1. NAME SOME SPECIFIC APPLICATION ORIENTED DATABASES.
   - Spatial databases,
   - Time-series databases,
   - Text databases and multimedia databases.

2. DEFINE SPATIAL DATABASES.
   Spatial databases contain spatial-related information. Such databases include geographic(map) databases, VLSI chip design databases, and medical and satellite image databases. Spatial data may be represented in raster format, consisting of n-dimensional bit maps or pixel maps.

3. WHAT IS TEMPORAL DATABASE?
   Temporal database store time related data. It usually stores relational data that include time related attributes. These attributes may involve several time stamps, each having different semantics.

4. WHAT IS TIME-SERIES DATABASES?
   A Time-Series database stores sequences of values that change with time, such as data collected regarding the stock exchange.

5. 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

6. WHAT IS APRIORI ALGORITHM?
   Apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules using prior knowledge. Apriori algorithm uses prior knowledge of frequent item set properties and it employs an iterative approach known as level-wise search where k-item sets are used to explore (k+1)-item sets.

7. DEFINE SPATIAL VISUALIZATION
   Spatial visualization depicts actual members of the population in their feature space.

8. WHAT ARE THE GOALS OF TIME SERIES ANALYSIS?
   - Finding Patterns in the data
   - Predicting future values

9. WHAT ARE THE CONTRIBUTION OF DATA MINING TO DNA ANALYSIS?
   - Semantic integration of heterogeneous, distributed genome databases
   - Similarity search and comparison among DNA sequences
   - Association analysis: identification of co-occurring gene sequences
Path analysis: linking genes to different stages of disease development
Visualization tools and genetic data analysis

10. 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

11. NAME SOME CONVENTIONAL VISUALIZATION TECHNIQUES
- Histogram
- Relationship tree
- Bar charts
- Pie charts
- Tables etc.

12. GIVE THE FEATURES INCLUDED IN MODERN VISUALIZATION TECHNIQUES
- Morphing
- Animation
- Multiple simultaneous data views
- Drill-Down
- Hyperlinks to related data source

13. DEFINE CONVENTIONAL VISUALIZATION
Conventional visualization depicts information about a population and not the population data itself.

14. NAME SOME EXAMPLES OF DATA MINING IN RETAIL INDUSTRY?
- Design and construction of data warehouses based on the benefits of data mining
- Multidimensional analysis of sales, customers, products, time and region
- Analysis of the effectiveness of sales campaigns
- Customer retention-analysis of customer loyalty
- Purchase recommendation and cross-reference of item

15. HOW CAN DATA VISUALIZATION HELP IN DECISION-MAKING?
Data visualization helps the analyst gain intuition about the data being observed. Visualization applications frequently assist the analyst in selecting display formats, viewer perspective and data representation schemas that faster deep intuitive understanding thus facilitating decision-making.

16. WHAT DO YOU MEAN BY HIGH PERFORMANCE DATA MINING?
Data mining refers to extracting or mining knowledge. It involves integration of techniques from multiple disciplines like database technology, statistics, and machine learning, neural networks, etc. When it involves techniques from high performance computing it is referred as high performance data mining.
17. WHAT IS MEANT BY SECURITY SAFEGUARDS?
   The personal data should be protected by reasonable security safeguards against such
   risks as loss or unauthorized access, destruction, use, modification, or disclosure of data.

18. 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.

19. WHAT IS MEANT BY VISUAL DATA MINING?
   Visual data mining is an effective way to discover knowledge from huge amounts of data.
   The systematic study and development of visual data mining techniques will facilitate the
   promotion and use of data mining as a tool for data analysis.

20. DEFINE TIME SERIES ANALYSIS
   There are many statistical techniques for analyzing time-series data, such as auto
   regression methods, unvaried ARIMA (autoregressive integrated moving average) modeling, and
   long-memory time-series modeling.

21. WHAT ARE THE CLASSIFICATIONS OF TOOLS FOR DATA MINING?
   - Commercial Tools
   - Public domain Tools
   - Research prototypes

22. WHAT ARE COMMERCIAL TOOLS?
   Commercial tools can be defined as the following products and usually are associated with the
   consulting activity by the same company:
   1. ‘Intelligent Miner’ from IBM
   2. ‘SAS’ System from SAS Institute
   3. ‘Thought’ from Right Information Systems. Etc

23. WHAT ARE PUBLIC DOMAIN TOOLS?
   Public domain Tools are largely freeware with just registration fees: ’Brute’ from
   University of Washington. ‘MC++’ from Stanford university, Stanford, California.

24. WHAT ARE RESEARCH PROTOTYPES?
   Some of the research products may find their way in to commercial market: ‘DB Miner’ from
   Simon Fraser University, British Columbia, ‘Mining Kernel System’ from
   University of Ulster, North Ireland.

25. WHAT IS THE DIFFERENCE BETWEEN GENERIC SINGLE-TASK TOOLS
    AND GENERIC MULTI-TASK TOOLS?
   Generic single-task tools generally use neural networks or decision trees. They cover
   only the data mining part and require extensive pre-processing and post processing steps.
   Generic multi-task tools offer modules for pre-processing and post processing steps and also
   offer a broad selection of several popular data mining algorithms as clustering.
26. WHAT ARE THE AREAS IN WHICH DATA WAREHOUSES ARE USED IN PRESENT AND IN FUTURE?

The potential subject areas in which data warehouses may be developed at present and also in future are

1. Census data:
The registrar general and census commissioner of India decennially compiles information of all individuals, villages, population groups, etc. This information is wide ranging such as the individual slip. A compilation of information of individual households, of which a database of 5% sample is maintained for analysis. A data warehouse can be built from this database upon which OLAP techniques can be applied, Data mining also can be performed for analysis and knowledge discovery

2. Prices of Essential Commodities
The ministry of food and civil supplies, Government of India compiles daily data for about 300 observation centers in the entire country on the prices of essential commodities such as rice, edible oil etc. A data warehouse can be built for this data and OLAP techniques can be applied for its analysis.

27. WHAT ARE THE OTHER AREAS FOR DATA WAREHOUSING AND DATA MINING?

- Agriculture
- Rural development
- Health
- Planning
- Education
- Commerce and Trade

28. SPECIFY SOME OF THE SECTORS IN WHICH DATA WAREHOUSING AND DATA MINING ARE USED?

- Tourism
- Program Implementation
- Revenue
- Economic Affairs
- Audit and Accounts

29. DESCRIBE THE USE OF DBMINER.
Used to perform data mining functions, including characterization, association, classification, prediction and clustering.

30. APPLICATIONS OF DBMINER.
The DBMiner system can be used as a general-purpose online analytical mining system for both OLAP and data mining in relational database and data warehouses. Used in medium to large relational databases with fast response time.

31. GIVE SOME DATA MINING TOOLS.

- DBMiner
- GeoMiner
Multimedia miner
WeblogMiner

32. MENTION SOME OF THE APPLICATION AREAS OF DATA MINING
- DNA analysis
- Financial data analysis
- Retail Industry
- Telecommunication industry
- Market analysis
- Banking industry
- Health care analysis.

33. DIFFERENTIATE DATA QUERY AND KNOWLEDGE QUERY
   A data query finds concrete data stored in a database and corresponds to a basic retrieval statement in a database system. A knowledge query finds rules, patterns and other kinds of knowledge in a database and corresponds to querying database knowledge including deduction rules, integrity constraints, generalized rules, frequent patterns and other regularities.

34. DIFFERENTIATE DIRECT QUERY ANSWERING AND INTELLIGENT QUERY ANSWERING.
   Direct query answering means that a query answers by returning exactly what is being asked. Intelligent query answering consists of analyzing the intent of query and providing generalized, neighborhood, or associated information relevant to the query.

35. DEFINE VISUAL DATA MINING
   Discovers implicit and useful knowledge from large data sets using data and knowledge visualization techniques. Integration of data visualization and data mining.

36. WHAT DOES AUDIO DATA MINING MEAN?
   Uses audio signals to indicate patterns of data or the features of data mining results. Patterns are transformed into sound and music.
   To identify interesting or unusual patterns by listening pitches, rhythms, tune and melody. Steps involved in DNA analysis Semantic integration of heterogeneous, distributed genome databases Similarity search and comparison among DNA sequences Association analysis: Identification of co-occurring gene sequences Path analysis: Linking genes to different stages of disease development Visualization tools and genetic data analysis

37. WHAT ARE THE FACTORS INVOLVED WHILE CHOOSING DATA MINING SYSTEM?
   Data types
   System issues
   Data sources
   Data Mining functions and methodologies
   Coupling data mining with database and/or data warehouse systems
   Scalability
   Visualization tools
   Data mining query language and graphical user interface.
38. DEFINE DMQL
Data Mining Query Language
It specifies clauses and syntaxes for performing different types of data mining tasks for example data classification, data clustering and mining association rules. Also it uses SQL-like syntaxes to mine databases.

39. DEFINE TEXT MINING
Extraction of meaningful information from large amounts free format textual data. Useful in Artificial intelligence and pattern matching Also known as text mining, knowledge discovery from text, or content analysis.

40. WHAT DOES WEB MINING MEAN
Technique to process information available on web and search for useful data. To discover web pages, text documents, multimedia files, images, and other types of resources from web. Used in several fields such as E-commerce, information filtering, fraud detection and education and research.

41. DEFINE SPATIAL DATA MINING.
Extracting undiscovered and implied spatial information. Spatial data: Data that is associated with a location Used in several fields such as geography, geology, medical imaging etc.

42. EXPLAIN MULTIMEDIA DATA MINING.
Mines large data bases. Does not retrieve any specific information from multimedia databases Derive new relationships, trends, and patterns from stored multimedia data mining. Used in medical diagnosis, stock markets, Animation industry, Airline industry, Traffic management systems, Surveillance systems etc.

43. WHAT IS HYPOTHESIS TESTING? EXPLAIN THE CONCEPTS OF NULL HYPOTHESIS AND ALTERNATIVE HYPOTHESIS?
A statistical discordancy test examines two hypotheses: a working hypothesis and an alternative hypothesis. A working hypothesis, \( H \), is a statement that the entire data set of \( n \) objects comes from an initial distribution Model. An alternative hypothesis, \( H \), which states that \( o_i \) comes from another distribution model

44. WRITE SHORT NOTES ON TEXT MINING
- Extraction of meaningful information from large amounts free format textual data.
- Useful in Artificial intelligence and pattern matching
- Also known as text mining, knowledge discovery from text, or content analysis.

45. WHAT KIND OF ASSOCIATION CAN BE MINED FOR MULTIMEDIA DATA?
- Associations between image content and non image content features
- Associations among image contents that are not related to spatial relationships
- Associations among image contents related to spatial relationships
46. EXPLAIN THE METHOD OF LEAST SQUARE METHOD

These coefficients can be solved for by the method of least squares, which estimates the best-fitting straight line as the one that minimizes the error between the actual data and the estimate of the line.

\[ w_1 = \frac{\sum_{i=1}^{D} (y_i - \bar{y})(\tilde{y} - \tilde{x})}{\sum_{i=1}^{D} (y_i - \tilde{x})^2} \]

47. DEFINE THE SENSITIVITY AND SPECIFICITY MEASURES.

Sensitivity is also referred to as the true positive (recognition) rate (that is, the proportion of positive tuples that are correctly identified), while specificity is the true negative rate (that is, the proportion of negative tuples that are correctly identified).

\[ \text{sensitivity} = \frac{t_{pos}}{p_{pos}} \]
\[ \text{specificity} = \frac{t_{neg}}{n_{neg}} \]

48. DEFINE THE PRECISION AND RECALL MEASURE IN TEXT MINING

**Precision:** This is the percentage of retrieved documents that are in fact relevant to the query (i.e., “correct” responses). It is formally defined as

\[ \text{precision} = \frac{|\text{Relevant} \cap \text{Retrieved}|}{|\text{Retrieved}|} \]

**Recall:** This is the percentage of documents that are relevant to the query and were, in fact, retrieved. It is formally defined as

\[ \text{recall} = \frac{|\text{Relevant} \cap \text{Retrieved}|}{|\text{Relevant}|} \]

49. DEFINE STOP LIST

To avoid indexing useless words, a text retrieval system often associates a stop list with a set of documents. A stop list is a set of words that are deemed “irrelevant.” For example, a, the, of, for, with, and so on are stop words, even though they may appear frequently. Stop lists may vary per document set.

50. WHAT IS MEANT BY AUTHORITATIVE WEB PAGES

Suppose you would like to search for Web pages relating to a given topic, such as financial investing. In addition to retrieving pages that are relevant, you also hope that the pages retrieved will be of high quality, or authoritative on the topic.

51. WHAT IS A WEB-LOG ENTRY?

A Web server usually registers a (Web) log entry, or Web log entry, for every access of a Web page. It includes the URL requested, the IP address from which the request originated, and a timestamp.
52. WHAT ARE SET VALUED ATTRIBUTES?
A set-valued attribute may be of homogeneous or homogeneous type. Typically set-valued data can be generalized by
(1) Generalization of each value in the set to its corresponding higher-level concept
(2) Derivation of the general behavior of the set, such as number of elements in the set, the types or values ranges in the set, the weighted average for numerical data, or the major clusters formed by the set.

53. DEFINE MULTIMEDIA DATA BASE SYSTEM
A multimedia data base may contain complex texts, graphics, images, video fragments, maps, voice, music, and other forms of audio/video information. Multimedia data are typically stored as sequences of bytes with variable length and segments of data are linked together or indexed in a multidimensional way for easy reference.

54. DEFINE –SPATIAL DATA MINING
Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial data base. Such mining demands an integration of data mining with spatial data base technologies.

55. WHAT ARE THE TWO TYPES OF MEASURES IN SPATIAL DATA CUBE?
1. Numerical Measure
2. Spatial measure

56. WHAT ARE THE THREE POSSIBLE CHOICEIN REGARD TO THE COMPUTATION OF SPATIAL MEASURES IN SPATIAL DATA CUBE CONSTRUCTION
1. A non spatial dimension
2. A spatial-to-non spatial dimension
3. A Spatial to Spatial dimension
57. DEFINE –DESCRIPTION BASED RETRIEVAL SYSTEM
Description based retrieval system, which build indices and perform object-retrieval based on image descriptions, such as keywords, captions, size and time of creation

58. DEFINE-CONTENT BASED RETRIEVAL SYSTEM
This supports retrieval based on image content, such as color histogram, texture, pattern image topology and shape of objects layouts and locations with in the image

59. WHAT ARE THE THREE CATEGORIES OBSERVED IN MINING ASSOCIATIONS IN THE MULTIMEDIA DATA?
1. Association between image content and non image content features
2. Associations among image contents that are not related to spatial relationship
3. Associations among image contents related to spatial relationships