Study Regular Notes For IBPS IT Officer Exam

1)GIF (Graphics Interchange Format) is a creation of CompuServe and is used to store multiple bitmap images in a single file for exchange between platforms and systems. In terms of number of files in existence, GIF is perhaps the most widely used format for storing multibit graphics and image data. Even a quick peek into the graphics file section of most BBSs and file archives seems to prove this true. Many of these are high-quality images of people, landscapes, cars, astro photographs, and anthropometric gynoidal data (you guess what that is). Shareware libraries and BBSs are filled with megabytes of GIF images.The vast majority of GIF files contain 16-color or 256-color near-photographic quality images. Gray-scale images, such as those produced by scanners, are also commonly stored using GIF, although monochrome graphics, such as clip art and document images, rarely are.GIF is not associated with any particular software application. GIF also was not created for any particular software application need, although most software applications that read and write graphical image data, such as paint programs, scanner and video software, and most image file display and conversion programs, usually support GIF. GIF was instead intended to allow the easy interchange and viewing of image data stored on local or remote computer systems. The original version of the GIF format was called **87a**.In 1989, CompuServe released an enhanced version, called **89a**, which added support for animation delays (multiple images in a stream were already supported in 87a), transparent background colors, and storage of application-specific metadata.

File Organization

GIF is different from many other common bitmap formats in the sense that it is stream-based. It consists of a series of data packets, called *blocks*, along with additional protocol information. Because of this arrangement, GIF files must be read as if they are a continuous stream of data. The various blocks and sub-blocks of data defined by GIF may be found almost anywhere within the file. This uncertainty makes it difficult to encapsulate every possible arrangement of GIF data in the form of C structures.

There are a number of different data block categories, and each of the various defined blocks falls into one of these categories. In GIF terminology, a Graphics Control Extension block is a type of Graphics Control block, for instance. In like manner, Plain Text Extension blocks and the Local Image Descriptor are types of Graphic Rendering blocks. The bitmap data is an Image Data block. Comment Extension and Application Extension blocks are types of Special Purpose blocks.

Blocks, in addition to storing fields of information, can also contain sub-blocks. Each data sub-block begins with a single count byte, which can be in the range of 1 to 255 and indicates the number of data bytes that follow the count byte. Multiple sub-blocks may occur in a contiguous grouping (count byte, data bytes, count byte, data bytes, and so on). A sequence of one or more data sub-blocks is terminated by a count byte with a value of zero.

The GIF format is capable of storing bitmap data with pixel depths of 1 to 8 bits. Images are always stored using the RGB color model and palette data. GIF is also capable of storing multiple images per file, but this capability is rarely utilized, and the vast majority of GIF files contain only a single image. Most GIF file viewers do not, in fact, support the display of multiple image GIF files or may display only the first image stored in the file. For these reasons, we recommend not creating applications that rely on multiple images per file, even though the specification allows this.

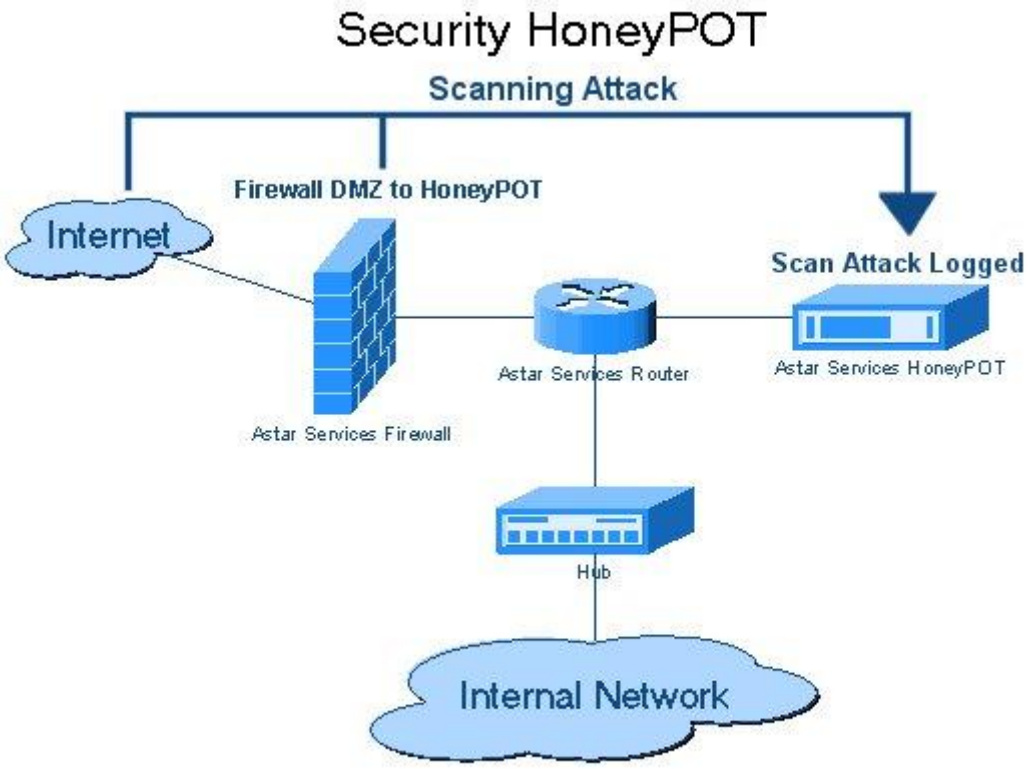
GIF files do not compress well when stored using file archivers such as pkzip and zoo. This is because the image data found in every GIF file is always compressed using the LZW (Lempel-Ziv-Welch) encoding scheme, the same compression algorithm used by most file archivers. (See the sidebar about LZW at the beginning of this article.) Compressing a GIF file is therefore a redundant operation, which rarely results in smaller files and is usually not worth the time and effort involved in the attempt.Normally when LZW-encoded image data is stored in a graphics file format, it is arranged as a continuous stream of data that is read from beginning to end. The GIF format, however, stores encoded image data as a series of data sub-blocks.Each data sub-block begins with a count byte. The value of the count byte may range from 1 to 255 and indicates

the number of data bytes in the sub-block. The data blocks immediately follow the count byte. A contiguous group of data blocks is terminated by a byte with a zero value. This may be viewed as either a terminator value or as a sub-block with a count byte value of zero; in either case, it indicates that no data bytes follow.Because GIF files do not contain a contiguous stream of LZW-encoded data, each sub-block must be read and the data sent to an LZW decoder. Most sub-blocks storing image data will be 255 bytes in length, so this is an excellent maximum size to use for the buffer that will hold the encoded image data. Also, the LZW encoding process does not keep track of where each scan line begins and ends. It is therefore likely that one scan line will end and another begin in the middle of a sub-block of image data.The format of the decoded GIF image data is fairly straightforward. Each pixel in a decoded scan line is always one byte in size and contains an index value into either a Global or Local Color Table. Although the structure of the GIF format is quite capable of storing color information directly in the image data (thus bypassing the need for a color table), the GIF specification does not specify this as a possible option. Therefore, even 1-bit image data must use 8-bit index values and a 2-entry color table.

GIF image data is always stored by scan line and by pixel. GIF does not have the capability to store image data as planes, so when GIF files are displayed using plane-oriented display adapters, quite a bit of buffering, shifting, and masking of image data must first occur before the GIF image can be displayed.The scan lines making up the GIF bitmap image data are normally stored in consecutive order, starting with the first row and ending with the last. The GIF format also supports an alternate way to store rows of bitmap data in an interlaced order. Interlaced images are stored as alternating rows of bitmap data. If you have ever viewed a GIF file that appeared on the screen as a series of four "wipes" that jumped across the screen as the image was displayed, you were viewing an interlaced GIF file.

2)A **honeypot** is a computer system that is set up to act as a decoy to lure cyber attackers, and to detect, deflect or study attempts to gain unauthorized access to information systems. Generally, it consists of a computer, applications, and data that simulate the behavior of a real system that appears to be part of a

network but is actually isolated and closely monitored. All communications with a honeypot are considered hostile, as there's no reason for legitimate users to access a honeypot. Viewing and logging this activity can provide an insight into the level and types of threat a network infrastructure faces while distracting attackers away from assets of real value.



Production honeypots are placed inside a production network with other production servers in the role of a decoy as part of a network intrusion detection system (IDS). They are designed to appear real and contain information or a resource of value with which to attract and occupy hackers. This ties up the attacker's time and resources, hopefully giving administrators time to assess and mitigate any vulnerabilities in their actual production systems. The information gathered from the honeypot can also be useful in catching and prosecuting those behind an attack. Researchers suspect that some cybercriminals also use honeypots to gather intelligence about researchers, act as decoys and to spread misinformation.

High-interaction honeypots imitate the activities of a production system and capture extensive information -- pure honeypots are full-fledged production systems using a tap on the honeypot's link to the network. The goal of high-interaction honeypots is for the attacker to gain root access on the machine, and then study what he or she does. An attacker with root access has access to all commands and files on a system, so this type of honeypot carries the greatest risk but also has the greatest potential for collecting information. Low-interaction honeypots simulate only the services frequently targeted by attackers and so are less risky and less complex to maintain. Virtual machines are often used to host honeypots so the honeypot can be restored more quickly if it is compromised. Two or more honeypots on a network form a honeynet, while a honeyfarm is a centralized collection of honeypots and analysis tools.

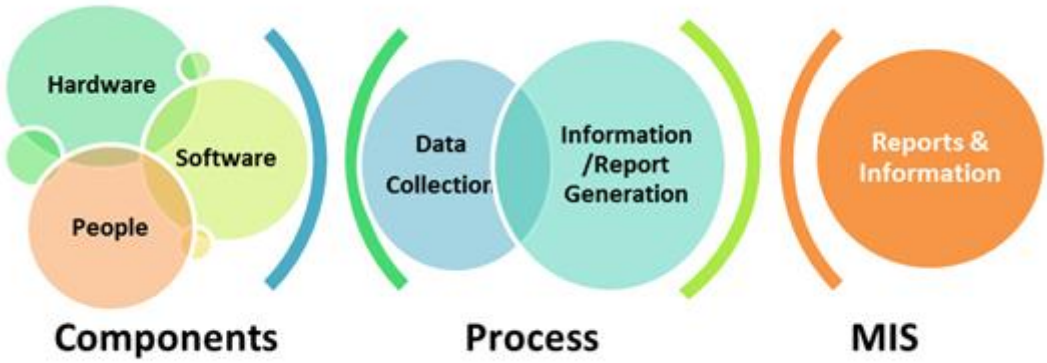
Databases often get attacked by intruders using SQL Injection.As such activities are not recognized by basic firewalls, companies often use database firewalls for protection. Some of the available SQl database firewalls provide/support honeypot architectures so that the intruder runs against a trap database while the web application remains functional.

3)MIS(Management Information System):Management Information System can be defined as a formal method of collecting timely information in a presentable form. in order to facilitate effective decision making and implementation, in order to carry out organizational operations for the purpose of achieving the organizational goal. A **management information system** is a system design to provide selected decision –orientation information needed by management plan, control and evaluate the activities of the corporation. It is designed within the frame work that emphasizes profit, planning, performance planning and control at all levels. It complements the ultimate integration of required business information sub system both financial with in the company.

According to Philip kolter- A **marketing information system** consist of people, equipment and procedures together,sort,analyse,evaluate and distribute the needed timely and accurate information and marketing decision makers.

Professor Allen S. Lee states that research in the information system field examines more than the technological system,or just the social system or even the two side by side in addition it investigates the phenomena that emerge when the two interact.

An information system can be any organized combination of people, hardware, software, communication network and data resources that collects, transforms and disseminates information in an organization.



Management information system concept is a vital to effective computer use in business of two or major reason:

1- It serves as a system framework for organizing business computer applications. business application of computer should be viewed as interrelated and integrated computer based information system and not as independent data processing job.

2- It emphasized the management orientation of electronics information processing in business the primary goal of computer based information should be the processing of data generated by business operations.

Advantages of Management Information System

1. **Management information system** helps the managers to make planning and control decision.

2. Facilitated planning- **management information system** improves the quality of plants by providing relevant information for sound decision making due to increase in size and complexity of organization managers have lost personal contact with the scenes of operations.

3. Minimize information overload- **management information system** changes the larger amount of data into summarized form and thereby avoids the confusion which may arise when manager are flooded with detailed facts.

4. Bring coordination-**management information system** facilities integration of specialized activities by keeping each department aware of the problem and requirement of their department. it connect all decision centers in the organization.

5. Make control easier-it serves as a link between managerial planning and control. It improves the ability of management to evaluate and improve performance. The use of computers has increased the data processing and storage capability and reduces the cost.

6. **Management information system** assembles, processes.stores, retrie, evaluate, and disseminate the information.

7. It insure that appropriate data is collected from the various sources, processed, and sent further to all the needy destinations.

8. **Management information system** helps in strategic planning, management control, operational control and transaction processing.

9. It helps the clerical personnel in the transaction processing and answer their queries on the data pertaining to the transaction the status of a particular record and reference on a variety of documents.

4)What is after-imaging (AI):

The after-imaging feature lets you recover a database that was damaged when a failure caused the loss of the database or primary recovery (before image / BI) area. When you enable after-imaging, the database engine writes notes containing a description of all database changes to the after-image (AI) files. You can use the AI files with the roll-forward recovery process to restore the database to the condition it was in before you lost the database, without losing completed transactions that occurred since the last backup.

After-image extent phases:

Empty means the extent is available for use.
Busy means the extent is currently in use.
Full means the extent has filled up and is ready to be archived and emptied.
Locked means that OpenEdge Replication is enabled for the database and the AI file contents have not been transmitted to the target database(s) yet. When all of the contents have been transmitted to the target database the extent will be marked Full.

Managing of after-imaging involves:

a) Managing your backups and ensuring that you are able to distinguish the date when the backup was made and that the backup is stored externally to the machine.

b) Using the following command to generate a new structure file and backing up a copy of the database structure (.st) file.

c) As AI extents become FULL, copy those FULL extents to a backup location, preferably external to the disk storage that contains the database data or bi files.

Roll-forward recovery using after-imaging:

The database should not be accessed until all after-image files have been rolled onto (applied) to the database, otherwise it will not be possible to continue rolling AI files.The roll-forward mechanism can either roll forward all transactions up to the end of the last available AI file, to a specified point in time or to a specific transaction number provided the last good backup and all required after image files are available.Rolling forward to a point in time or to a specific transaction number can be useful if needing to recover to a state prior to an undesirable event, such as a programming error that adversely affected the database.

5)**Slice and dice** refers to a strategy for segmenting, viewing and understanding data in a database. Users slices and dice by cutting a large segment of data into smaller parts, and repeating this process until arriving at the right level of detail for analysis. Slicing and dicing helps provide a closer view of data for analysis and presents data in new and diverse perspectives.The term is typically used with OLAP databases that present information to the user in the form of multidimensional cubes similar to a 3D spreadsheet.

6) Multiple Choice Questions Of Data Warehousing:

1. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of
management decisions.
A. Data Mining.
B. Data Warehousing.
C. Web Mining.

D. Text Mining.
ANSWER: B
2. The data Warehouse is_____.
A. read only.
B. write only.
C. read write only.
D. none.
ANSWER: A
3. Expansion for DSS in DW is_____.
A. Decision Support system.
B. Decision Single System.
C. Data Storable System.
D. Data Support System.
ANSWER: A
4. The important aspect of the data warehouse environment is that data found within the data warehouse
is_____.
A. subject-oriented.
B. time-variant.
C. integrated.
D. All of the above.
ANSWER: D
5. The time horizon in Data warehouse is usually _____.
A. 1-2 years.
B. 3-4years.
C. 5-6 years.
D. 5-10 years.
ANSWER: D
6. The data is stored, retrieved & updated in _____.
A. OLAP.
B. OLTP.
C. SMTP.
D. FTP.

ANSWER: B

7. _____describes the data contained in the data warehouse.

A. Relational data.

B. Operational data.

C. Metadata.

D. Informational data.

ANSWER: C

8. _____predicts future trends & behaviors, allowing business managers to make proactive,
knowledge-driven decisions.

A. Data warehouse.

B. Data mining.

C. Datamarts.

D. Metadata.

ANSWER: B

9. _____ is the heart of the warehouse.

A. Data mining database servers.

B. Data warehouse database servers.

C. Data mart database servers.

D. Relational data base servers.

ANSWER: B

10. _____ is the specialized data warehouse database.

A. Oracle.

B. DBZ.

C. Informix.

D. Redbrick.

ANSWER: D

11. _____defines the structure of the data held in operational databases and used by
operational applications.

A. User-level metadata.

B. Data warehouse metadata.

C. Operational metadata.

D. Data mining metadata.
ANSWER: C
12. _____ is held in the catalog of the warehouse database system.
A. Application level metadata.
B. Algorithmic level metadata.
C. Departmental level metadata.
D. Core warehouse metadata.
ANSWER: B
13. _____maps the core warehouse metadata to business concepts, familiar and useful to end users.
A. Application level metadata.
B. User level metadata.
C. Enduser level metadata.
D. Core level metadata.
ANSWER: A
14. _____consists of formal definitions, such as a COBOL layout or a database schema.
A. Classical metadata.
B. Transformation metadata.
C. Historical metadata.
D. Structural metadata.
ANSWER: A
15. _____consists of information in the enterprise that is not in classical form.
A. Mushy metadata.
B. Differential metadata.
C. Data warehouse.
D. Data mining.
ANSWER: A
16. . _____databases are owned by particular departments or business groups.
A. Informational.
B. Operational.

C. Both informational and operational.
D. Flat.
ANSWER: B
17. The star schema is composed of _____ fact table.
A. one.
B. two.
C. three.
D. four.
ANSWER: A
18. The time horizon in operational environment is _____.
A. 30-60 days.
B. 60-90 days.
C. 90-120 days.
D. 120-150 days.
ANSWER: B
19. The key used in operational environment may not have an element of_____.
A. time.
B. cost.
C. frequency.
D. quality.
ANSWER: A
20. Data can be updated in _____environment.
A. data warehouse.
B. data mining.
C. operational.
D. informational.
ANSWER: C
21. Record cannot be updated in _____.
A. OLTP
B. files
C. RDBMS
D. data warehouse

ANSWER: D

22. The source of all data warehouse data is the_____.

A. operational environment.

B. informal environment.

C. formal environment.

D. technology environment.

ANSWER: A

23. Data warehouse contains_____data that is never found in the operational environment.

A. normalized.

B. informational.

C. summary.

D. denormalized.

ANSWER: C

24. The modern CASE tools belong to _____ category.

A. a. analysis.

B. b.Development

C. c.Coding

D. d.Delivery

ANSWER: A

25. Bill Inmon has estimated_____of the time required to build a data warehouse, is consumed in
the conversion process.

A. 10 percent.

B. 20 percent.

C. 40 percent

D. 80 percent.

ANSWER: D

26. Detail data in single fact table is otherwise known as_____.

A. monoatomic data.

B. diatomic data.

C. atomic data.

D. multiatomic data.

ANSWER: C

27. _____test is used in an online transactional processing environment.

A. MEGA.

B. MICRO.

C. MACRO.

D. ACID.

ANSWER: D

28. _____ is a good alternative to the star schema.

A. Star schema.

B. Snowflake schema.

C. Fact constellation.

D. Star-snowflake schema.

ANSWER: C

29. The biggest drawback of the level indicator in the classic star-schema is that it limits_____.

A. quantify.

B. qualify.

C. flexibility.

D. ability.

ANSWER: C

30. A data warehouse is _____.

A. updated by end users.

B. contains numerous naming conventions and formats

C. organized around important subject areas.

D. contains only current data.

ANSWER: C

31. An operational system is _____.

A. used to run the business in real time and is based on historical data.

B. used to run the business in real time and is based on current data.

C. used to support decision making and is based on current data.

D. used to support decision making and is based on historical data.

ANSWER: B

32. The generic two-level data warehouse architecture includes _____.

A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. far real-time updates.
ANSWER: C
33. The active data warehouse architecture includes _____
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. all of the above.
ANSWER: D
34. Reconciled data is _____.
A. data stored in the various operational systems throughout the organization.
B. current data intended to be the single source for all decision support systems.
C. data stored in one operational system in the organization.
D. data that has been selected and formatted for end-user support applications.
ANSWER: B
35. Transient data is _____.
A. data in which changes to existing records cause the previous version of the records to be eliminated.
B. data in which changes to existing records do not cause the previous version of the records to be
eliminated.
C. data that are never altered or deleted once they have been added.
D. data that are never deleted once they have been added.
ANSWER: A
36. The extract process is _____.
A. capturing all of the data contained in various operational systems.
B. capturing a subset of the data contained in various operational systems.
C. capturing all of the data contained in various decision support systems.
D. capturing a subset of the data contained in various decision support systems.
ANSWER: B
37. Data scrubbing is _____.

A. a process to reject data from the data warehouse and to create the necessary indexes.
B. a process to load the data in the data warehouse and to create the necessary indexes.
C. a process to upgrade the quality of data after it is moved into a data warehouse.
D. a process to upgrade the quality of data before it is moved into a data warehouse
ANSWER: D
38. The load and index is _____.
A. a process to reject data from the data warehouse and to create the necessary indexes.
B. a process to load the data in the data warehouse and to create the necessary indexes.
C. a process to upgrade the quality of data after it is moved into a data warehouse.
D. a process to upgrade the quality of data before it is moved into a data warehouse.
ANSWER: B
39. Data transformation includes _____.
A. a process to change data from a detailed level to a summary level.
B. a process to change data from a summary level to a detailed level.
C. joining data from one source into various sources of data.
D. separating data from one source into various sources of data.
ANSWER: A
40. _____ is called a multifield transformation.
A. Converting data from one field into multiple fields.
B. Converting data from fields into field.
C. Converting data from double fields into multiple fields.
D. Converting data from one field to one field.
ANSWER: A
41. The type of relationship in star schema is _____.
A. many-to-many.

B. one-to-one.
C. one-to-many.
D. many-to-one.
ANSWER: C
42. Fact tables are _____.
A. completely demoralized.
B. partially demoralized.
C. completely normalized.
D. partially normalized.
ANSWER: C
43. _____ is the goal of data mining.
A. To explain some observed event or condition.
B. To confirm that data exists.
C. To analyze data for expected relationships.
D. To create a new data warehouse.
ANSWER: A
44. Business Intelligence and data warehousing is used for _____.
A. Forecasting.
B. Data Mining.
C. Analysis of large volumes of product sales data.
D. All of the above.
ANSWER: D
45. The data administration subsystem helps you perform all of the following, except_____.
A. backups and recovery.
B. query optimization.
C. security management.
D. create, change, and delete information.
ANSWER: D
46. The most common source of change data in refreshing a data warehouse is _____.
A. queryable change data.
B. cooperative change data.

C. logged change data.
D. snapshot change data.
ANSWER: A
47. _____ are responsible for running queries and reports against data warehouse tables.
A. Hardware.
B. Software.
C. End users.
D. Middle ware.
ANSWER: C
48. Query tool is meant for _____.
A. data acquisition.
B. information delivery.
C. information exchange.
D. communication.
ANSWER: A
49. Classification rules are extracted from _____.
A. root node.
B. decision tree.
C. siblings.
D. branches.
ANSWER: B
50. Dimensionality reduction reduces the data set size by removing _____.
A. relevant attributes.
B. irrelevant attributes.
C. derived attributes.
D. composite attributes.
ANSWER: B
51. _____ is a method of incremental conceptual clustering.
A. CORBA.
B. OLAP.
C. COBWEB.

D. STING.
ANSWER: C
52. Effect of one attribute value on a given class is independent of values of other attribute is called
_____.
A. value independence.
B. class conditional independence.
C. conditional independence.
D. unconditional independence.
ANSWER: A
53. The main organizational justification for implementing a data warehouse is to provide _____.
A. cheaper ways of handling transportation.
B. decision support.
C. storing large volume of data.
D. access to data.
ANSWER: C
54. Multidimensional database is otherwise known as_____.
A. RDBMS
B. DBMS
C. EXTENDED RDBMS
D. EXTENDED DBMS
ANSWER: B
55. Data warehouse architecture is based on _____.
A. DBMS.
B. RDBMS.
C. Sybase.
D. SQL Server.
ANSWER: B
56. Source data from the warehouse comes from _____.
A. ODS.
B. TDS.
C. MDDB.

D. ORDBMS.
ANSWER: A
57. _____ is a data transformation process.
A. Comparison.
B. Projection.
C. Selection.
D. Filtering.
ANSWER: D
58. The technology area associated with CRM is _____.
A. specialization.
B. generalization.
C. personalization.
D. summarization.
ANSWER: C
59. SMP stands for _____.
A. Symmetric Multiprocessor.
B. Symmetric Multiprogramming.
C. Symmetric Metaprogramming.
D. Symmetric Microprogramming.
ANSWER: A
60. _____ are designed to overcome any limitations placed on the warehouse by the nature of the
relational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C
61. _____ are designed to overcome any limitations placed on the warehouse by the nature of the
relational data model.
A. Operational database.
B. Relational database.

C. Multidimensional database.
D. Data repository.
ANSWER: C
62. MDDB stands for _____.
A. multiple data doubling.
B. multidimensional databases.
C. multiple double dimension.
D. multi-dimension doubling.
ANSWER: B
63. _____ is data about data.
A. Metadata.
B. Microdata.
C. Minidata.
D. Multidata.
ANSWER: A
64. _____ is an important functional component of the metadata.
A. Digital directory.
B. Repository.
C. Information directory.
D. Data dictionary.
ANSWER: C
65. EIS stands for _____.
A. Extended interface system.
B. Executive interface system.
C. Executive information system.
D. Extendable information system.
ANSWER: C
66. _____ is data collected from natural systems.
A. MRI scan.
B. ODS data.
C. Statistical data.
D. Historical data.
ANSWER: A

67. _____ is an example of application development environments.
A. Visual Basic.
B. Oracle.
C. Sybase.
D. SQL Server.
ANSWER: A
68. The term that is not associated with data cleaning process is _____.
A. domain consistency.
B. deduplication.
C. disambiguation.
D. segmentation.
ANSWER: D
69. _____ are some popular OLAP tools.
A. Metacube, Informix.
B. Oracle Express, Essbase.
C. HOLAP.
D. MOLAP.
ANSWER: A
70. Capability of data mining is to build _____ models.
A. retrospective.
B. interrogative.
C. predictive.
D. imperative.
ANSWER: C
71. _____ is a process of determining the preference of customer's majority.
A. Association.
B. Preferencing.
C. Segmentation.
D. Classification.
ANSWER: B
72. Strategic value of data mining is _____.
A. cost-sensitive.

B. work-sensitive.
C. time-sensitive.
D. technical-sensitive.
ANSWER: C
73. _____ proposed the approach for data integration issues.
A. Ralph Campbell.
B. Ralph Kimball.
C. John Raphlin.
D. James Gosling.
ANSWER: B
74. The terms equality and roll up are associated with _____.
A. OLAP.
B. visualization.
C. data mart.
D. decision tree.
ANSWER: C
75. Exceptional reporting in data warehousing is otherwise called as _____.
A. exception.
B. alerts.
C. errors.
D. bugs.
ANSWER: B
76. _____ is a metadata repository.
A. Prism solution directory manager.
B. CORBA.
C. STUNT.
D. COBWEB.
ANSWER: A
77. _____ is an expensive process in building an expert system.
A. Analysis.
B. Study.
C. Design.

D. Information collection.
ANSWER: D
78. The full form of KDD is _____.
A. Knowledge database.
B. Knowledge discovery in database.
C. Knowledge data house.
D. Knowledge data definition.
ANSWER: B
79. The first International conference on KDD was held in the year _____.
A. 1996.
B. 1997.
C. 1995.
D. 1994.
ANSWER: C
80. Removing duplicate records is a process called _____.
A. recovery.
B. data cleaning.
C. data cleansing.
D. data pruning.
ANSWER: B
81. _____ contains information that gives users an easy-to-understand perspective of the
information stored in the data warehouse.
A. Business metadata.
B. Technical metadata.
C. Operational metadata.
D. Financial metadata.
ANSWER: A
82. _____ helps to integrate, maintain and view the contents of the data warehousing system.
A. Business directory.
B. Information directory.

C. Data dictionary.
D. Database.
ANSWER: B
83. Discovery of cross-sales opportunities is called _____.
A. segmentation.
B. visualization.
C. correction.
D. association.
ANSWER: D
84. Data marts that incorporate data mining tools to extract sets of data are called _____.
A. independent data mart.
B. dependent data marts.
C. intra-entry data mart.
D. inter-entry data mart.
ANSWER: B
85. _____ can generate programs itself, enabling it to carry out new tasks.
A. Automated system.
B. Decision making system.
C. Self-learning system.
D. Productivity system.
ANSWER: D
86. The power of self-learning system lies in _____.
A. cost.
B. speed.
C. accuracy.
D. simplicity.
ANSWER: C
87. Building the informational database is done with the help of _____.
A. transformation or propagation tools.
B. transformation tools only.
C. propagation tools only.

D. extraction tools.
ANSWER: A
88. How many components are there in a data warehouse?
A. two.
B. three.
C. four.
D. five.
ANSWER: D
89. Which of the following is not a component of a data warehouse?
A. Metadata.
B. Current detail data.
C. Lightly summarized data.
D. Component Key.
ANSWER: D
90. _____ is data that is distilled from the low level of detail found at the current detailed leve.
A. Highly summarized data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: B
91. Highly summarized data is _____.
A. compact and easily accessible.
B. compact and expensive.
C. compact and hardly accessible.
D. compact.
ANSWER: A
92. A directory to help the DSS analyst locate the contents of the data warehouse is seen in _____.
A. Current detail data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.

ANSWER: C

93. Metadata contains atleast _____.

A. the structure of the data.

B. the algorithms used for summarization.

C. the mapping from the operational environment to the data warehouse.

D. all of the above.

ANSWER: D

94. Which of the following is not a old detail storage medium?

A. Phot Optical Storage.

B. RAID.

C. Microfinche.

D. Pen drive.

ANSWER: D

95. The data from the operational environment enter _____ of data warehouse.

A. Current detail data.

B. Older detail data.

C. Lightly summarized data.

D. Highly summarized data.

ANSWER: A

96. The data in current detail level resides till _____ event occurs.

A. purge.

B. summarization.

C. archieved.

D. all of the above.

ANSWER: D

97. The dimension tables describe the _____.

A. entities.

B. facts.

C. keys.

D. units of measures.

ANSWER: B

98. The granularity of the fact is the _____ of detail at which it is recorded.

A. transformation.
B. summarization.
C. level.
D. transformation and summarization.
ANSWER: C
99. Which of the following is not a primary grain in analytical modeling?
A. Transaction.
B. Periodic snapshot.
C. Accumulating snapshot.
D. All of the above.
ANSWER: B
100. Granularity is determined by _____.
A. number of parts to a key.
B. granularity of those parts.
C. both A and B.
D. none of the above.
ANSWER: C
101. _____ of data means that the attributes within a given entity are fully dependent on the entire
primary key of the entity.
A. Additivity.
B. Granularity.
C. Functional dependency.
D. Dimensionality.
ANSWER: C
102. A fact is said to be fully additive if _____.
A. it is additive over every dimension of its dimensionality.
B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: A
103. A fact is said to be partially additive if _____.
A. it is additive over every dimension of its dimensionality.

B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: B
104. A fact is said to be non-additive if _____.
A. it is additive over every dimension of its dimensionality.
B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: C
105. Non-additive measures can often combined with additive measures to create new _____.
A. additive measures.
B. non-additive measures.
C. partially additive.
D. All of the above.
ANSWER: A
106. A fact representing cumulative sales units over a day at a store for a product is a _____.
A. additive fact.
B. fully additive fact.
C. partially additive fact.
D. non-additive fact.
ANSWER: B
107. _____ of data means that the attributes within a given entity are fully dependent on the entire
primary key of the entity.
A. Additivity.
B. Granularity.
C. Functional Dependency.
D. Dependency.
ANSWER: C
108. Which of the following is the other name of Data mining?

A. Exploratory data analysis.
B. Data driven discovery.
C. Deductive learning.
D. All of the above.
ANSWER: D
109. Which of the following is a predictive model?
A. Clustering.
B. Regression.
C. Summarization.
D. Association rules.
ANSWER: B
110. Which of the following is a descriptive model?
A. Classification.
B. Regression.
C. Sequence discovery.
D. Association rules.
ANSWER: C
111. A _____ model identifies patterns or relationships.
A. Descriptive.
B. Predictive.
C. Regression.
D. Time series analysis.
ANSWER: A
112. A predictive model makes use of _____.
A. current data.
B. historical data.
C. both current and historical data.
D. assumptions.
ANSWER: B
113. _____ maps data into predefined groups.
A. Regression.
B. Time series analysis
C. Prediction.

D. Classification.
ANSWER: D
114. _____ is used to map a data item to a real valued prediction variable.
A. Regression.
B. Time series analysis.
C. Prediction.
D. Classification.
ANSWER: B
115. In _____, the value of an attribute is examined as it varies over time.
A. Regression.
B. Time series analysis.
C. Sequence discovery.
D. Prediction.
ANSWER: B
116. In _____ the groups are not predefined.
A. Association rules.
B. Summarization.
C. Clustering.
D. Prediction.
ANSWER: C
117. Link Analysis is otherwise called as _____.
A. affinity analysis.
B. association rules.
C. both A & B.
D. Prediction.
ANSWER: C
118. _____ is a the input to KDD.
A. Data.
B. Information.
C. Query.
D. Process.
ANSWER: A

119. The output of KDD is _____.
A. Data.
B. Information.
C. Query.
D. Useful information.
ANSWER: D
120. The KDD process consists of _____ steps.
A. three.
B. four.
C. five.
D. six.
ANSWER: C
121. Treating incorrect or missing data is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: B
122. Converting data from different sources into a common format for processing is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: C
123. Various visualization techniques are used in _____ step of KDD.
A. selection.
B. transformaion.
C. data mining.
D. interpretation.
ANSWER: D
124. Extreme values that occur infrequently are called as _____.
A. outliers.

B. rare values.
C. dimensionality reduction.
D. All of the above.
ANSWER: A
125. Box plot and scatter diagram techniques are _____.
A. Graphical.
B. Geometric.
C. Icon-based.
D. Pixel-based.
ANSWER: B
126. _____ is used to proceed from very specific knowledge to more general information.
A. Induction.
B. Compression.
C. Approximation.
D. Substitution.
ANSWER: A
127. Describing some characteristics of a set of data by a general model is viewed as _____
A. Induction.
B. Compression.
C. Approximation.
D. Summarization.
ANSWER: B
128. _____ helps to uncover hidden information about the data.
A. Induction.
B. Compression.
C. Approximation.
D. Summarization.
ANSWER: C
129. _____ are needed to identify training data and desired results.
A. Programmers.
B. Designers.

C. Users.
D. Administrators.
ANSWER: C
130. Overfitting occurs when a model _____.
A. does fit in future states.
B. does not fit in future states.
C. does fit in current state.
D. does not fit in current state.
ANSWER: B
131. The problem of dimensionality curse involves _____.
A. the use of some attributes may interfere with the correct completion of a data mining task.
B. the use of some attributes may simply increase the overall complexity.
C. some may decrease the efficiency of the algorithm.
D. All of the above.
ANSWER: D
132. Incorrect or invalid data is known as _____.
A. changing data.
B. noisy data.
C. outliers.
D. missing data.
ANSWER: B
133. ROI is an acronym of _____.
A. Return on Investment.
B. Return on Information.
C. Repetition of Information.
D. Runtime of Instruction
ANSWER: A
134. The _____ of data could result in the disclosure of information that is deemed to be
confidential.
A. authorized use.
B. unauthorized use.

C. authenticated use.
D. unauthenticated use.
ANSWER: B
135. _____ data are noisy and have many missing attribute values.
A. Preprocessed.
B. Cleaned.
C. Real-world.
D. Transformed.
ANSWER: C
136. The rise of DBMS occurred in early _____.
A. 1950's.
B. 1960's
C. 1970's
D. 1980's.
ANSWER: C
137. SQL stand for _____.
A. Standard Query Language.
B. Structured Query Language.
C. Standard Quick List.
D. Structured Query list.
ANSWER: B
138. Which of the following is not a data mining metric?
A. Space complexity.
B. Time complexity.
C. ROI.
D. All of the above.
ANSWER: D
139. Reducing the number of attributes to solve the high dimensionality problem is called as _____.
A. dimensionality curse.
B. dimensionality reduction.
C. cleaning.
D. Overfitting.

ANSWER: B

140. Data that are not of interest to the data mining task is called as _____.

A. missing data.

B. changing data.

C. irrelevant data.

D. noisy data.

ANSWER: C

141. _____ are effective tools to attack the scalability problem.

A. Sampling.

B. Parallelization

C. Both A & B.

D. None of the above.

ANSWER: C

142. Market-basket problem was formulated by _____.

A. Agrawal et al.

B. Steve et al.

C. Toda et al.

D. Simon et al.

ANSWER: A

143. Data mining helps in _____.

A. inventory management.

B. sales promotion strategies.

C. marketing strategies.

D. All of the above.

ANSWER: D

144. The proportion of transaction supporting X in T is called _____.

A. confidence.

B. support.

C. support count.

D. All of the above.

ANSWER: B

145. The absolute number of transactions supporting X in T is called _____.

A. confidence.
B. support.
C. support count.
D. None of the above.
ANSWER: C
146. The value that says that transactions in D that support X also support Y is called _____.
A. confidence.
B. support.
C. support count.
D. None of the above.
ANSWER: A
147. If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam,
10000 transaction contain both bread and jam. Then the support of bread and jam is _____.
A. 2%
B. 20%
C. 3%
D. 30%
ANSWER: A
148. 7 If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam,
10000 transaction contain both bread and jam. Then the confidence of buying bread with jam is _____.
A. 33.33%
B. 66.66%
C. 45%
D. 50%
ANSWER: D
149. The left hand side of an association rule is called _____.
A. consequent.
B. onset.

C. antecedent.
D. precedent.
ANSWER: C
150. The right hand side of an association rule is called _____.
A. consequent.
B. onset.
C. antecedent.
D. precedent.
ANSWER: A
151. Which of the following is not a desirable feature of any efficient algorithm?
A. to reduce number of input operations.
B. to reduce number of output operations.
C. to be efficient in computing.
D. to have maximal code length.
ANSWER: D
152. All set of items whose support is greater than the user-specified minimum support are called as
_____.
A. border set.
B. frequent set.
C. maximal frequent set.
D. lattice.
ANSWER: B
153. If a set is a frequent set and no superset of this set is a frequent set, then it is called _____.
A. maximal frequent set.
B. border set.
C. lattice.
D. infrequent sets.
ANSWER: A
154. Any subset of a frequent set is a frequent set. This is _____.
A. Upward closure property.
B. Downward closure property.

C. Maximal frequent set.
D. Border set.
ANSWER: B
155. Any superset of an infrequent set is an infrequent set. This is _____.
A. Maximal frequent set.
B. Border set.
C. Upward closure property.
D. Downward closure property.
ANSWER: C
156. If an itemset is not a frequent set and no superset of this is a frequent set, then it is _____.
A. Maximal frequent set
B. Border set.
C. Upward closure property.
D. Downward closure property.
ANSWER: B
157. A priori algorithm is otherwise called as _____.
A. width-wise algorithm.
B. level-wise algorithm.
C. pincer-search algorithm.
D. FP growth algorithm.
ANSWER: B
158. The A Priori algorithm is a _____.
A. top-down search.
B. breadth first search.
C. depth first search.
D. bottom-up search.
ANSWER: D
159. The first phase of A Priori algorithm is _____.
A. Candidate generation.
B. Itemset generation.
C. Pruning.
D. Partitioning.

ANSWER: A

160. The second phaase of A Priori algorithm is _____.

A. Candidate generation.

B. Itemset generation.

C. Pruning.

D. Partitioning.

ANSWER: C

161. The _____ step eliminates the extensions of (k-1)-itemsets which are not found to be frequent, from
being considered for counting support.

A. Candidate generation.

B. Pruning.

C. Partitioning.

D. Itemset eliminations.

ANSWER: B

162. The a priori frequent itemset discovery algorithm moves _____ in the lattice.

A. upward.

B. downward.

C. breadthwise.

D. both upward and downward.

ANSWER: A

163. After the pruning of a priori algorithm, _____ will remain.

A. Only candidate set.

B. No candidate set.

C. Only border set.

D. No border set.

ANSWER: B

164. The number of iterations in a priori _____.

A. increases with the size of the maximum frequent set.

B. decreases with increase in size of the maximum frequent set.

C. increases with the size of the data.

D. decreases with the increase in size of the data.

ANSWER: A
165. MFCS is the acronym of _____.
A. Maximum Frequency Control Set.
B. Minimal Frequency Control Set.
C. Maximal Frequent Candidate Set.
D. Minimal Frequent Candidate Set.
ANSWER: C
166. Dynamuc Itemset Counting Algorithm was proposed by ____.
A. Bin et al.
B. Argawal et at.
C. Toda et al.
D. Simon et at.
ANSWER: A
167. Itemsets in the _____ category of structures have a counter and the stop number with them.
A. Dashed.
B. Circle.
C. Box.
D. Solid.
ANSWER: A
168. The itemsets in the _____category structures are not subjected to any counting.
A. Dashes.
B. Box.
C. Solid.
D. Circle.
ANSWER: C
169. Certain itemsets in the dashed circle whose support count reach support value during an iteration
move into the _____.
A. Dashed box.
B. Solid circle.
C. Solid box.

D. None of the above.
ANSWER: A
170. Certain itemsets enter afresh into the system and get into the _____, which are essentially the
supersets of the itemsets that move from the dashed circle to the dashed box.
A. Dashed box.
B. Solid circle.
C. Solid box.
D. Dashed circle.
ANSWER: D
171. The itemsets that have completed on full pass move from dashed circle to _____.
A. Dashed box.
B. Solid circle.
C. Solid box.
D. None of the above.
ANSWER: B
172. The FP-growth algorithm has _____ phases.
A. one.
B. two.
C. three.
D. four.
ANSWER: B
173. A frequent pattern tree is a tree structure consisting of _____.
A. an item-prefix-tree.
B. a frequent-item-header table.
C. a frequent-item-node.
D. both A & B.
ANSWER: D
174. The non-root node of item-prefix-tree consists of _____ fields.
A. two.
B. three.
C. four.

D. five.
ANSWER: B
175. The frequent-item-header-table consists of _____ fields.
A. only one.
B. two.
C. three.
D. four.
ANSWER: B
176. The paths from root node to the nodes labelled 'a' are called _____.
A. transformed prefix path.
B. suffix subpath.
C. transformed suffix path.
D. prefix subpath.
ANSWER: D
177. The transformed prefix paths of a node 'a' form a truncated database of pattern which co-occur with a
is called _____.
A. suffix path.
B. FP-tree.
C. conditional pattern base.
D. prefix path.
ANSWER: C
178. The goal of _____ is to discover both the dense and sparse regions of a data set.
A. Association rule.
B. Classification.
C. Clustering.
D. Genetic Algorithm.
ANSWER: C
179. Which of the following is a clustering algorithm?
A. A priori.
B. CLARA.
C. Pincer-Search.

D. FP-growth.
ANSWER: B
180. _____ clustering technique start with as many clusters as there are records, with each cluster having
only one record.
A. Agglomerative.
B. divisive.
C. Partition.
D. Numeric.
ANSWER: A
181. _____ clustering techniques starts with all records in one cluster and then try to split that cluster
into small pieces.
A. Agglomerative.
B. Divisive.
C. Partition.
D. Numeric.
ANSWER: B
182. Which of the following is a data set in the popular UCI machine-learning repository?
A. CLARA.
B. CACTUS.
C. STIRR.
D. MUSHROOM.
ANSWER: D
183. In _____ algorithm each cluster is represented by the center of gravity of the cluster.
A. k-medoid.
B. k-means.
C. STIRR.
D. ROCK.
ANSWER: B
184. In _____ each cluster is represented by one of the objects of the

cluster located near the
center.
A. k-medoid.
B. k-means.
C. STIRR.
D. ROCK.
ANSWER: A
185. Pick out a k-medoid algoithm.
A. DBSCAN.
B. BIRCH.
C. PAM.
D. CURE.
ANSWER: C
186. Pick out a hierarchical clustering algorithm.
A. DBSCAN
B. BIRCH.
C. PAM.
D. CURE.
ANSWER: A
187. CLARANS stands for _____.
A. CLARA Net Server.
B. Clustering Large Application RAnge Network Search.
C. Clustering Large Applications based on RANdomized Search.
D. CLustering Application Randomized Search.
ANSWER: C
188. BIRCH is a _____.
A. agglomerative clustering algorithm.
B. hierarchical algorithm.
C. hierarchical-agglomerative algorithm.
D. divisive.
ANSWER: C
189. The cluster features of different subclusters are maintained in a tree called
_____.

A. CF tree.
B. FP tree.
C. FP growth tree.
D. B tree.
ANSWER: A
190. The _____ algorithm is based on the observation that the frequent sets are normally very few in
number compared to the set of all itemsets.
A. A priori.
B. Clustering.
C. Association rule.
D. Partition.
ANSWER: D
191. The partition algorithm uses _____ scans of the databases to discover all frequent sets.
A. two.
B. four.
C. six.
D. eight.
ANSWER: A
192. The basic idea of the apriori algorithm is to generate_____ item sets of a particular size & scans
the database.
A. candidate.
B. primary.
C. secondary.
D. superkey.
ANSWER: A
193. _____is the most well known association rule algorithm and is used in most commercial products.
A. Apriori algorithm.
B. Partition algorithm.
C. Distributed algorithm.

D. Pincer-search algorithm.
ANSWER: A
194. An algorithm called_____is used to generate the candidate item sets for each pass after the first.
A. apriori.
B. apriori-gen.
C. sampling.
D. partition.
ANSWER: B
195. The basic partition algorithm reduces the number of database scans to _____ & divides it into
partitions.
A. one.
B. two.
C. three.
D. four.
ANSWER: B
196. _____and prediction may be viewed as types of classification.
A. Decision.
B. Verification.
C. Estimation.
D. Illustration.
ANSWER: C
197. _____can be thought of as classifying an attribute value into one of a set of possible classes.
A. Estimation.
B. Prediction.
C. Identification.
D. Clarification.
ANSWER: B
198. Prediction can be viewed as forecasting a_____value.
A. non-continuous.
B. constant.

C. continuous.
D. variable.
ANSWER: C
199. _____data consists of sample input data as well as the classification assignment for the data.
A. Missing.
B. Measuring.
C. Non-training.
D. Training.
ANSWER: D
200. Rule based classification algorithms generate _____ rule to perform the classification.
A. if-then.
B. while.
C. do while.
D. switch.
ANSWER: A
201. _____ are a different paradigm for computing which draws its inspiration from neuroscience.
A. Computer networks.
B. Neural networks.
C. Mobile networks.
D. Artificial networks.
ANSWER: B
202. The human brain consists of a network of _____.
A. neurons.
B. cells.
C. Tissue.
D. muscles.
ANSWER: A
203. Each neuron is made up of a number of nerve fibres called _____.
A. electrons.
B. molecules.

C. atoms.
D. dendrites.
ANSWER: D
204. The _____is a long, single fibre that originates from the cell body.
A. axon.
B. neuron.
C. dendrites.
D. strands.
ANSWER: A
205. A single axon makes _____ of synapses with other neurons.
A. ones.
B. hundreds.
C. thousands.
D. millions.
ANSWER: C
206. _____ is a complex chemical process in neural networks.
A. Receiving process.
B. Sending process.
C. Transmission process.
D. Switching process.
ANSWER: C
207. _____ is the connectivity of the neuron that give simple devices their real power. a. b. c. d.
A. Water.
B. Air.
C. Power.
D. Fire.
ANSWER: D
208. _____ are highly simplified models of biological neurons.
A. Artificial neurons.
B. Computational neurons.
C. Biological neurons.
D. Technological neurons.

ANSWER: A

209. The biological neuron's _____ is a continuous function rather than a step function.

A. read.

B. write.

C. output.

D. input.

ANSWER: C

210. The threshold function is replaced by continuous functions called _____ functions.

A. activation.

B. deactivation.

C. dynamic.

D. standard.

ANSWER: A

211. The sigmoid function also knows as _____functions.

A. regression.

B. logistic.

C. probability.

D. neural.

ANSWER: B

212. MLP stands for _____.

A. mono layer perception.

B. many layer perception.

C. more layer perception.

D. multi layer perception.

ANSWER: D

213. In a feed- forward networks, the conncetions between layers are _____ from input to output.

A. bidirectional.

B. unidirectional.

C. multidirectional.

D. directional.

ANSWER: B
214. The network topology is constrained to be _____.
A. feedforward.
B. feedbackward.
C. feed free.
D. feed busy.
ANSWER: A
215. RBF stands for _____.
A. Radial basis function.
B. Radial bio function.
C. Radial big function.
D. Radial bi function.
ANSWER: A
216. RBF have only _____ hidden layer.
A. four.
B. three.
C. two.
D. one.
ANSWER: D
217. RBF hidden layer units have a receptive field which has a _____; that is, a particular input
value at which they have a maximal output.
A. top.
B. bottom.
C. centre.
D. border.
ANSWER: C
218. _____ training may be used when a clear link between input data sets and target output values
does not exist.
A. Competitive.
B. Perception.
C. Supervised.

D. Unsupervised.
ANSWER: D
219. _____ employs the supervised mode of learning.
A. RBF.
B. MLP.
C. MLP & RBF.
D. ANN.
ANSWER: C
220. _____ design involves deciding on their centres and the sharpness of their Gaussians.
A. DR.
B. AND.
C. XOR.
D. RBF.
ANSWER: D
221. _____ is the most widely applied neural network technique.
A. ABC.
B. PLM.
C. LMP.
D. MLP.
ANSWER: D
222. SOM is an acronym of _____.
A. self-organizing map.
B. self origin map.
C. single organizing map.
D. simple origin map.
ANSWER: A
223. _____ is one of the most popular models in the unsupervised framework.
A. SOM.
B. SAM.
C. OSM.
D. MSO.

ANSWER: A

224. The actual amount of reduction at each learning step may be guided by _____.

A. learning cost.

B. learning level.

C. learning rate.

D. learning time.

ANSWER: C

225. The SOM was a neural network model developed by _____.

A. Simon King.

B. Teuvokohonen.

C. Tomoki Toda.

D. Julia.

ANSWER: B

226. SOM was developed during _____.

A. 1970-80.

B. 1980-90.

C. 1990 -60.

D. 1979 -82.

ANSWER: D

227. Investment analysis used in neural networks is to predict the movement of _____ from previous
data.

A. engines.

B. stock.

C. patterns.

D. models.

ANSWER: B

228. SOMs are used to cluster a specific _____ dataset containing information about the patient's
drugs etc.

A. physical.

B. logical.

C. medical.
D. technical.
ANSWER: C
229. GA stands for _____.
A. Genetic algorithm
B. Gene algorithm.
C. General algorithm.
D. Geo algorithm.
ANSWER: A
230. GA was introduced in the year _____.
A. 1955.
B. 1965.
C. 1975.
D. 1985.
ANSWER: C
231. Genetic algorithms are search algorithms based on the mechanics of natural_____.
A. systems.
B. genetics.
C. logistics.
D. statistics.
ANSWER: B
232. GAs were developed in the early _____.
A. 1970.
B. 1960.
C. 1950.
D. 1940.
ANSWER: A
233. The RSES system was developed in _____.
A. Poland.
B. Italy.
C. England.
D. America.

ANSWER: A
234. Crossover is used to _____.
A. recombine the population's genetic material.
B. introduce new genetic structures in the population.
C. to modify the population's genetic material.
D. All of the above.
ANSWER: A
235. The mutation operator _____.
A. recombine the population's genetic material.
B. introduce new genetic structures in the population.
C. to modify the population's genetic material.
D. All of the above.
ANSWER: B
236. Which of the following is an operation in genetic algorithm?
A. Inversion.
B. Dominance.
C. Genetic edge recombination.
D. All of the above.
ANSWER: D
237. . _____ is a system created for rule induction.
A. RBS.
B. CBS.
C. DBS.
D. LERS.
ANSWER: D
238. NLP stands for _____.
A. Non Language Process.
B. Nature Level Program.
C. Natural Language Page.
D. Natural Language Processing.
ANSWER: D
239. Web content mining describes the discovery of useful information from the _____contents.

A. text.
B. web.
C. page.
D. level.
ANSWER: B
240. Research on mining multi-types of data is termed as _____ data.
A. graphics.
B. multimedia.
C. meta.
D. digital.
ANSWER: B
241. _____ mining is concerned with discovering the model underlying the link structures of the web.
A. Data structure.
B. Web structure.
C. Text structure.
D. Image structure.
ANSWER: B
242. _____ is the way of studying the web link structure.
A. Computer network.
B. Physical network.
C. Social network.
D. Logical network.
ANSWER: C
243. The _____ propose a measure of standing a node based on path counting.
A. open web.
B. close web.
C. link web.
D. hidden web.
ANSWER: B
244. In web mining, _____ is used to find natural groupings of users, pages, etc.

A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: A
245. In web mining, _____ is used to know the order in which URLs tend to be accessed.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: C
246. In web mining, _____ is used to know which URLs tend to be requested together.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: B
247. _____ describes the discovery of useful information from the web contents.
A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. All of the above.
ANSWER: A
248. _____ is concerned with discovering the model underlying the link structures of the web.
A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. All of the above.
ANSWER: B

249. The _____ engine for a data warehouse supports query-triggered usage of data
A. NNTP
B. SMTP
C. OLAP
D. POP
ANSWER: C
250. _____ displays of data such as maps, charts and other graphical representation allow data to be
presented compactly to the users.
A. Hidden
B. Visual
C. Obscured
D. Concealed
ANSWER: B

7)**Hot Site:**

Hot sites are essentially mirrors of your datacenter infrastructure. The backup site is populated with servers, cooling, power, and office space (if applicable). The most important feature offered from a hot site is that the production environment(s) are running concurrently with your main datacenter. This syncing allows for minimal impact and downtime to business operations. In the event of a significant outage event to your main datacenter, the hot site can take the place of the impacted site immediately. However, this level of redundancy does not come cheap, and businesses will have to weigh the cost-benefit-analysis (CBA) of hot site utilization.

Cold Site:

A cold site is essentially office or datacenter space without any server-related equipment installed. The cold site provides power, cooling, and/or office space which waits in the event of a significant outage to the main work site or datacenter. The cold site will require extensive support from engineering and IT

personnel to get all necessary servers and equipment migrated and functional. Cold sites are the cheapest cost-recovery option for businesses to utilize.

Warm Site:

A warm site is the middle ground of the two disaster recovery options. Warm sites offer office space/datacenter space and will have some pre-installed server hardware. The difference between a hot site and a warm site is that while the hot site provides a mirror of the production data-center and its environment(s), a warm site will contain only servers ready for the installation of production environments. Warm sites make sense for aspect of the business which is not critical, but requires a level of redundancy (ex. Administrative roles). A CBA conducted on whether to use a warm site versus a hot site should include the downtime associated with the software-loading/configuration requirements for engineering.

8)Ingress filtering is a method of verifying that inbound packets arriving at a network are from the source computer they claim to be from before entry (or ingress) is granted. Ingress filtering takes advantage of the Layer 2IP address filtering capability of a router at the network's edge and blocks traffic that has a high probability of being malicious. At its simplest, ingress filtering involves establishing an access control list that contains the Internet Protocol addresses (IP addresses) of permitted source addresses. Conversely, the access control list may also be used to block prohibited source addresses.

Source IP addresses that are commonly blocked with ingress filter include:

a)IP addresses that are already in use as an IP address within the internal network. This helps prevent an attacker from spoofing an internal IP address to take advantage of a poorly written firewall rule.

b)IP addresses that are private. This helps prevent malicious traffic coming in from an improperly configured Internet-based host or an attacker's spoofed address.

c)IP addresses that are loopbacks. This helps prevent traffic from an attacker who is spoofing a loopback address to take advantage of a poorly written firewall rule base.

d)IP addresses that are multicast addresses. This helps prevent undesired multicast traffic that is likely to be spam.

e)IP addresses that are service or management network addresses. This helps prevent an attacker from using the public Internet to gain unauthorized access to network services running at the network and above.

Egress filtering is a process in which outbound data is monitored or restricted, usually by means of a firewall that blocks packets that fail to meet certain security requirements. The word "egress" means "outgoing"; an egress router, for example, is one through which packets leave a network on their way to another network. The main purpose of egress filtering is to ensure that unwanted or destructive traffic (such as malware, unauthorized e-mail messages, or requests to Web sites) do not leave a particular network. An example might be the use of a firewall to keep students on a university's campus network from sending out malware or copyrighted material from any of the computers within that network. Egress filtering can also be used to allow only certain servers or computers within an organization's network to send data out of that network. This precaution can prevent, for example, employee use of corporate computers for casual Web surfing or excessive personal correspondences.

9)**Database Recovery:**A major responsibility of the database administrator is to prepare for the possibility of hardware, software, network, process, or system failure. If such a failure affects the operation of a database system, you must usually recover the database and return to normal operation as quickly as possible. Recovery should protect the database and associated users from unnecessary problems and avoid or reduce the possibility of having to duplicate work manually.Recovery processes vary depending on the type of failure that occurred, the structures affected, and the type of recovery that you perform. If

no files are lost or damaged, recovery may amount to no more than restarting an instance. If data has been lost, recovery requires additional steps.

Errors and Failures

Several problems can halt the normal operation of an Oracle database or affect database I/O to disk. The following sections describe the most common types. For some of these problems, recovery is automatic and requires little or no action on the part of the database user or database administrator.

User Error

A database administrator can do little to prevent user errors (for example, accidentally dropping a table). Usually, user error can be reduced by increased training on database and application principles. Furthermore, by planning an effective recovery scheme ahead of time, the administrator can ease the work necessary to recover from many types of user errors.

Statement Failure

Statement failure occurs when there is a logical failure in the handling of a statement in an Oracle program. For example, assume all extents of a table (in other words, the number of extents specified in the MAXEXTENTS parameter of the CREATE TABLE statement) are allocated, and are completely filled with data; the table is absolutely full. A valid INSERT statement cannot insert a row because there is no space available. Therefore, if issued, the statement fails.

If a statement failure occurs, the Oracle software or operating system returns an error code or message. A statement failure usually requires no action or recovery steps; Oracle automatically corrects for statement failure by rolling back the effects (if any) of the statement and returning control to the application. The user can simply re-execute the statement after correcting the problem indicated by the error message.

Process Failure

A process failure is a failure in a user, server, or background process of a database instance (for example, an abnormal disconnect or process

termination). When a process failure occurs, the failed subordinate process cannot continue work, although the other processes of the database instance can continue.

The Oracle background process PMON detects aborted Oracle processes. If the aborted process is a user or server process, PMON resolves the failure by rolling back the current transaction of the aborted process and releasing any resources that this process was using. Recovery of the failed user or server process is automatic. If the aborted process is a background process, the instance usually cannot continue to function correctly. Therefore, you must shut down and restart the instance.

Network Failure

When your system uses networks (for example, local area networks, phone lines, and so on) to connect client workstations to database servers, or to connect several database servers to form a distributed database system, network failures (such as aborted phone connections or network communication software failures) can interrupt the normal operation of a database system. For example:

A network failure might interrupt normal execution of a client application and cause a process failure to occur. In this case, the Oracle background process PMON detects and resolves the aborted server process for the disconnected user process, as described in the previous section.

A network failure might interrupt the two-phase commit of a distributed transaction. Once the network problem is corrected, the Oracle background process RECO of each involved database server automatically resolves any distributed transactions not yet resolved at all nodes of the distributed database system.

Database Instance Failure

Database instance failure occurs when a problem arises that prevents an Oracle database instance (SGA and background processes) from continuing to work. An instance failure can result from a hardware problem, such as a power outage, or

a software problem, such as an operating system crash. Instance failure also results when you issue a SHUTDOWN ABORT or STARTUP FORCE command.

Recovery from Instance Failure

a)Crash or instance recovery recovers a database to its transaction-consistent state just before instance failure. *Crash recovery* recovers a database in a single-instance configuration and *instance recovery* recovers a database in an Oracle Parallel Server configuration.

b)Recovery from instance failure is automatic. For example, when using the Oracle Parallel Server, another instance performs instance recovery for the failed instance. In single-instance configurations, Oracle performs crash recovery for a database when the database is restarted (mounted and opened to a new instance). The transition from a mounted state to an open state automatically triggers crash recovery, if necessary.

Crash or instance recovery consists of the following steps:

a)Rolling forward to recover data that has not been recorded in the datafiles, yet has been recorded in the online redo log, including the contents of rollback segments. This is called *cache recovery*.

b)Opening the database. Instead of waiting for all transactions to be rolled back before making the database available, Oracle allows the database to be opened as soon as cache recovery is complete. Any data that is not locked by unrecovered transactions is immediately available.

c)Marking all transactions system-wide that were active at the time of failure as DEAD and marking the rollback segments containing these transactions as PARTLY AVAILABLE.

Rolling back dead transactions as part of SMON recovery. This is called *transaction recovery*.Resolving any pending distributed transactions undergoing a two-phase commit at the time of the instance failure.As new transactions encounter rows locked by dead transactions, they can automatically roll back the dead transaction to release the locks. If you are using

Fast-Start Recovery, just the data block is immediately rolled back, as opposed to the entire transaction.

Media (Disk) Failure

An error can arise when trying to write or read a file that is required to operate an Oracle database. This occurrence is called *media failure* because there is a physical problem reading or writing to files on the storage medium.A common example of media failure is a disk head crash, which causes the loss of all files on a disk drive. All files associated with a database are vulnerable to a disk crash, including data files, online redo log files, and control files.The appropriate recovery from a media failure depends on the files affected.

Recovery and Atomicity:

When a system crashes, it may have several transactions being executed and various files opened for them to modify the data items. Transactions are made of various operations, which are atomic in nature. But according to ACID properties of DBMS, atomicity of transactions as a whole must be maintained, that is, either all the operations are executed or none.

When a DBMS recovers from a crash, it should maintain the following −

a)It should check the states of all the transactions, which were being executed.

b)A transaction may be in the middle of some operation; the DBMS must ensure the atomicity of the transaction in this case.

c)It should check whether the transaction can be completed now or it needs to be rolled back.

d)No transactions would be allowed to leave the DBMS in an inconsistent state.

There are two types of techniques, which can help a DBMS in recovering as well as maintaining the atomicity of a transaction −

a)Maintaining the logs of each transaction, and writing them onto some stable storage before actually modifying the database.

b)Maintaining shadow paging, where the changes are done on a volatile memory, and later, the actual database is updated.

Log-based Recovery: Log is a sequence of records, which maintains the records of actions performed by a transaction. It is important that the logs are written prior to the actual modification and stored on a stable storage media, which is failsafe.

Log-based recovery works as follows –

a)The log file is kept on a stable storage media.

b)When a transaction enters the system and starts execution, it writes a log about it.

The database can be modified using two approaches −

a)Deferred database modification − All logs are written on to the stable storage and the database is updated when a transaction commits.

b)Immediate database modification − Each log follows an actual database modification. That is, the database is modified immediately after every operation.

Recovery with Concurrent Transactions: When more than one transaction are being executed in parallel, the logs are interleaved. At the time of recovery, it would become hard for the recovery system to backtrack all logs, and then start recovering. To ease this situation, most modern DBMS use the concept of 'checkpoints'.

Checkpoint: Keeping and maintaining logs in real time and in real environment may fill out all the memory space available in the system. As time passes, the log file may grow too big to be handled at all. Checkpoint is a mechanism where all the previous logs are removed from the system and stored permanently in a storage disk. Checkpoint declares a point before which the DBMS was in consistent state, and all the transactions were committed.

All the transactions in the undo-list are then undone and their logs are removed. All the transactions in the redo-list and their previous logs are removed and then redone before saving their logs.

10)**Concurrency Control**:When multiple transactions are trying to access the same sharable resource, there could arise many problems if the access control is not done properly. There are some important mechanisms to which access control can be maintained. Earlier we talked about theoretical concepts like serializability, but the practical concept of this can be implemented by using **Locks** and **Timestamps**. Here we shall discuss some protocols where Locks and Timestamps can be used to provide an environment in which concurrent transactions can preserve their Consistency and Isolation properties.

In order to maintain the concurrent access of transactions, two protocols are introduced.

a)Lock Based Protocol: - Lock is in other words called as access. In this type of protocol any transaction will not be processed until the transaction gets the lock on the record. That means any transaction will not retrieve or insert or update or delete the data unless it gets the access to that particular data.

These locks are broadly classified as Binary locks and shared / exclusive locks.

In **binary lock** data can either be locked or unlocked. It will have only these two states. It can be locked for retrieve or insert or update or delete the data or unlocked for not using the data.

In **shared / exclusive lock** technique the data is said to be exclusively locked if for insert / update /delete. When it is exclusively locked no other transaction can read or write the data. When a data is read from the database, then its lock is shared i.e.; the data can be read by other transaction too but it cannot be changed while retrieving the data.The relationship between Shared and Exclusive Lock can be represented by the following table which is known as Lock Matrix.

Two Phase Locking Protocol

The use of locks has helped us to create neat and clean concurrent schedule. The Two Phase Locking Protocol defines the rules of how to acquire the locks on a data item and how to release the locks.
The Two Phase Locking Protocol assumes that a transaction can only be in one of two phases.

Growing Phase: In this phase the transaction can only acquire locks, but cannot release any lock. The transaction enters the growing phase as soon as it acquires the first lock it wants. From now on it has no option but to keep acquiring all the locks it would need. It cannot release any lock at this phase even if it has finished working with a locked data item. Ultimately the transaction reaches a point where all the lock it may need has been acquired. This point is called **Lock Point**.

Shrinking Phase: After Lock Point has been reached, the transaction enters the shrinking phase. In this phase the transaction can only release locks, but cannot acquire any new lock. The transaction enters the shrinking phase as soon as it releases the first lock after crossing the Lock Point. From now on it has no option but to keep releasing all the acquired locks.

There are two different versions of the Two Phase Locking Protocol. One is called the Strict Two Phase Locking Protocol and the other one is called the Rigorous Two Phase Locking Protocol.

Strict Two Phase Locking Protocol

In this protocol, a transaction may release all the shared locks after the Lock Point has been reached, but it cannot release any of the exclusive locks until the transaction commits. This protocol helps in creating cascade less schedule.

A **Cascading Schedule** is a typical problem faced while creating concurrent schedule.

b)Time Stamp Based Protocol: - As we have seen above in lock based protocol, it acquires locks at the time of execution. But in this method, as soon as a transaction is created it assigns the order of the transaction. The order of the transaction is nothing but the ascending order of the transaction creation. The priority for older transaction is given to execute first. This protocol uses system time or logical counter to determine the time stamp of the transaction.Suppose there are two transactions T1 and T2. Suppose T1 has entered the system at time 0005 and T2 has entered the system at 0008 clock time. Priority will be given to T1 to execute first as it is entered the system first.

In addition to the timestamp of a transaction, this protocol also maintains the timestamp of last 'read' and 'write' operation on a data. Based on the timestamp of transaction and the data which it is accessing a timestamp ordering protocol is defined.

11)

Web Server -

Server on which your website is hosted. This server will have installed web servers such as IIS, apache, etc.

Application Server -

Server on which your created applications which are utilizing your database, web service, etc. This application server will host business layer (wrapped with web services), scheduled jobs, windows services, etc.It provides middleware services for security and state maintenance, along with data access and persistence.

More descriptively, it can be viewed as dividing an application into:

1)A first-tier, front-end, Web browser-based graphical user interface, usually at a personal computer or workstation

2)A middle-tier business logic application or set of applications, possibly on a local area network or intranet server

3)A third-tier, back-end database and transaction server, sometimes on a mainframe or large server

Database Server -

Database server will have your one or more database hosted such as Oracle, Sql Server, MySql, etc

12)**A RFID chip** is another term used to define a RFID tag. It is a tag, label or card that can exchange data with a reader using radio frequency (RF) signals. It usually has a built-in antenna and an integrated circuit IC). The antenna can send and receive radio waves, while the IC takes care of modulating and demodulating the radio signals, as well as processing and storing data.RFID chips are quite similar to bar code labels in that they typically work with a corresponding scanner or reader. However, RFID chips have significant advantages. Because a RFID chip communicates with a reader through radio waves (not infrared, which is being used by bar code technology), the chip doesn't have to be positioned right in front of the reader. That is, line-of-sight is not needed.

Also, unlike a bar code reader/label pair, which have to be really close (about a few centimeters), some RFID reader/chip pairs can function even if they are a few meters apart. Furthermore, while a bar code label can only be read by a single reader at a time, a RFID chip can transmit data to multiple readers simultaneously.
There are different kinds of RFID chips. Some require batteries, known as active chips, while others don't (passive). Others are designed for indoor use or built for rugged, outdoor applications. The most common applications include object tracking and identification.
Chips can also differ in the kind of radio frequencies they operate on. Some communicate via Ultra High Frequency (UHF), High Frequency (HF) or Low Frequency (LF).RFID chips can be attached just about anywhere: clothes, shoes,

vehicles, containers and even plants, animals and human beings (as implants). Miniaturized chips have even been attached to insects.

13)**Cloud Computing** is a broad term that describes a broad range of services. As with other significant developments in technology, many vendors have seized the term “Cloud” and are using it for products that sit outside of the common definition. In order to truly understand how the Cloud can be of value to an organization, it is first important to understand what the Cloud really is and its different components. Since the Cloud is a broad collection of services, organizations can choose where, when, and how they use Cloud Computing. In this report we will explain the different types of Cloud Computing services commonly referred to as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) and give some examples and case studies to illustrate how they all work. We will also provide some guidance on situations where particular flavors of Cloud Computing are not the best option for an organization.

Infrastructure as a Service (IaaS)

Providing the fundamental building blocks of computing resources, IaaS takes the traditional physical computer hardware, such as servers, storage arrays, and networking, and lets you build virtual infrastructure that mimics these resources, but which can be created, reconfigured, resized, and removed within moments, as and when a task requires it. The most well known IaaS provider, Amazon Web Services, offers a variety of options, including their “EC2” computing platform, and “S3” storage platform.

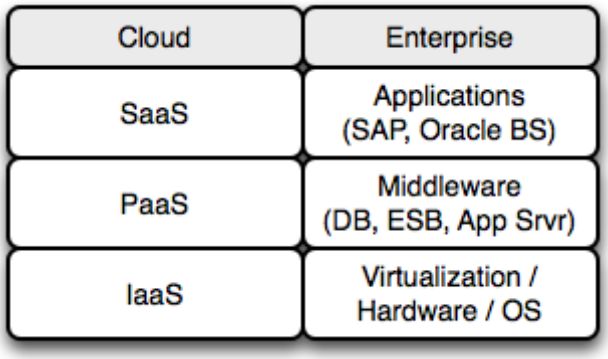
Platform as a Service (PaaS)

Operating at the layer above raw computing hardware, whether physical or virtual, PaaS provides a method for programming languages to interact with services like databases, web servers, and file storage, without having to deal with lower level requirements like how much space a database needs, whether the data must be protected by making a copy between 3 servers, or distributing the workload across servers that can be spread throughout the world. Typically, applications must be written for a specific PaaS offering to take full advantage

of the service, and most platforms only support a limited set of programming languages. Often, PaaS providers also have a Software as a Service offering (see below), and the platform has been initially built to support that specific software. Some examples of PaaS solutions are the “Google App Engine” system, “Heroku” which operates on top of the Amazon Web Services IaaS system.

Software as a Service (SaaS)

The top layer of cloud computing, Software as a Service is typically built on top of a Platform as a Service solution, whether that platform is publicly available or not, and provides software for end-users such as email, word processing, or a business CRM. Software as a Service is typically charged on a per-user and per-month basis, and companies have the flexibility to add or remove users at any time without addition costs beyond the monthly per-user fee. Some of the most well known SaaS solutions are “Google Apps”.



13)**Common Gateway Interface:**The common gateway interface (CGI) is a standard way for a Web server to pass a Web user's request to an application program and to receive data back to forward to the user. When the user requests a Web page (for example, by clicking on a highlighted word or entering a Web site address), the server sends back the requested page. However, when a user fills out a form on a Web page and sends it in, it usually needs to be processed by an application program. The Web server typically passes the form information to a small application program that processes the data and may send back a confirmation message. This method or convention for passing data back and forth between the server and the application is called the common

gateway interface (CGI). It is part of the Web's Hypertext Transfer Protocol (HTTP).Another increasingly common way to provide dynamic feedback for Web users is to include scripts or programs that run on the user's machine rather than the Web server. These programs can be Java applets, Java scripts, or ActiveX controls. These technologies are known collectively as *client-side* solutions, while the use of CGI is a *server-side* solution because the processing occurs on the Web server.

One problem with CGI is that each time a CGI script is executed, a new process is started. For busy Web sites, this can slow down the server noticeably. A more efficient solution, but one that it is also more difficult to implement, is to use the server's API, such as ISAPI or NSAPI. Another increasingly popular solution is to use Java servlets.

14)**Operating system virtualization (OS virtualization)** is a server virtualization technology that involves tailoring a standard operating system so that it can run different applications handled by multiple users on a single computer at a time. The operating systems do not interfere with each other even though they are on the same computer.
In OS virtualization, the operating system is altered so that it operates like several different, individual systems. The virtualized environment accepts commands from different users running different applications on the same machine. The users and their requests are handled separately by the virtualized operating system.Also known as operating system-level virtualization. OS virtualization can also be used to migrate critical applications to another running operating system instance. Patches and updates to the underlying operating system are done in a timely way, and have little or no impact on the availability of application services. The processes in the OS virtualized environment are isolated and their interactions with the underlying OS instance are monitored.

15)An **extranet** is a private network that uses Internet technology and the public telecommunication system to securely share part of a business's information or operations with suppliers, vendors, partners, customers, or other businesses. An extranet can be viewed as part of a company's intranet that is extended to users outside the company. It has also been described as a "state of mind" in which the Internet is perceived as a way to do business with other companies as well as to sell products to customers.An extranet requires security and privacy. These can include firewall server management, the issuance and use of digital certificates or similar means of user authentication, encryption of messages, and the use of virtual private networks (VPNs) that tunnel through the public network.

16)A **disaster recovery plan** (DRP) is a documented process or set of procedures to recover and protect a business IT infrastructure in the event of a disaster.Such a plan, ordinarily documented in written form, specifies procedures an organization is to follow in the event of a disaster. It is "a comprehensive statement of consistent actions to be taken before, during and after a disaster.The disaster could be natural, environmental or man-made. Man-made disasters could be intentional (for example, an act of a terrorist) or unintentional (that is, accidental, such as the breakage of a man-made dam).

There are benefits that can be obtained from the drafting of a disaster recovery plan. Some of these benefits are:

a)Providing a sense of security

b)Minimizing risk of delays

c)Guaranteeing the reliability of standby systems

d)Providing a standard for testing the plan

e)Minimizing decision-making during a disaster

f)Reducing potential legal liabilities

g)Lowering unnecessarily stressful work environment

A **backup site** or **work area recovery site** is a location where an organization can relocate following a disaster, such as fire, flood, terrorist threat or other disruptive event. This is an integral part of the disaster recovery plan and wider business continuity planning of an organization.

A backup, or alternate, site can be another data center location operated by the organization, or contracted via a company that specializes in disaster recovery services. In some cases, one organization will have an agreement with a second organization to operate a joint backup site. In addition, an organization may have a reciprocal agreement with another organization to set up a warm site at each of their data centers.

There are three types of backup sites, including cold sites, warm sites, and hot sites. The differences between the types are determined by the costs and effort required to implement each.

17)**Brute force attack** is a trial-and-error method used to obtain information such as a user password or personal identification number (PIN). In a brute force attack, automated software is used to generate a large number of consecutive guesses as to the value of the desired data. Brute force attacks may be used by criminals to crack encrypted data, or by security analysts to test an organization's network security. A brute force attack may also be referred to as brute force cracking.

For example, a form of brute force attack known as a dictionary attack might try all the words in a dictionary. Other forms of brute force attack might try commonly-used passwords or combinations of letters and numbers.

Strong hashing algorithms can slow down brute-force attacks. Essentially, hashing algorithms perform additional mathematical work on a password before storing a value derived from the password on disk. If a slower hashing algorithm is used, it will require thousands of times as much mathematical work to try each password and dramatically slow down brute-force attacks. However, the more work required, the more work a server or other computer has to do

each time as user logs in with their password. Software must balance resilience against brute-force attacks with resource usage.

18)A **traceability matrix** is a document that co-relates any two-baseline documents that require a many-to-many relationship to check the completeness of the relationship.It is used to track the requirements and to check the current project requirements are met.

What is RTM (Requirement Traceability Matrix):

Requirement Traceability Matrix or RTM captures all requirements proposed by the client or development team and their traceability in a single document delivered at the conclusion of the life-cycle.In other words, it is a document that maps and traces user requirement with test cases. The main purpose of Requirement Traceability Matrix is to see that all test cases are covered so that no functionality should miss while testing.

Requirement Traceability Matrix – Parameters include

Requirement ID

Risks

Requirement Type and Description

Trace to design specification

Unit test cases

Integration test cases

System test cases

User acceptance test cases

Trace to test script

Types of Traceability Test Matrix

Forward traceability: This matrix is used to check whether the project progresses in the desired direction and for the right product. It makes sure that each requirement is applied to the product and that each requirement is tested thoroughly. It maps requirements to test cases.

Backward or reverse traceability: It is used to ensure whether the current product remains on the right track. The purpose behind this type of traceability is to verify that we are not expanding the scope of the project by adding code, design elements, test or other work that is not specified in the requirements. It maps test cases to requirements.

Bi-directional traceability (Forward+Backward): This traceability metrics ensures that all requirements are covered by test cases. It analyzes the impact of a change in requirements affected by the defect in a work product and vice versa.

Requirements tracing, a process of documenting the links between the requirements and the work products developed to implement and verify those requirements. The RTM captures all requirements and their traceability in a single document delivered at the conclusion of the life cycle.

Advantage of Requirement Traceability Matrix

a)It confirms 100% test coverage

b)It highlights any requirements missing or document inconsistencies

c)It shows the overall defects or execution status with a focus on business requirements

d)It helps in analyzing or estimating the impact on the QA team's work with respect to revisiting or re-working on the test cases

19)A **PKI** allows you to bind public keys (contained in SSL certificates) with a person in a way that allows you to trust the certificate. Public Key Infrastructures, like the one used to secure the Internet, most commonly use a Certificate Authority (also called a Registration Authority) to verify the identity of an entity and create unforgeable certificates. Web browsers, web servers, email clients, smart cards, and many other types of hardware and software all have integrated, standards-based PKI support that can be used with each other.Without PKI, sensitive information can still be encrypted (ensuring confidentiality) and exchanged, but there would be no assurance of the identity (authentication) of the other party. Any form of sensitive data exchanged over the Internet is reliant on PKI for security.

A typical PKI includes the following key elements:

a)A trusted party, called a certificate authority (CA), acts as the root of trust and provides services that authenticate the identity of individuals, computers and other entities

b)A registration authority, often called a subordinate CA, certified by a root CA to issue certificates for specific uses permitted by the root

c)A certificate database, which stores certificate requests and issues and revokes certificates

d)A certificate store, which resides on a local computer as a place to store issued certificates and private keys

A CA issues digital certificates to entities and individuals after verifying their identity. It signs these certificates using its private key; its public key is made available to all interested parties in a self-signed CA certificate. CAs use this trusted root certificate to create a "chain of trust" -- many root certificates are embedded in Web browsers so they have built-in trust of those CAs. Web servers, email clients, smartphones and many other types of hardware and software also support PKI and contain trusted root certificates from the major CAs.

20)An **internet exchange point (IXP)** is a physical network access point through which major network providers connect their networks and exchange traffic. The primary focus of an exchange point is to facilitate network interconnection through an exchange access point instead of third-party networks.
Internet exchange points were created to minimize the part of an Internet service provider's (ISP) network traffic that had to go through an upstream provider. IXPs provide a common place for ISPs to exchange their Internet traffic between autonomous network systems. The exchange points are often established in the same city to avoid latency.

The advantages of Internet exchange points include:

a)Allowing high speed data transfer

b)Reducing latency

c)Providing fault tolerance

d)Improving routing efficiency

e)Improving bandwidth

An exchange point is usually an Ethernet switch, much like the Ethernet switches that connect computers in an office network. Each network connecting to the exchange point connects one or more of its routers to the exchange point's Ethernet switch, and they send traffic across the Ethernet switch to routers belonging to other networks.Some exchange points use more complex designs, connecting multiple Ethernet switches in multiple buildings, and sometimes using other network protocols to carry traffic between those switches. Even in those cases, the hand-off to other networks is still done as an Ethernet connection.

21)The **Constructive Cost Model (COCOMO)** is the most widely used software estimation model in the world. It was developed by Barry Boehm of TRW and first published in his book Software Engineering Economics in 1981. The COCOMO model predicts the effort and duration of a project based on inputs relating to the size of the resulting systems and a number of "cost drives" that

affect productivity. The most fundamental calculation in the COCOMO model is the use of the Effort Equation (Equation1) to estimate the number of Person-Months required developing a project. Most of the other COCOMO results, including the estimates for Requirements and Maintenance, are derived from this quantity.

COCOMO is defined in terms of three different models:

a)The Basic Model

b)The Intermediate Model

c)The Detailed Model

The Basic Model:

Basic COCOMO model estimates the software development effort using only a single predictor variable (size in DSI) and three software development modes. Basic COCOMO is good for quick, early, rough order of magnitude estimates of software costs, but its accuracy is necessarily limited because of its lack of factors which have a significant influence on software costs. The Basic COCOMO estimates are within a factor of 1.3 only 29% of the time, and within a factor of 2 only 60% of the time.
The Basic COCOMO Model estimate for annual software maintenance is calculated in terms of **Annual Change Traffic** (**ACT**). **ACT** is the fraction of the software product's source instructions which undergo change during a (typical) year, either through addition or modification.

The Intermediate Model:

The Intermediate model use an **Effort Adjustment Factor (EAF)** and slightly different coefficients for the effort equations than the Basic model. You can apply Intermediate COCOMO across the entire software product for easily and roughly cost estimation during the early stage, or apply it at the software product component level for more accurate cost estimation in more detailed stages.

Detailed COCOMO:

It an extension of the Intermediate model that adds effort multipliers for each phase of the project to determine the cost driver impact on each step.