

Maria College of Engineering and Technology, Attoor
Department of M.Sc SOFTWARE ENGINEERING

Academic Year: 2013-2014 (Even)

DATA MINING & DATA WAREHOUSING (ESE514)

Two Mark Questions

UNIT-1

1. Define data mining?

Data mining is an extracting or “mining” knowledge from large amount of data.

2. What is data learning?

Data learning means the task of modern scientist to explain and predict.

3. Write the rules in basic structure of data warehouse?

Some specific rules for the basic structure of the data warehouse are

- Time Dependent**
- Non-Volatile**
- Subject oriented**
- Integrated**
-

4. What is metadata?

In setting up a data warehouse the end-user and administrator must have access to all information in tables and attributes. Metadata used by end-users for querying purposes.

5. Write the functions of data warehouse?

- To supply the information needed to make correct decisions.**
- Integration with data warehouse is Relationship between operational data, data marts and a data warehouse.**

6. Define multiprocessing machine and its type?

Very large database sites, multiprocessing machines are needed for data mining projects. Several types of multiprocessing machines are

- i.) Symmetric multiprocessing machine**
- ii.) Massively multiprocessing machine**

7. Write the methods of information and production factor?

The methods of information and production factor are

- Filtering Data
- Selecting Data
- Interpreting Data

8. What is self learning?

“A Self learning computer can generate programs itself enabling it to carry out new tasks.

9. Write the limitations of self learning?

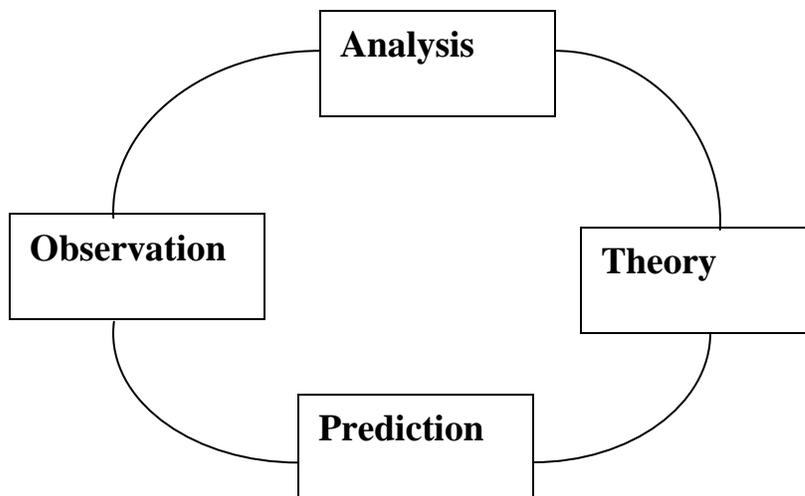
The limitations of self learning is

- Lack of creativity

10. What is data warehouse and why do we need it?

Rapid access in all kinds of information before you can make any logical decisions. Right choices for your organization is essential able to research the past and identify relevant trends.

11. Draw the empirical cycle of scientific research?



12. What is learning?

Learning denotes changes in the system that enables the system to do the same task more efficiently the next time. Learning is making useful changes or modifying what is being experienced.

13. Why machine learning is done?

- **To understand and improve the efficiency of human learning.**
- **To discover new things or structure that is unknown to human beings.**
- **To fill in skeletal or computer specifications about a domain.**

UNIT-2

1. Write the six stages of knowledge discovery process?

The six stages of knowledge discovery process are

- **Data selection**
- **Cleaning**
- **Enrichment**
- **Coding**
- **Data mining**
- **Reporting**

2. What is reporting?

Reporting means the result of data mining can take many forms. In KDD process the constant interplay between the selections of

- **Data**
- **Cleaning**
- **Data mining**
- **Reporting of results**

3. Write the various different techniques used for different purpose in data mining?

The various different techniques used for different purpose in data mining are

- **Query tools**
- **Statistical techniques**
- **Visualization**
- **Neural networks**
- **Association rule**
- **Genetic algorithm**

4. Write the advantages of OLAP tools?

- **Data mining algorithm not need in special form of storage**
- **Work directly in data stored in relational database**

5. What are Input nodes and Output nodes?

Input nodes mean receive the input signals. Output nodes means give the output signals

6. What are the nodes used in neural networks?

The nodes used in neural networks are

- **Input nodes**
- **Output nodes**

7. Write the different forms of the neural network technique?

The different forms of the neural network technique are

- **Perceptrons**
- **Back propagation networks**
- **Kohonen self organizing map**

8. What is Visualization?

Visualization is for depiction of data and to gain intuition about data being observed. It assists the analysts in selecting display formats, viewer perspectives and data representation schema.

9. Name some conventional visualization techniques?

Some conventional visualization techniques are

- **Histogram**
- **Relationship tree**
- **Bar charts**
- **Pie charts**
- **Tables etc**

10. Explain OLAP?

The general activity of querying and presenting text and number data from Data Warehouses, as well as a specifically dimensional style of querying and presenting that is exemplified by a number of “OLAP Vendours” .The OLAP vendours technology is non relational and is almost always biased on an explicit multidimensional cube of data. OLAP databases are also known as multidimensional cube of databases

11. Name some of the data mining applications?

- **Data mining for Biomedical and DNA data analysis**
- **Data mining for financial data analysis**
- **Data mining for the Retail industry**
- **Data mining for the Telecommunication industry**

12. What are the various OLAP operations?

The various OLAP operations are

- a) Roll-up
- b) Drill-down
- c) Slice

UNIT-3

1. What is data warehouse?

Data warehouse architecture is a repository of information collected from multiple sources and stored under a unified schema.

2. List some of the Data Warehouse tools?

- *OLAP (On Line Analytic Processing)
- *ROLAP (Relational OLAP)
- *End User Data Access tool
- *Ad Hoc Query tool
- *Data Transformation services
- *Replication

3. Write the difference between normalization and row splitting?

Rows splitting	Normalization
1.) One to one	1.) One to many
2.)	2.)

4. Define metadata?

In setting up a data warehouse the end-user and administrator must have access to all information in tables and attributes. Metadata used by end-users for querying purposes.

5. Write the steps to design a summary table?

- **Determine which dimensions are aggregated**
- **Determine the aggregation of multiple values**
- **Aggregate multiple facts into the sum table**
- **Determine the level of aggregation**
- **Determine the extent of embedding dimension data in the summary**
- **Design time into sum table**
- **Index the sum table**

6. When is data mart appropriate?

- **To speed up queries by reducing the volume of data to be scanned.**
- **To struct data in a form suitable for a user access tool.**
- **To partition data in order to impose access control categories**
- **To segment data into different hardware platforms.**

7. Write the components of data warehouse architecture?

The components of data ware house are

- **Load manager**
- **Ware house manager**
- **Reporting or query manager**

8. Define a data mart?

Data mart is a pragmatic collection of related facts, but does not have to be exhaustive or exclusive. A data mart is both a kind of subject area and an application. Data mart is a collection of numeric fact.

9. Write the Merits of Data Warehouse?

- * **Ability to make effective decisions from database**
- * **Better analysis of data and decision support**
- * **Discover trends and correlations that benefits business**
- * **Handle huge amount of data**

10. What are the characteristics of data warehouse?

The characteristics of data warehouse are

- * **Separate**
- * **Available**
- * **Integrated**
- * **Subject Oriented**
- * **Not Dynamic**
- * **Consistency**
- * **Iterative Development**
- * **Aggregation Performance**

11. How are association rules mined from large databases?

Association rule mining is a two-step process.

- Find all frequent item sets.
- Generate strong association rules from the frequent item sets

UNIT-4

1. Write the advantages of MPP?

The advantages of MPP are

- Nothing is shared
- They do not suffer from same restriction as
 - SMP
 - Cluster
 - Systems

2. Define hot backup?

Hot backup means any backup. i.e., not cold considered to be hot.

3. What is RAID technology?

RAID stands for redundant array of inexpensive disks.

4. What is a partial backup?

Partial backup means any backup. i.e., not complete.

5. What is documentation?

Documentation contains all the information gathered on

- Data classification
- User classification
- Network requirements
- Data movements and storage

6. Define complete backup?

Complete backup means entire database backed up at the same time. It includes

- All database data files
- Control files
- Journal files

7. What is SLA?

SLA stands for Service Level Agreement; It can be further divided into two categories

- **User requirements**
- **System requirements**

8. Define cold backup?

Cold backup means backup. i.e., taken while the database completely shutdown.

9. Define online backup?

Online backup means synonym for hot backup.

10. What are the tasks needed for overnight processing?

The tasks needed for overnight processing are

- **Data rollup**
- **Obtaining the data**
- **Data transformation**
- **Daily load**
- **Data cleanup**
- **Index creation**
- **Aggregation creation and maintenance**
- **Backup**
- **Data archiving**
- **Data mart fresh**

11. Define Relational databases?

A relational databases 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

12. What is clustering?

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters

6. What is the term Ad hoc?

Ad hoc means aggregations are those that are created to solve particular performance problems.

7. What is a passive monitoring tool?

Passive monitoring tools works by taking snapshots of statistics at various stages.

8. What are active monitoring tools?

Active monitoring tools gather data continuously.

9. What are the tools used in management tools?

Tools used in management tools such as

- Event manager**
- System manager**
- Configuration manager**
- Backup recovery manager**
- Database manager**

10. What are the events used in management tools?

Events such used in management tools as

- Running out of space on certain key disks**
- A processing dying**
- A process returning an error**
- Hardware failure**
- Management and monitoring tools**
- Scheduling software**