(2M)</li> <li>• What to test (Scope of the tests) (1M)</li> <li>• Who will test? (2M)</li> <li>• How to test? (2M)</li> <li>• When to test? (2M)</li> <li>• When to stop Testing? (2M)</li> </ul>
7	<p><b>Explain in detail about Mutation testing. (13M) (April/May 2017) BTL2</b></p> <p><b>Answer: Page : 58 - Notes</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Mutation testing is a method of software testing in which program or source code is deliberately manipulated(4M)</li> <li>• Followed by suite of testing against the mutated code(5M)</li> <li>• The mutations introduced to source code are designed to imitate common programming errors.(4M)</li> </ul>
8	<p><b>Discuss in detail about the various skills needed by test specialist.(13M) (Nov/dec2017)BTL2</b></p> <p><b>Answer:Page : 352 - Srinivasan &amp; Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• <b>Personal and managerial Skills(7M)</b> <ul style="list-style-type: none"> <li>➢ Organizational, and planning skills, work with others, resolve conflicts, mentor and train others, written /oral communication skills, think creatively.</li> </ul> </li> <li>• <b>Technical Skills(6M)</b> <ul style="list-style-type: none"> <li>• General software engineering principles and practices, understanding of testing principles and practices, ability to plan, design, and execute test cases, knowledge of networks, database, and operating System.</li> </ul> </li> </ul>
9	<p><b>Explain the organizational structure for testing in single product companies.(13M) BTL2 (April/May 2017)</b></p> <p><b>Answer:Page :321 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Exploits the rear loading nature of testing activities.(2M)</li> <li>• Enables Engineers to gain experience in all aspects of life cycle(4M)</li> <li>• Is amenable to the fact that the organization mostly has informal processes.(2M)</li> <li>• Some defects may be detected earlier.(3M)</li> <li>• Accountability for testing quality reduces.(1M)</li> <li>• Schedule pressures normally compromise testing.(1M)</li> </ul>
	<b>PART* C</b>
1	<p><b>Describe pesticide paradox and how bring in new people to look at the software helps solve it.(15M)BTL5</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <p>This is the situation that occurs if you continue to test (3M)</p> <ul style="list-style-type: none"> <li>• Software with the same tests or same people.(4M)</li> <li>• Eventually, the software seems to build up immunity to the test because no new bugs are found.(3M)</li> </ul>

	<ul style="list-style-type: none"> <li>• If you change the tests or bring in new testers ,you will find new bugs.(2M)</li> <li>• The bugs are already there,it's the new technique which made the bugs visible.(3M)</li> </ul>
2	<p><b>Why is the process of creating the test plan matters ,not the plan itself?(15M)BTL5</b></p> <p><b>Answer:Page: Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• Because all the issues and the questions defined in the test plan either impact or influenced by other project functional groups or team members.(4M)</li> <li>• Getting everyone to understand and agree to the contents of the plan is what matters.(4M)</li> <li>• Privately creating a paper document and putting it on a shelf is not just a waste of time, but also jeopardizes the project.(7M)</li> </ul>
3	<p><b>Justify the statement “ A schedule should be made to meet absolute dates ,so that there s no question when a testing task or phase is to start and when it is to end”.(15M)BTL6</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• The statement is false (3M)</li> <li>• Because testing depends so much on other aspects of the project(5M)</li> <li>• For example ,you can't test something until its coded), a test schedule is best made relative to the delivery status.(7M)</li> </ul>
4	<p><b>Name a few typical testing resources that should be considered when test planning.(15M)BTL6</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• People, Equipment, Offices, Labs , Software ,Outsourcing Companies and miscellaneous supplies.(3M)</li> <li>• What are the entrance and exit criteria?(4M)</li> <li>• The requirements must be met to move from one testing place to another.(3M)</li> <li>• A Phase can't be left until its exit criteria are met.(3M)</li> <li>• A new phase can't be entered until its entrance criteria are met.(2M)</li> </ul>

# JIT

<b>UNIT V TEST AUTOMATION</b>	
Software test automation – skill needed for automation – scope of automation – design and architecture for automation – requirements for a test tool – challenges in automation – Test metrics and measurements – project, progress and productivity metrics.	
<b>PART * A</b>	
1	<b>Define the term Project monitoring.</b> BTL1 <ul style="list-style-type: none"> <li>• Project Monitoring refers to activities and tasks managers engage in to periodically check the status of each project.</li> <li>• Reports are prepared that compare the actual work done to the work that was planned.</li> </ul>
2	<b>Define the term Project controlling.</b> BTL1 <p>Project Controlling consists of developing and applying a set of corrective actions to get a project on track when monitoring shows a deviation from what was planned.</p>
3	<b>Define Milestones. (Nov/Dec2016)</b> BTL1 <p>Milestones are tangible events that are expected to occur at a certain time in the project's lifetime. Managers use them to determine project status.</p>
4	<b>Differentiate version control and change control.</b> BTL2 <ul style="list-style-type: none"> <li>• Version Control combines procedures and tools to manage different versions of configuration objects that are created during software process.</li> <li>• Change control is a set of procedures to evaluate the need of change and apply the changes requested by the user in a controlled manner.</li> </ul>
5	<b>What are the goals of Reviewers?</b> BTL1 <ul style="list-style-type: none"> <li>• Identify problem components or components in the software artifact that need improvement.</li> <li>• Identify components of the software artifact that do not need improvement.</li> <li>• Identify specific errors or defects in the software artifact.</li> <li>• Ensure that the artifact conforms to organizational standards.</li> </ul>
6	<b>What are the benefits of a Review program?</b> BTL1 <ul style="list-style-type: none"> <li>• Higher quality software</li> <li>• Increased productivity</li> <li>• Increased awareness of quality issues</li> <li>• Reduced maintenance costs</li> <li>• Higher customer satisfaction</li> </ul>
7	<b>What are the Various types of Reviews?</b> BTL1 <ul style="list-style-type: none"> <li>• Inspections</li> <li>• Walk Throughs</li> </ul>
8	<b>Conclude on the need of Integration testing.(AU Nov/Dec2016)</b> BTL2 <ul style="list-style-type: none"> <li>• Component integration testing that checks the interconnections between various parts (components) in a product.</li> <li>• System integration testing that tests the connections between the product and external systems.</li> </ul>
9	<b>What is Inspections?</b> BTL1 <p>It is a type of review that is formal in nature and requires prereview preparation on the</p>

	part of the review team. The Inspection leader prepares is the checklist of items that serves as the agenda for the review.
10	<p><b>What is Walkthrough?(Nov/Dec 2017) BTL1</b></p> <p>It is a type of technical review where the producer of the reviewed material serves as the review leader and actually guides the progression of the review .It have traditionally been applied to design and code.</p>
11	<p><b>List out the members present in the Review Team.BTL1</b></p> <ul style="list-style-type: none"> <li>• SQA(Software Quality Assurance) staff</li> <li>• Testers</li> <li>• Developers</li> <li>• Users /Clients.</li> <li>• Specialists.</li> </ul>
12	<p><b>List the components of review plans.(AU April/May 2015)BTL1</b></p> <ul style="list-style-type: none"> <li>• Review Goals</li> <li>• Items being reviewed</li> <li>• Preconditions for the review.</li> <li>• Rolls, Team size, participants</li> <li>• Training requirements.</li> <li>• Review steps.</li> <li>• Time requirement</li> </ul>
13	<p><b>What are the advantages of review approach.BTL1</b></p> <p>There are two pass approach for detect detection.</p> <ul style="list-style-type: none"> <li>• Pass 1 has individuals first reading reviewed item</li> <li>• Pass 2 has the item read by the group as a whole.</li> </ul>
14	<p><b>What are the various roles in review program?BTL1</b></p> <ul style="list-style-type: none"> <li>• Review Leader</li> <li>• Review Recorder</li> <li>• Reader Reviewer</li> </ul>
15	<p><b>List the various review team membership constituencyReview Team Members.BTL1</b></p> <ul style="list-style-type: none"> <li>• SQA Staff</li> <li>• Testers</li> <li>• Developers</li> <li>• Users / Clients</li> <li>• Specialists</li> </ul>
16	<p><b>What are the various different types of software artifacts.BTL1</b></p> <ul style="list-style-type: none"> <li>• Requirement Reviews</li> <li>• Design Reviews</li> <li>• Code Reviews</li> <li>• Test Plan reviews</li> </ul>
17	<p><b>Define Change Control Board (CCB).BTL1</b></p> <ul style="list-style-type: none"> <li>• There are 2 aspects of change control – one is tool based, the other term based.</li> <li>• The team involved is called CCB.</li> </ul>
18	<p><b>Define Project monitoring.BTL1</b></p> <p>Project monitoring refers to the activities and tasks managers engage into periodically check the status of each project.Reports are prepared that compare the actual work done to the work that was planned or tracking.</p>

19	<p><b>Define Project Controlling.BTL1</b> It is the process of developing and applying a set of corrective actions to get a project on track when monitoring shows a deviation from what was planned.</p>
20	<p><b>Define Defect Removal Leverage (DRL).BTL1</b> This is a ratio of the defect detection rates from two review or test phases and can be expressed as</p> $DRL = \frac{\text{Defects / hour (review or test phase X)}}{\text{Defects / hour (review or test phase Y)}}$
21	<p><b>What are the various steps in the inspection process?BTL1</b></p> <ul style="list-style-type: none"> <li>• Entry Criteria</li> <li>• Initiation</li> <li>• Preparation</li> <li>• Inspection Meeting</li> <li>• Reporting results</li> <li>• Rework &amp; follow up</li> </ul>
22	<p><b>What is the Role of process in Software quality?BTL1</b></p> <ul style="list-style-type: none"> <li>• Capability Maturity Model.</li> <li>• Testing Maturity model ( TMM )</li> </ul>
23	<p><b>List the measurements and milestones for monitoring and controlling.BTL1</b></p> <ul style="list-style-type: none"> <li>• Measurements for monitoring testing status</li> <li>• Coverage measures</li> <li>• Test case development</li> <li>• Test execution</li> <li>• Test harness development</li> <li>• Measurements to monitor tester productivity</li> <li>• Measurements for monitoring testing costs</li> <li>• Measurements for monitoring errors, faults, and failures</li> <li>• Monitoring test effectiveness</li> </ul>
24	<p><b>Overview of the Testing Maturity Model(TMM)&amp; the test related activities that should be done for V-model architecture.BTL1</b></p> <ul style="list-style-type: none"> <li>• Test related issues</li> <li>• Benefits of test process improvement</li> <li>• Introduction to TMM</li> <li>• TMM levels</li> </ul>
25	<p><b>List the criteria for test completion.BTL1</b></p> <ul style="list-style-type: none"> <li>• All the planned tests that were developed have been executed and passed</li> <li>• All specified coverage goals have been met</li> <li>• The detection of a specific number of defects has been accomplished</li> <li>• The rates of defect detection for a certain time period have fallen below a specified</li> </ul>

	level , Fault seeding ratios are favorable
--	--

<b>PART * B</b>
-----------------

- |   |  |
|---|--|
| 1 | <b>Illustrate with a sketch describe the design and architecture for test automation. (13M)(Nov/Dec ,2016)BTL1</b> |
|---|--|

**Answer:Page : 396 - Srinivasan & Ramaswamy**

**Explanation:**

- External modules.(3M)
- Scenario and configuration file modules.(3M)
- Test cases and test framework modules.(3M)
- Tools and results modules.(2M)
- Report generator and report metrics modules.(2M)

- |   |   |
|---|---|
| 2 | <b>Explain the various generations of automations and the skills for each.(13M)(Nov/Dec,2017)BTL1</b> |
|---|---|

**Answer:Page :392 - Srinivasan & Ramaswamy**

**Explanation:**

- First Generation – Record and playback(4M)
- Second Generation – Data Driven(5M)
- Third Generation - Action Driven(4M)

- |   |   |
|---|---|
| 3 | <b>Explain the design and architecture of test automation and list the challenges.(13M)(April/May ,2017).BTL2</b> |
|---|---|

**Answer:Page :396 - Srinivasan & Ramaswamy**

**Explanation:**

- External modules.(2M)
- Scenario and configuration file modules.(2M)
- Test cases and test framework modules.(2M)
- Tools and results modules.(2M)
- Report generator and report metrics modules.(2M)
- Challenges(3M)
- Certain types of testing cannot be executed without automation.
- Automation means end to end not test execution alone.

- |   |  |
|---|--|
| 4 | <b>Discuss in detail about the controlling and monitoring: three critical views. (13M)BTL2</b> |
|---|--|

**Answer:Page : 71 – Notes**

**Explanation:**

- Measurements for monitoring testing status(1M)
- Coverage measures(1M)
- Test case development(2M)
- Test execution(1M)

	<ul style="list-style-type: none"> <li>• Test harness development(2M)</li> <li>• Measurements to monitor tester productivity(2M)</li> <li>• Measurements for monitoring testing costs(1M)</li> <li>• Measurements for monitoring errors, faults, and failures(1M)</li> <li>• Monitoring test effectiveness(2M)</li> </ul>
5	<p><b>Explain in detail about the role of reviews in testing software deliverables.(13M) BTL2</b></p> <p><b>Answer:Page : 68 - Notes</b></p> <p><b>Planning the Review (5M)</b></p> <ul style="list-style-type: none"> <li>• The role and responsibilities of the review leader</li> <li>• Identifying the deliverable to review and its review criteria</li> <li>• Developing review checklists for the reviewers based on requirements</li> <li>• Selecting the review team and assign review duties</li> </ul> <p><b>Conducting the Review (4M)</b></p> <ul style="list-style-type: none"> <li>• The role and responsibilities of the review leader</li> <li>• Inform the reviewers of their review duties, tasks, and schedule</li> <li>• Collect the reviews in a review meeting</li> <li>• Dealing with interpersonal issues</li> <li>• Common review pitfalls and how to avoid them</li> </ul> <p><b>Report and Follow-up on the Review(4M)</b></p> <ul style="list-style-type: none"> <li>• The role and responsibilities of the review leader</li> <li>• Compile the review findings into a single review report</li> <li>• Track review findings or issues</li> <li>• Follow-up on review findings or issues</li> </ul>
6	<p><b>Describe the various metrics and measurements in software testing and explain the various areas of metrics. (13M) (Nov/Dec 2016) BTL2</b></p> <p><b>Answer:Page: 420 - Srinivasan &amp; Ramaswamy</b></p> <p><b>Explanation:</b></p> <ul style="list-style-type: none"> <li>• Project metrics(2M)</li> <li>• Effort variance(3M)</li> <li>• Schedule Variance(3M)</li> <li>• Effort Distribution across phase(5M)</li> </ul>
1	<p><b>How will you differentiate tools and automation? Name the few benefits and drawbacks of using software test tools and automation. (15M)BTL6</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>• A testing tool will help you test .making it easier for you to perform a manual testing task.(3M)</li> <li>• Automation is also a tool but it will run without your intervention.(3M)</li> <li>• Think power saw and hammer building a house while the carpenter sleeps.(3M)</li> </ul> <p><b>Benefits:(3M)</b></p> <ul style="list-style-type: none"> <li>• Speed up the amount of time it takes to run your test process.</li> <li>• Precise and relentless.</li> </ul> <p><b>Drawbacks:(3M)</b></p>

	<ul style="list-style-type: none"> <li>Because software can change during the product's development, your test tools will be need to change.</li> <li>It is easy to rely on automation much.</li> </ul>
2	<p><b>If you were using metrics from the bug – tracking database to measure your progress or success at testing, why would just counting the number of bugs you find per day or computing your average find rate be an insufficient measure? (15M)BTL6</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>It does not tell the entire story. You could be testing the complex area of the software.(4M)</li> <li>Your area could have been written by the most experienced programmer.(4M)</li> <li>It could have been written by the least experienced programmer.(4M)</li> <li>The code that you are testing may already have been tested or may be brand new.(3M)</li> </ul>
3	<p><b>"The test team is responsible for the quality of the product" Does the statement make sense, Justify your answer with necessary explanation. (15M)BTL6</b></p> <p><b>Answer:Page : Appendix - Srinivasan Ramaswamy</b></p> <ul style="list-style-type: none"> <li>False! Testing looks for bugs .(7M)</li> <li>Testers didn't put the bugs in the product and can't guarantee when they are done testing that no more bugs exist.(8M)</li> </ul>

JIT