Q: Define data warehouse?

A : Data warehouse is a subject oriented, integrated, time-variant, and nonvolatile collection of data that supports management's decision-making process.

Q: What does subject-oriented data warehouse signify?

A : Subject oriented signifies that the data warehouse stores the information around a particular subject such as product, customer, sales, etc.

Q: List any five applications of data warehouse.

A : Some applications include financial services, banking services, customer goods, retail sectors, controlled manufacturing.

Q: What do OLAP and OLTP stand for?

A : OLAP is an acronym for **Online Analytical Processing** and OLTP is an acronym of Online Transactional Processing.

Q: What is the very basic difference between data warehouse and operational databases?

A : A data warehouse contains historical information that is made available for analysis of the business whereas an operational database contains current information that is required to run the business.

Q: List the Schema that a data warehouse system can implements.

A : A data Warehouse can implement star schema, snowflake schema, and fact constellation schema.

Q: What is Data Warehousing?

A : Data Warehousing is the process of constructing and using the data warehouse.

Q: List the process that are involved in Data Warehousing.

A : Data Warehousing involves data cleaning, data integration and data consolidations.

Q: List the functions of data warehouse tools and utilities.

A : The functions performed by Data warehouse tool and utilities are Data Extraction, Data Cleaning, Data Transformation, Data Loading and Refreshing.

Q: What do you mean by Data Extraction?

A : Data extraction means gathering data from multiple heterogeneous sources.

Q: Define metadata?

A : Metadata is simply defined as data about data. In other words, we can say that metadata is the summarized data that leads us to the detailed data.

Q: What does Metadata Respiratory contain?

A : Metadata respiratory contains definition of data warehouse, business metadata, operational metadata, data for mapping from operational environment to data warehouse, and the algorithms for summarization.

Q: How does a Data Cube help?

A : Data cube helps us to represent the data in multiple dimensions. The data cube is defined by dimensions and facts.

Q: Define dimension?

A : The dimensions are the entities with respect to which an enterprise keeps the records.

Q: Explain data mart.

A : Data mart contains the subset of organization-wide data. This subset of data is valuable to specific groups of an organization. In other words, we can say that a data mart contains data specific to a particular group.

Q: What is Summary Information?

A : Summary Information is the area in data warehouse where the predefined aggregations are kept.

Q: What does the Query Manager responsible for?

A : Query Manager is responsible for directing the queries to the suitable tables.

Q: List the types of OLAP server

A : There are four types of OLAP servers, namely Relational OLAP, Multidimensional OLAP, Hybrid OLAP, and Specialized SQL Servers.

Q: Which one is faster, Multidimensional OLAP or Relational OLAP?

A : Multidimensional OLAP is faster than Relational OLAP.

Q: List the functions performed by OLAP.

A : OLAP performs functions such as roll-up, drill-down, slice, dice, and pivot.

Q: How many dimensions are selected in Slice operation?

A : Only one dimension is selected for the slice operation.

Q: How many dimensions are selected in dice operation?

A : For dice operation two or more dimensions are selected for a given cube.

Q: How many fact tables are there in a star schema?

A : There is only one fact table in a star Schema.

Q: What is Normalization?

A : Normalization splits up the data into additional tables.

Q: Out of star schema and snowflake schema, whose dimension table is normalized?

A : Snowflake schema uses the concept of normalization.

Q: What kind of costs are involved in Data Marting?
A : Data Marting involves hardware & software cost, network access cost, and time cost.

Q: Define data cleaning

A: Data cleaning means removing the inconsistent data or noise and collecting necessary information

Q: Define data mining

A: Data mining is a process of extracting or mining knowledge from huge amount of data

Q: What are the steps in the data mining process?

A: Data cleaning

Data integration

Data selection

Data transformation

Data mining

Pattern evaluation

Q: What is Descriptive and predictive data mining?

A: Descriptive data mining describes the data set in a concise and summertime manner and Presents interesting general properties of the data. Predictive data mining analyses the data in order to construct one or set of models and attempts to predict the behaviour of new data sets.

Q: What is Association rule?

A: Association rule finds interesting association or correlation relationships among a large set of data items, which is used for decision-making processes. Association rules analyzes buying patterns that are frequently associated or purchased together.

Q: Define support.

A: Support is the ratio of the number of transactions that include all items in the antecedent and consequent parts of the rule to the total number of transactions. Support is an association rule interestingness measure.

Q: Define Confidence.

A: Confidence is the ratio of the number of transactions that include all items in the consequent as well as antecedent to the number of transactions that include all items in antecedent. Confidence is an association rule interestingness measure.

Q: How are association rules mined from large databases?

A: Association rule mining is a two-step process.

Find all frequent itemsets.

Generate strong association rules from the frequent itemsets.

Q: Explain OLAP?

A: 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 Vendors" .The OLAP vendors technology is no relational and is almost always biased on an explicit multidimensional cube of data. LAP databases are also known as multidimensional cube of databases

Q: What is the difference between "supervised" and unsupervised" learning scheme.

A: In data mining during classification the class label of each training sample is provided, this type of training is called supervised learning. E.g. Classification

in unsupervised learning the class label of each training sample is not known and the member or set of classes to be learned may not be known in advance. E.g. Clustering

Q: Explain the various OLAP operations.

A: a) Roll-up: The roll-up operation performs aggregation on a data cube, either by

   Climbing up a concept hierarchy for a dimension.

  b) Drill-down: It is the reverse of roll-up. It navigates from less detailed data to more

   Detailed data.

  c) Slice: Performs a selection on one dimension of the given cube, resulting in a

   Sub cube.

Q: Why is data quality so important in a data warehouse environment?

A: Data quality is important in a data warehouse environment to facilitate decision-making. In order to support decision-making, the stored data should provide information from a historical perspective and in a summarized manner.

Q: Explain the architecture of data warehouse.

A: Steps for the design and construction of DW

Top-down view

Data source view

Data warehouse view

Business query view

3tier DW architecture

Q: What is Data Mining? Explain the steps in Knowledge Discovery?

A: Data mining refers to extracting or mining knowledge from large amount of data. The steps in knowledge discovery are:

Data cleaning

Data integration

Data selection

Data transformation

Data mining

Pattern Evolution

Knowledge Discovery

1. **Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation?**

    a. KDD process
    b. ETL process
    c. KTL process
    d. MDX process
    e. None of the above

    ANSWER : KDD process

2. **Data modeling technique used for data marts is**

    a. Dimensional modelling
    b. ER model
    c. Extended ER model
    d. physical model
    e. Logical model.

    ANSWER : 1 Dimensional modelling

3. **Which of the following employees data mining techniques to analyze the intent of a user query, provided additional generalized or associated information relevant to the query?**
    a. Iceberg query method
    b. Data analyser
    c. Intelligent query answering

d. DBA

e. Query parser.

4. **Which of the following statements is true?**

a. A fact table describes the transactions stored in a DWH

b. A fact table describes the granularity of data held in a DWH

c. The fact table of a data warehouse is the main store of descriptions of the transactions stored in a DWH

d. The fact table of a data warehouse is the main store of all of the recorded transactions over time

e. A fact table maintains the old records of the database

5. **Data warehouse bus matrix is a combination of**

a. Dimensions and data marts

b. Dimensions and facts

c. Dimensions and detailed facts

d. Facts and data marts

e. All above

6. **Which of the following is not related to dimension table attributes?**

a. Verbose

b. Descriptive

c. Equally unavailable

d. Complete

e. Indexed

7. **The Synonym for data mining is**

a. Data warehouse

b. Data warehouse

c. ETL

d. Business intelligence

e. OLAP

8. **An OLAP tool provides for**

a. Multidimensional analysis

b. Roll-up and drill-down

c. Slicing and dicing

d. Setting up only relations.

e. Rotation

9. **The process of removing the deficiencies and loopholes in the data is called as #2. The process of removing the deficiencies and loopholes in the data is called as**
   a. Aggregation of data
   b. Extracting of data
   c. Cleaning up of data
   d. Loading of data
   e. Compression of data.

10. **.Which one manages both current and historic transactions?**
    a. OLAP
    b. OLTP
    c. XML
    d. Spread sheet
    e. All above

11. **Which of the following is the most important when deciding on the data structure of a data mart?**
    (a) XML data exchange standards
    (b) Data access tools to be used
    (c) Metadata naming conventions
    (d) Extract, Transform, and Load (ETL) tool to be used
    (e) All (a), (b), (c) and (d) above.

Answer: (B) Data access tools to be used when deciding on the data structure of a data mart.

12. **. Which of the following is the collection of data objects that are similar to one another within the same group?**
    (a) Partitioning
    (b) Grid
    (c) Cluster
    (d) Table
    (e) Data source.

Answer: (c) Cluster is the collection of data objects that are similar to one another within the same group.

13. **A warehouse architect is trying to determine what data must be included in the warehouse. A meeting has been arranged with a business analyst to understand the data requirements, which of the following should be included in the agenda?**
    (a) Number of users
    (b) Corporate objectives
    (c) Database design

(d) Routine reporting

(e) Budget.

Answer: (d) Routine reporting should be included in the agenda

14. A Business Intelligence system requires data from:

(a) Data warehouse

(b) Operational systems

(c) All possible sources within the organization and possibly from external sources

(d) Web servers

(e) Database servers.

Answer: (a) Data warehouse

15. Data mining application domains are

(a) Biomedical

(b) DNA data analysis

(c) Financial data analysis

(d) Retail industry and telecommunication industry

(e) All (a), (b), (c) and (d) above.

Answer: (e)

16. In a data warehouse, if D1 and D2 are two conformed dimensions, then

(a) D1 may be an exact replica of D2

(b) D1 may be at a rolled up level of granularity compared to D2

(c) Columns of D1 may be a subset of D2 and vice versa

(d) Rows of D1 may be a subset of D2 and vice versa

(e) All (a), (b), (c) and (d) above.

Answer: (a) D1 may be an exact replica of D2

17. Which of the following is not an ETL tool?

(a) Informatica

(b) Oracle warehouse builder

(c) Datastage

(d) Visual studio

(e) DT/studio

Answer: (d) Visual studio

18. Which of the following is/are the Data mining tasks?

(a) Regression

(b) Classification

(c) Clustering

(d) inference of associative rules

(e) All (a), (b), (c) and (d) above.

Answer: (e) All (a), (b), (c) and (d) above

19. **Which of the following tools a business intelligence system will have?**
    (a) OLAP tool
    (b) Data mining tool
    (c) Reporting tool
    (d) Both(a) and (b) above
    (e) (a), (b) and (c) above.

Answer: (a) OLAP tool

20. **The apriori property means**
    a. If a set cannot pass a test, all of its supersets will fail the same test as well
    b. To improve the efficiency the level-wise generation of frequent item sets
    c. f a set can pass a test, all of its supersets will fail the same test as well
    d. To decrease the efficiency the level-wise generation of frequent item sets
    e. All (opt1), (opt2), (opt3) and (opt4) above

ANSWER: 2 to improve the efficiency the level-wise generation of frequent item sets

21. **What is/are the different types of Meta data?**
    I. Administrative.
    II. Business.
    III. Operational.
       a. Only (I) above
       b. Both (II) and (III) above
       c. Both (I) and (II) above
       d. Both (I) and (III) above
       e. All (I), (II) and (III) above

ANSWER: All (I), (II) and (III) above.