

**UNIT I****BUSINESS INTELLIGENCE**

**Effective and timely decisions – Data, information and knowledge – Role of mathematical models – Business intelligence architectures: Cycle of a business intelligence analysis – Enabling factors in business intelligence projects – Development of a business intelligence system – Ethics and business intelligence.**

**PART-A****1. Define Business Intelligence.(APRIL/MAY 2017)**

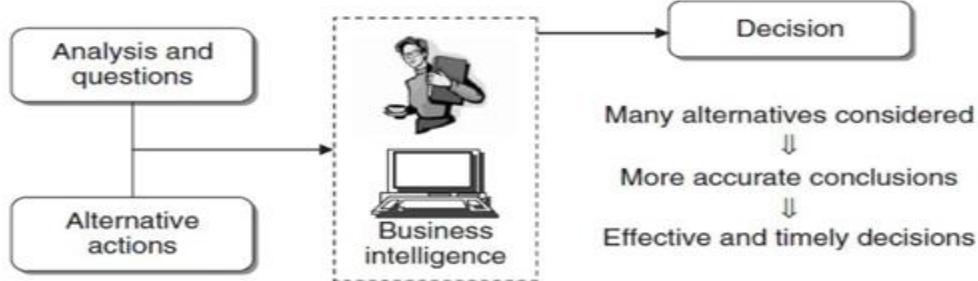
Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.

**2. What are effective decisions?**

The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way.

**3. What are timely decisions?**

Enterprises operate in economic environments characterized by growing levels of competition and high dynamism. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company.

**4. What are the benefits of a business intelligence system? (APRIL/MAY 2017)**

The decision makers ask themselves a series of questions and develop the corresponding analysis. Hence, they examine and compare several options, selecting among them the best decision, given the conditions at hand. If decision makers can rely on a business intelligence system facilitating their activity, we can expect that the overall quality of the decision-making process will be greatly improved. With the help of mathematical models and algorithms, it is actually possible to analyze a larger number of alternative actions, achieve more accurate conclusions and reach effective and timely decisions.

**5. Define data.**

Data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities. For example, for a retailer data refer to primary entities such as customers, points of sale and items. They need to be processed by means of appropriate extraction tools and analytical methods capable of transforming them into information and knowledge that can be subsequently used by decision makers.

**6. Difference between information and knowledge.**

Information	Knowledge
Information is the outcome of extraction and processing activities carried out on data, and it appears meaningful for those who receive it in a specific domain	Information is transformed into knowledge when it is used to make decisions and develop the corresponding actions.
For example: To the sales manager of a retail company, the number of customers holding a loyalty card who have reduced by more than 50% the monthly amount spent in the last three months, represent meaningful pieces of information that can be extracted from raw stored data.	For example: For a retail company, a sales analysis may detect that a group of customers, living in an area where a competitor has recently opened a new point of sale, have reduced their usual amount of business.

## 6. What is attrition ?

Low customer loyalty, also known as customer attrition or churn, is a critical factor for many companies operating in service industries.

## 7. Who are knowledge workers?

In complex organizations, public or private, decisions are made on a continual basis. Such decisions may be more or less critical, have long- or short-term effects and involve people and roles at various hierarchical levels. The ability of these knowledge workers to make decisions, both as individuals and as a community, is one of the primary factors that influence the performance and competitive strength of a given organization.

## 8. What do you mean by knowledge management?

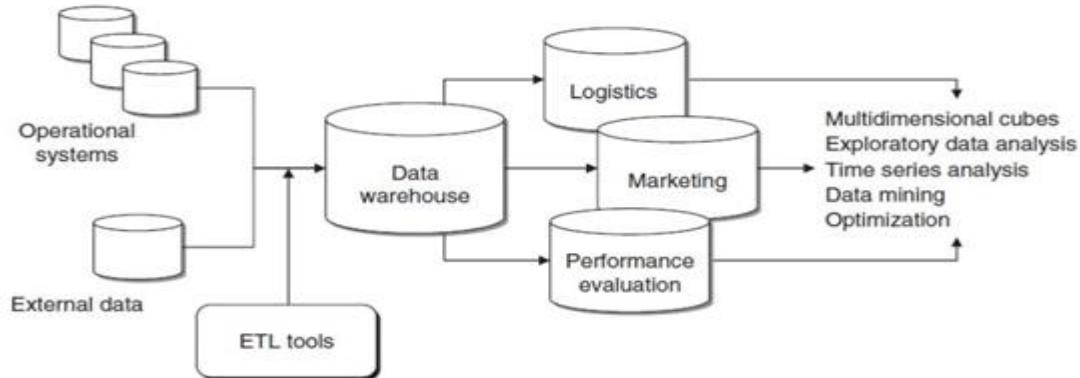
Several public and private enterprises and organizations have developed in recent years formal and systematic mechanisms to gather store and share their wealth of knowledge, which is now perceived as an invaluable intangible asset. The activity of providing support to knowledge workers through the integration of decision-making processes and enabling information technologies is usually referred to as knowledge management.

## 9. How to summarize a typical business intelligence analysis?

- First, the objectives of the analysis are identified and the performance indicators that will be used to evaluate alternative options are defined.
- Mathematical models are then developed by exploiting the relationships among system control variables, parameters and evaluation metrics.
- Finally, what-if analyses are carried out to evaluate the effects on the performance determined by variations in the control variables and changes in the parameters.

## 10. What are the three major components of business intelligence architecture?

- Data Sources
- Data warehouses and data marts
- Business intelligence methodologies



## 11. What are data sources and data warehouses?

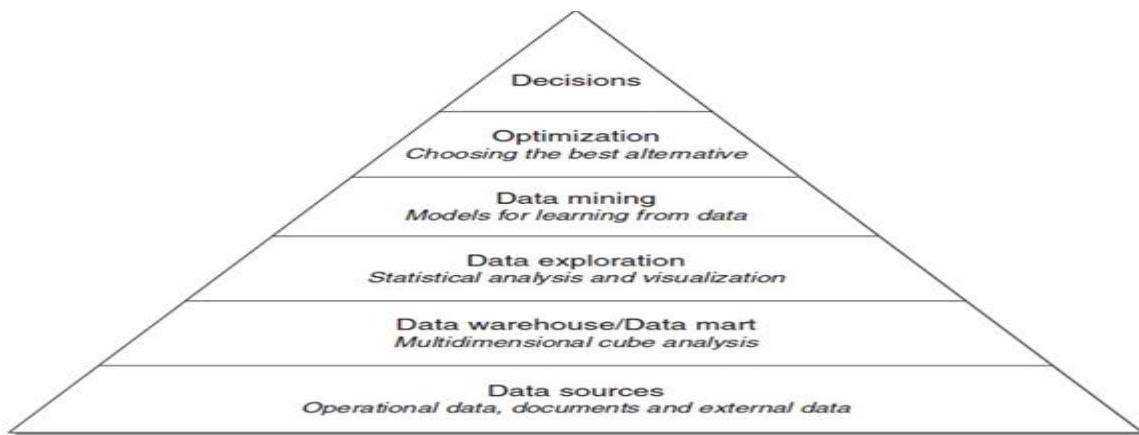
- Data sources-** In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type. The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such as emails and data received from external providers. Generally speaking, a major effort is required to unify and integrate the different data sources.
- Data warehouses -** Using extraction and transformation tools known as extract, transform and load (ETL), the data originating from the different sources are stored in databases intended to support business intelligence analyses. These databases are usually referred to as data warehouses.

## 12. What are business intelligence methodologies?

Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers. In a business intelligence system, several decision support applications may be implemented, such as:

- multidimensional cube analysis.
- exploratory data analysis.
- time series analysis.
- inductive learning models for data mining.
- optimization models.

**13. What are the main components of business intelligence system?**



**14. What is data exploration?**

At this level of pyramid, find the tools for performing a *passive* business intelligence analysis, which consist of query and reporting systems, as well as statistical methods. These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis tools to find answers and confirm their original insight

**15. What is data mining?**

At this level, it includes active business intelligence methodologies, whose purpose is the extraction of information and knowledge from data. These include mathematical models for pattern recognition, machine learning and data mining techniques. Unlike the tools described at the previous level of the pyramid, the models of an active kind do not require decision makers to formulate any prior hypothesis to be later verified. Their purpose is instead to expand the decision makers' knowledge.

**16. What are decisions?**

A decision corresponds to the choice and the actual adoption of a specific decision and in some way represents the natural conclusion of the decision-making process. Even when business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers, who may also take advantage of informal and unstructured information available to adapt and modify the recommendations and the conclusions achieved through the use of mathematical models.

**17. Who are all responsible for functioning of a good business intelligence system?**

The required competencies are provided for the most part by the information systems specialists within the organization, usually referred to as database administrators. Analysts and experts in mathematical and statistical models are responsible for the intermediate phases. Finally, the activities of decision makers responsible for the application domain appear dominant at the top.

**18. What are the phases involve in the cycle of business intelligence analysis?**

- (i) Analysis
- (ii) Insight
- (iii) Decision
- (iv) Evaluation

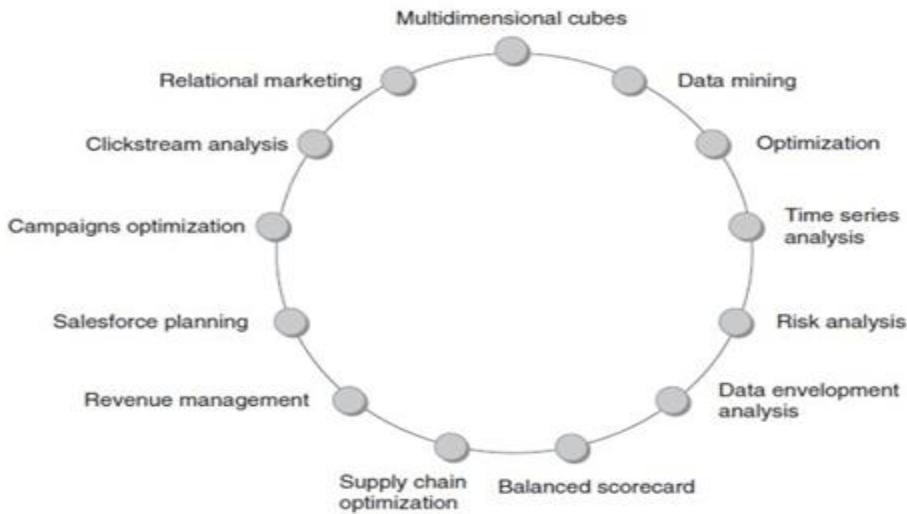
**19. What is analysis phase in the cycle of Business intelligence analysis?**

During the analysis phase, it is necessary to recognize and accurately spell out the problem at hand. Decision makers must then create a mental representation of the phenomenon being analyzed, by identifying the critical factors that are perceived as the most relevant. The availability of business intelligence methodologies may help already in this stage, by permitting decision makers to rapidly develop various paths of investigation. Thus, the first phase in the business intelligence cycle leads decision makers to ask several questions and to obtain quick responses in an interactive way.

**20. What are the phases of the development of a business intelligence system?**

- (i) Analysis
- (ii) Design
- (iii) Planning
- (iv) Implementation and control.

**21. What are the methodologies available in business intelligence system?**



**22. What is the need for decision support system?( NOV/DEC 2017)**

Decision support system is a computer solution using statistical data to help managers and operation planners overcome strategic deficiencies in order to implement streamlined, efficient solutions. DSS help those in positions of authority compile enough information to make an informed decision about changes in policy, implementation and so forth.

**23. What are structured, unstructured and semi-structured decisions NOV/DEC 2017)**

- Structured decisions are routine are repetitive decisions often taken at operational management level.
- Semi-structured decisions involve combination of standard procedures and unstructured elements taken at middle management level.
- Unstructured decisions are fuzzy , strategic and complex taken at top management level.

### PART-B

- 1) Explain the need and benefits of business intelligence system.
- 2) Explain business intelligence architecture in detail (or) Explain the framework of business intelligence in detail.(APRIL/MAY 2017)(NOV/DEC 2017)
- 3) Explain in detail about the role of mathematical models.
- 4) Explain all the phases in the cycle of business intelligence analysis
- 5) Explain all the phases in the development of business intelligence system in detail.(APRIL/MAY 2017)
- 6) Discuss about ethics to be followed in business intelligence.
- 7) Discuss briefly about the user interface in business intelligence system. (APRIL/MAY 2017)
- 8) Explain the scope for business intelligence and business value in the following applications  
(i) Customer segmentation (ii) Customer profitability (iii) Fraud detection (iv) Channel optimization. (NOV/DEC 2017)

### UNIT II

### KNOWLEDGE DELIVERY

The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Parameterized Reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message.

### PART-A

**1. What is Knowledge Delivery .**

Knowledge Delivery has evolved into data visualization, executive reporting, and real-time dashboards. The mainstream of Knowledge Delivery has adopted some strict rules, or grammar, associated with the presentation of information in dashboards, reports, etc.

**2. What are the Business Intelligence User types.**

Power users, Business users, Casual users, Data aggregators or Information Providers, Operational analytics users, extended enterprise users, IT users.

**3. What are the different modes of presentation that are relevant to different user types.**

- Standard reports
- Interactive Analysis and Ad Hoc Querying
- Parameterized Reports and Self-Service Reporting

**4. What are the groups of items being measured in standard reports.**

- The number of owner-occupied housing units with a mortgage
- Value of the houses
- Mortgage status
- Household income for the previous 12 months

**5. What are standard reports in BI.**

**Standard reports** usually have a fixed format, are parameter-driven and, in their simplest form, are pre-run. Standard reports provide a core set of information about what's going on in a particular business area. These reports are the backbone of BI applications.

**6. What is adhoc querying.**

An Ad-Hoc Query is a query that cannot be determined prior to the moment the query is issued. It is created in order to get information when need arises. Ad hoc queries are used intensively in the internet. Search engines process millions of queries every single second from different data sources. Any keywords typed the internet user are dynamically generated with an ad hoc query against virtually any database back end. As the basic structure of an SQL statement consist of SELECT keyword FROM table WHERE conditions, an ad hoc query dynamically supplies the keyword, data source and the conditions without the user knowing it.

**7. Write some of the caveats allowing users to formulate and execute adhoc queries.**

- **Performance.** Writing efficient queries is a skill, and many queries involve joins across multiple tables that can bring a system's performance to its knees.
- **Semantic consistency.** Allowing users to write their own queries implies they know and understand the meanings of the data elements they have selected to include in their result sets.
- **Repeatability.** The ad hoc process involves a sequence consisting of multiple iterations of the two-phased query and review of the result set process.

**8. What is parameterized report. (NOV/DEC 2017)**

The parameterized approach is particularly beneficial in operational scenarios in which similar queries and drill-downs are done over and over again. Parameterized reports bridges the gap between static, canned reports and free flowing ad hoc queries.

**9. What is self service business intelligence.**

Self-service business intelligence (SSBI) is an approach to data analytics that enables business users to access and work with corporate data even though they do not have a background in statistical analysis, business intelligence (BI) or data mining.

**10. What is Canned reports.**

prebuilt or “canned” report. These are typically distributed across a wide audience with polished formatting. Canned reports are short and sweet, batched and scripted data for often monthly or weekly metrics. Canned reports are the Pre defined reports ,They are reports that run daily, weekly, monthly and don't change.

**11. What are the different types of visualization modes for data.**

Bar chart, line chart, Pie chart, Scatter plot, Bubble chart, Gauge, Directional indicators (arrows up or down), Heat map, Spider or radar chart and Sparkline.

**12. What is adhoc analysis and adhoc reports.**

Ad hoc analysis may be used to create a report that does not already exist, or drill deeper into a static report to get details about accounts, transactions, or records. The process may be also used to get more current data for the existing areas covered by a static report. OLAP dashboards are specifically designed to facilitate ad hoc analysis by providing quick, easy access to data from the original report. The idea of ad hoc reporting is used to talk about 'one-off' or one-time reports that are done in a customized way, to provide results for one specific question or objective.

**13. Define OLAP.**

OLAP (online analytical processing) is computer processing that enables a user to easily and selectively extract and view data from different points of view. OLAP data is stored in a multidimensional\_database. It can be used for datamining or the discovery of previously undiscerned relationships between data items. OLAP products are typically designed for multiple-user environments.

**14. What is OLAP cube.**

A cube can be considered a multi-dimensional generalization of a two- or three-dimensional spreadsheet. For example, a company might wish to summarize financial data by product, by time-period, and by city to compare actual and budget expenses. Product, time, city and scenario (actual and budget) are the data's dimensions.

**Slice** is the act of picking a rectangular subset of a cube by choosing a single value for one of its dimensions, creating a new cube with one fewer dimension.

**Dice:** The dice operation produces a subcube by allowing the analyst to pick specific values of multiple dimensions

**15. What is Gauge.**

A gauge is an indicator of magnitude in the context of critical value ranges. A gauge is good for conveying relative status of critical variables and points that should trigger some action. A traditional example is an automobile's fuel gauge, which indicates the relative fullness of the tank, as well as an area close to the "empty" measure marked as red to indicate the need for refueling.

**16. What is scorecards and dashboards.**

**Score cards:** A scorecard is a data visualization tool that helps organizations, individuals, or groups of individuals to reach goals by displaying progress toward objectives against the objectives themselves. A report that gives information about the status, condition, or success of someone or something.

**Dashboards:** Dashboards often provide at-a-glance views of KPIs (key performance indicators) relevant to a particular objective or business process (e.g. sales, marketing, human resources, or production). In real-world terms, "dashboard" is another name for "progress report" or "report."

**17. List Some characteristics of business processes that are nicely suited to integrated analytics.**

- The business process has distinct performance objectives.
- The business process involves decision points by one or more actors.
- The process's performance can be impaired by absence of information.
- The process's performance can be impaired by ill-informed decisions.
- The process can be improved with well-informed decision-making.
- Participants do not need to be "tech-savvy" to be informed.

**18. Who are Operational analytics users?**

Who indirectly rely on the results of analytics embedded within operational applications. Examples include call center representatives whose scripts are adjusted interactively in relation to customer profiles, predicted behavioural predispositions, and real-time customer responses, web site offers and ad placement, or users of retail shelf management systems that adjust stock levels based on demand across multiple regions.

**19. What is sparkline.**

Sparklines are small line graphs without axes or coordinates. Many sparklines can be used in relative comparison regarding trends. As an example, the trends of different stock price histories for similar companies can be compared to determine if there are industry trends relating to stock price.

**20. What is integrated analysis? (APRIL/MAY 2017)**

- It implies the need for real-time integration of data from multiple sources of both analytics results and operational data. In turn, the delivery of the actionable knowledge must be seamlessly presented in a way that is best suited to business operations and is seamlessly integrated with the common suites of desktop productivity tools.

**21. Write the major issues in modeling. (APRIL/MAY 2017)**

Some models are more opaque than others; that is, it's hard to understand the logic the model used to identify relevant patterns and relationships in the data. The problem with these models is that business people often have a hard time trusting them until they see quantitative results, such as reduced costs or higher revenues. Getting business users to understand and trust the output of analytical models is perhaps the biggest challenge in data mining.

**22. What is interactive analysis? (NOV/DEC 2017)**

Interactive analysis often manifested as a pivot table. These pivot tables enable broader flexibility in grouping data within ordered hierarchies, development of static graphs and charts, or just perusing the data from different angles.

**23. Write some guidelines for laying out a BI dashboard.**

- Choose the right visualization graphic
- Manage your “real estate.”
- Maintain context
- Be consistent
- Keep it simple

**PART B**

- 1. Explain in detail on business intelligence user types.**
- 2. Explain standard report with an example. (NOV/DEC 2017)**
- 3. Explain Interactive Analysis and Ad Hoc Querying with an example. (APRIL/MAY 2017) (NOV/DEC 2017)**
- 4. Explain in detail on Parameterized Reports and Self-Service Reporting. (APRIL/MAY 2017) (NOV/DEC 2017)**
- 5. Explain Alerts/Notification method.**
- 6. Explain some of the visualization modes for data.**
- 7. (i) Explain Scorecards and Dashboards with example. (APRIL/MAY 2017) (NOV/DEC 2017)**  
**(ii) Explain Geographic visualization in detail. (APRIL/MAY 2017) (NOV/DEC 2017)**
- 8. Explain in detail about Integrated Analytics.**
- 9. How to optimize the presentation of the right message?**
- 10. Explain OLAP with slice and dice example.**

**UNIT III**

**EFFICIENCY**

**Efficiency measures – The CCR model: Definition of target objectives- Peer groups – Identification of good operating practices; cross efficiency analysis – virtual inputs and outputs – Other models. Pattern matching – cluster analysis, outlier analysis.**

**PART-A**

**1. What is data envelopment analysis?**

Data envelopment analysis is to compare the operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plants, or transportation systems. In order for the comparison to be meaningful, the units being investigated must be homogeneous.

**2. What do you mean by productivity indicator?**

Data envelopment analysis relies on a productivity indicator that provides a measure of the efficiency that characterizes the operating activity of the units being compared. This measure is based on the results obtained by each unit, which will be referred to as outputs, and on the resources utilized to achieve these results, which will be generically designated as inputs or production factors.

**3. What are decision making units (DMU)?**

In DEA the units being compared are called decision making units (DMUs), as they enjoy a certain decisional autonomy.

**4. How to find the efficiency of a decision making unit?**

Lets evaluate the efficiency of  $n$  units, where  $N = \{1, 2, \dots, n\}$  denote the set of units being compared.

If the units produce a single output using a single input only, the efficiency of the  $j$ th decision making unit  $DMU_j, j \in N$ , is defined as  $\theta_j = y_j/x_j$ , in which  $y_j$  is the output value produced by  $DMU_j$  and  $x_j$  the input value used.

**5. How to find the efficiency of a DMU with multiple outputs?**

If the units produce multiple outputs using various input factors, the efficiency of  $DMU_j$  is defined as the ratio between a weighted sum of the outputs and a weighted sum of the inputs. Denote by  $H = \{1, 2, \dots, s\}$  the set of production factors and by  $K = \{1, 2, \dots, m\}$  the corresponding set of outputs. If  $x_{ij}, i \in H$ , denotes the quantity of input  $i$  used by  $DMU_j$  and  $y_{rj}, r \in K$ , the quantity of output  $r$  obtained,

The efficiency of  $DMU_j$  is defined as  $\theta_j = (u_1y_{1j} + u_2y_{2j} + \dots + u_my_{mj})/(v_1x_{1j} + v_2x_{2j} + \dots + v_sx_{sj}) = (\sum_{r \in K} u_r y_{rj}) / (\sum_{i \in H} v_i x_{ij})$ , for weights  $u_1, u_2, \dots, u_m$  associated with the outputs and  $v_1, v_2, \dots, v_s$  assigned to the inputs.

## 6. What is efficient frontier?

The efficient frontier, also known as production function, expresses the relationship between the inputs utilized and the outputs produced. It indicates the maximum quantity of outputs that can be obtained from a given combination of inputs. At the same time, it also expresses the minimum quantity of inputs that must be used to achieve a given output level. Hence, the efficient frontier corresponds to technically efficient operating methods

## 7. What are called the production possibility sets?

slope of the line connecting each point to the origin represents the efficiency value associated with the corresponding branch. The line with the maximum slope, represented by a solid line, is the efficient frontier for all branches being analyzed. The branches that are on this line correspond to efficient units, while the branches that are below the efficient frontier are inefficient units. The area between the efficient frontier and the positive horizontal semi-axis is called the production possibility set.

## 8. What is input oriented efficiency and output oriented efficiency.

Efficient frontier provides some indications for improving the performance of inefficient units. Indeed, it identifies for each input level the output level that can be achieved in conditions of efficiency. By the same token, it identifies for each output level the minimum level of input that should be used in conditions of efficiency. (i.e)for each  $DMU_j, j \in N$ ,

- (i) Input oriented efficiency  $\theta_{Ij}$  can be defined as the ratio between the ideal input quantity  $x^*$  that should be used by the unit if it were efficient and the actually used quantity  $x_j$  : 
$$\boxed{\theta_{Ij} = x^*/x_j}$$
- (ii) Similarly, the output oriented efficiency  $\theta^{Oj}$  is defined as the ratio between the quantity of output  $y_j$  actually produced by the unit and the ideal quantity  $y^*$  that it should produce in conditions of efficiency: 
$$\boxed{\theta^{Oj} = y^*/y_j}$$

## 9. How to convert inefficient unit to efficient?

Inefficient unit can be brought close to the efficient frontier. In this case, the inefficiency of a given unit is evaluated by the length of the segment connecting the unit to the efficient frontier along the line passing through the origin of the axes.

$$\boxed{\theta_A = \frac{OP}{OA},}$$

The inefficient unit may be made efficient by a displacement along segment  $OA$  that moves it onto the efficient frontier. Such displacement is tantamount to progressively decreasing the quantity of both inputs while keeping unchanged the quantity of output.

## 10. What is CCR model ?

Using data envelopment analysis, the choice of the optimal system of weights for a generic  $DMU_j$  involves solving a mathematical optimization model whose decision variables are represented by the weights  $u_r$ ,  $r \in K$ , and  $v_i$ ,  $i \in H$ , associated with each output and input. Various formulations have been proposed, the best-known of which is probably the Charnes-Cooper-Rhodes (CCR) model. The CCR model formulated for  $DMU_j$  takes the form

$$\begin{aligned} \max \quad & \vartheta = \sum_{r \in K} u_r y_{rj}, \\ \text{s.to} \quad & \sum_{i \in H} v_i x_{ij} = 1, \\ & \sum_{r \in K} u_r y_{rj} - \sum_{i \in H} v_i x_{ij} \leq 0, \quad j \in N, \end{aligned}$$

## 11. What is the formulation taken by the CCR model for decision making units?

- (i) The objective function involves the maximization of the efficiency measure for  $DMU_j$
- (ii) Constraints require that the efficiency values of all the units, calculated by means of the weights system for the unit being examined, be lower than one.
- (iii) Conditions guarantee that the weights associated with the inputs and the outputs are non-negative.

## 12. What are target objectives?

Based on the efficiency values, data envelopment analysis therefore gives a measure for each unit being compared of the savings in inputs or the increases in outputs required for the unit to become efficient. To determine the target values, it is possible to follow an input- or output oriented strategy.

**13. How to determine the target values?**

Target values can be determined in two cases.

(i) In the first case, the improvement objectives primarily concern the resources used, and the target values for inputs and outputs are given by

$$x_{ij}^{\text{target}} = \vartheta^* x_{ij} - s_i^{-*}, \quad i \in \mathcal{H},$$

$$y_{rj}^{\text{target}} = y_{rj} + s_r^{+*}, \quad r \in \mathcal{K}.$$

(ii) In the second case, target values for inputs and outputs are given by

$$x_{ij}^{\text{target}} = x_{ij} - \frac{s_i^{-*}}{\vartheta^*}, \quad i \in \mathcal{H},$$

$$y_{rj}^{\text{target}} = \frac{y_{rj} + s_r^{+*}}{\vartheta^*}, \quad r \in \mathcal{K}.$$

**14. What are the other performance improvement strategies?**

The target values for the inputs are set in such a way as to minimize the quantity used of the resources to which the highest priority has been assigned, without allowing variations in the level of other inputs or in the outputs produced; The target values for the outputs are set in such a way as to maximize the quantity produced of the outputs to which highest priority has been assigned, without allowing variations in the level of other outputs or inputs used; Preferences expressed by the decision makers with respect to a decrease in some inputs or an increase in specific outputs.

**15. What are peer groups in DEA?**

DEA identifies for each inefficient unit a set of excellent units, called a peer group, which includes those units that are efficient if evaluated with the optimal system of weights of an inefficient unit. The peer group, made up of DMUs which are characterized by operating methods similar to the inefficient unit being examined, is a realistic term of comparison which the unit should aim to imitate in order to improve its performance.

**16. Define outlier analysis?**

The data objects which do not comply the general behavior or model of the data, then such data are grossly different or inconsistent with the remaining set of data called outliers. Thus the outlier detection is a data mining task known as outlier analysis.

**17. What is the use of weight restrictions?**

The units that are really efficient are separated from those whose efficiency score largely depends on the selected weights system, we may impose some restrictions on the value of the weights to be associated with inputs and outputs. In general, these restrictions translate into the definition of maximum thresholds for the weight of specific outputs or minimum thresholds for the weight of specific inputs.

**18. What is cross efficiency analysis? (NOV/DEC 2017)**

The analysis of cross efficiency is based on the definition of the efficiency matrix which provides information on the nature of the weights system adopted by the units for their own efficiency evaluation. The square efficiency matrix contains as many rows and columns as there are units being compared. The generic element  $\theta_{ij}$  of the matrix represents the efficiency of DMU $j$  evaluated through the optimal weights structure for DMU $i$ , while the element  $\theta_{jj}$  provides the efficiency of DMU $j$  calculated using its own optimal weights.

**19. Define cluster analysis?**

Cluster analysis groups data objects based only on information found in the data that describes the object ts and their relationships. The goal is that the objects within a group be similar(or related) to one another and different from the objects in other groups. The greater the similarity within a group and greater the difference between the groups makes the better, or more distinct the clustering.

**20. What are virtual inputs and outputs in DEA? (NOV/DEC 2017)**

Virtual inputs and virtual outputs provide information on the relative importance that each unit attributes to each individual input and output, for the purpose of maximizing its own efficiency score. The virtual inputs of a DMU are defined as the product of the inputs used by the unit and the corresponding optimal weights. Similarly, virtual outputs are given by the product of the outputs of the unit and the associated optimal weights.

**21. Define linear regression. (APRIL/MAY 2017)**

Linear regression is a statistical technique for analyzing data in order to obtain a measure of correlation between two variables where the relationship between the variables is expected to be linear. Regression is mostly used in predictive analysis.

**22. What is link analysis? (APRIL/MAY 2017)**

Link analysis is based on a branch of mathematics called graph theory, which represents relationships between different objects as edges in a graph. Link analysis is not a specific modeling technique, so it can be used for both directed and undirected data mining. It is often used for creating new derived variables for use by other modeling techniques. It can also be used for undirected data mining, by exploring the properties of the graphs themselves.

**PART-B**

- 1. Explain in detail about Efficiency measures and Efficient frontier.**
- 2. Explain the CCR model in detail. (APRIL/MAY 2017)**
- 3. Discuss in detail about the methods of identifying good operating practices.**
- 4. Explain the ways of converting inefficient units to efficient unit?**
- 5. Explain cluster analysis in detail. (APRIL/MAY 2017)**
- 6. Explain outlier analysis in detail. (APRIL/MAY 2017)**
- 7. Write short notes on cross efficiency analysis. (APRIL/MAY 2017)**
- 8. Explain about hierarchical clustering analysis in detail with an example. (NOV/DEC 2017)**
- 9. Explain about peer group arrangement and identification of good operating practices for Case-Based Reasoning (CBR). (NOV/DEC 2017)**

**UNIT IV****BUSINESS INTELLIGENCE APPLICATIONS**

Marketing models – Logistic and Production models – Case studies.

**PART-A****1. Write short notes on Marketing decision processes?**

Marketing decision processes are characterized by a high level of complexity due to the simultaneous presence of multiple objectives and countless alternative actions resulting from the combination of the major choice options available to decision makers

**2. What is relational marketing?**

The aim of a relational marketing strategy is to initiate, strengthen, intensify and preserve over time the relationships between a company and its stakeholders, represented primarily by its customers, and involves the analysis, planning, execution and evaluation of the activities carried out to pursue these objectives.

**3. What is customer relationship management (CRM)? (NOV/DEC 2017)**

Customer relationship management (CRM) is a term that refers to practices, strategies and technologies that companies use to manage and analyze customer interactions and data throughout the customer lifecycle, with the goal of improving business relationships with customers, assisting in customer retention and driving sales growth.

**4. Write something about sales force automation (SFA)?**

Sales force automation (SFA) is an integrated application of customizable customer relationship management (CRM) tools that automate and streamline sales inventory, leads, forecasting, performance and analysis. SFA tools include Web-based (hosted CRM) and in-house systems. SFA is also known as sales force management system.

**5. Why do we have relational marketing analysis?**

Relational marketing analyses can be used to identify customers who are more likely to take up the offer of additional services and products (cross-selling), or of alternative services and products of a higher level and with a greater profitability for the enterprise (up-selling).

**6. How the Acquisition can be done, why?**

Although retention plays a prominent role in relational marketing strategies, for many companies the acquisition of new customers also represents a critical factor for growth. The acquisition process requires the identification of new prospects, as they are potential customers who may be totally or partially unaware of the products and services offered by the company.

**7. What is Cross-selling?**

The term cross-selling refers to the attempt to sell an additional product or service to an active customer, already involved in a long-lasting commercial relationship with the enterprise. By means of classification models, it is possible to identify the customers characterized by a high probability of accepting a cross-selling offer, starting from the information contained in the available attributes.

**8. Write brief notes on Market basket analysis?**

The purpose of market basket analysis is to gain insight from the purchases made by customers in order to extract useful knowledge to plan marketing actions. It is mostly used to analyze purchases in the retail industry and in e-commerce activities, and is generally amenable to unsupervised learning problems. It may also be applied in other domains to analyze the purchases made using credit cards, the complementary services activated by mobile or fixed telephone customers, the policies or the checking accounts acquired by a same household. Each transaction consists of a list of purchased items. This list is called a basket.

**9. What is web mining?**

The web is a critical channel for the communication and promotion of a company's image. Moreover, e-commerce sites are important sales channels. Hence, it is natural to use web mining methods in order to analyze data on the activities carried out by the visitors to a website.

Web mining methods are mostly used for three main purposes,

- (i) Content mining (ii) Structure mining (iii) Usage mining.

**10. What does meant by salesforce?**

The term salesforce is generally taken to mean the whole set of people and roles that are involved, with different tasks and responsibilities, in the sales process

**11. Write short notes on content mining?**

Content mining involves the analysis of the content of web pages to extract useful information. Search engines primarily perform content mining activities to provide the links deemed interesting in relation to keywords supplied by users. Content mining methods can be traced back to data mining problems for the analysis of texts, both in free format or HTML and XML formats, images and multimedia content.

**12. Write short notes on structure mining?**

The aim of this type of analysis is to explore and understand the topological structure of the web. Using the links presented in the various pages, it is possible to create graphs where the nodes correspond to the web pages and the oriented arcs are associated with links to other pages. Results and algorithms from graph theory are used to characterize the structure of the web.

**13. Write short notes on usage mining?**

Analyses aimed at usage mining are certainly the most relevant from a relational marketing standpoint, since they explore the paths followed by navigators and their behaviors during a visit to a company website. Methods for the extraction of association rules are useful in obtaining correlations between the different pages visited during a session.

**14. Write down the taxonomy of salesforces?**

A preliminary taxonomy of salesforces is based on the type of activity carried out, as indicated below.

- **Residential.** Residential sales activities take place at one or more sites managed by a company supplying some products or services, where customers go to make their purchases. This category includes sales at retail outlets as well as wholesale trading centers and cash-and-carry shops.
- **Mobile.** In mobile sales, agents of the supplying company go to the customers' homes or offices to promote their products and services and collect orders. Sales in this category occur mostly within B2B relationships, even though they can also be found in B2C contexts.
- **Telephone.** Telephone sales are carried out through a series of contacts by telephone with prospective customer.

**15. List the decision-making processes relative to salesforce management?**

The decision-making processes relative to salesforce management can be grouped into three categories:

- (i) Design. (ii) Planning. (iii)Assessment.

**16. What are Response functions?**

Response functions play a key role in the formulation of models for designing and planning a sales network. In general terms, a response function describes the elasticity of sales in terms of the intensity of the sales actions, and is a formal method to describe the complex relationship existing between sales actions and market reactions

**17. Write about Sales territory design?**

Sales territory design involves allocating sales coverage units to individual agents so as to minimize a weighted sum of two terms, representing respectively the total distance between areas belonging to the same territory and the imbalance of sales opportunities for the agents.

**18. Define Supply Chain?**

A supply chain may be defined as a network of connected and interdependent organizational units that operate in a coordinated way to manage, control and improve the flow of materials and information originating from the suppliers and reaching the end customers, after going through the procurement, processing and distribution subsystems of a company.

**19. What is Backlogging?**

The term backlog refers to the possibility that a portion of the demand due in a given period may be satisfied in a subsequent period, incurring an additional penalty cost. Backlogs are a feature of production systems more likely to occur in B2B or make-to-order manufacturing contexts. In B2C industries, such as mass production consumer goods, on the other hand, one is more likely to find a variant of the backlog, known as lost sales, in which unfulfilled demand in a period cannot be transferred to a subsequent period and is lost.

**20. State minimum lot conditions?**

A further feature often appearing in manufacturing systems is represented by minimum lot conditions: for technical or scale economy reasons, it is sometimes necessary that the production volume for one or more products be either equal to 0 (i.e. the product is not manufactured in a specific period) or not less than a given threshold value, the minimum lot.

**21. What is Revenue management?**

Revenue management is a managerial policy whose purpose is to maximize profits through an optimal balance between demand and supply. It is mainly intended for marketing as well as logistic activities and has found growing interest in the service industry, particularly in the air transportation, tourism and hotel sectors. More recently these methods have also begun to spread within the manufacturing and distribution industries.

**22. State the use of data mining and business intelligence in health care department. (APRIL/MAY 2017)**

Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. Some experts believe the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. Data mining involves uncovering patterns from vast data stores and using that information to build predictive models.

**23. Define mental models. (APRIL/MAY 2017)**

Mental model is a explanation of someone's thought process about how something works in the real world. It is a representation of the relationships between its various parts and a person's intuitive perception about his or her own acts and their consequences.

**24. Comment on information broadcasting tools for business intelligence. (NOV/DEC 2017)**

Information broadcasting allows you to make objects with business intelligence content available to a wide spectrum of users, according to your own requirements. It also offers functions to optimize performance and for exception reporting. BEx Information Broadcasting is the Business Explorer component you use to distribute BI objects that were generated with the various BEx tools. With the BEx Broadcaster, you can precalculate queries, query views, Web templates, reports and workbooks and broadcast them by e-mail or to the portal.

**PART-B**

- 1. Explain in detail about Relational Marketing components and decision making option with suitable examples.**
- 2. Explain in detail about an environment for relational marketing analysis.**
- 3. Describe in brief given below,**
  - I. Acquisition**
  - II. Retention**
  - III. Cross-selling and up-selling**
  - IV. Market basket analysis.**
- 4. Explain in detail the Taxonomy of web mining analysis.**

5. **Describe salesforce management.** Explain in detail the decision processes, models in salesforce management.
6. Explain in detail calls and product presentations planning with optimization model?
7. **Describe supply chain optimization.** Explain in detail the Optimization models for logistics planning.
8. **What is Revenue management system?** Explain decision processes in revenue management.
9. Do and explain the case studies of Logistics planning in the food industry.
10. Explain about marketing models. (NOV/DEC 2017)
11. Explain in detail about Logistics and production model for business intelligence. (APRIL/MAY 2017)(NOV/DEC 2017)

**UNIT V****FUTURE OF BUSINESS INTELLIGENCE**

**Future of business intelligence – Emerging Technologies, Machine Learning, Predicting the Future, BI Search & Text Analytics – Advanced Visualization – Rich Report, Future beyond Technology.**

**PART-A****1. Write about workload automation?**

Sophisticated automation tools are at the forefront of technological advances that are paving the way for the simplification of BI and play a considerable role in sustaining hybrid or centralized BI architecture. Formerly, lengthy querying and reporting jobs that involved time-consuming scripts can now be expedited due to self-service and script less automation job scheduling. With the proper IT configuration, end users are able to issue on-demand queries in close to real time with some of the most viable applications for BI such as Big Data and data from conventional sources. These queries greatly simplify the integration of BI with other platforms.

**2. What is Data Discovery?**

Discovery tools have lessened the need for conventionally lengthy (and time-consuming) BI reports, and significantly augmented them in cases in which they're essential. Dashboards and interactive visualizations graphically represent data and results from BI in ways in which trends are readily discernible, data smashups and in-memory analytics enable users to quickly query a variety of disparate sources, and search tools offer text-derived analysis of either structured or unstructured quantitative and qualitative data.

**3. What are BI technology evaluation?**

- Embrace
- Adopt Where Appropriate
- Evaluate and Test
- Monitor and Understand

**4. What are the Emerging Trends in Business Intelligence?**

- Data Discovery Accelerates Self-Service BI and Analytics
- Unified Access and Analysis of All Types of Information Improves User Productivity
- Big Data Generated by Social Media Drives Innovation in Customer Analytics
- Text Analytics Enables Organizations to Interpret Social Media Sentiment Trends and Commentary
- Decision Management Enables Organizations to be Predictive and Proactive in Real Time

**5. What is the difference between BI Analytics and Predictive Analytics?**

BI Analytics	Predictive Analytics
Business intelligence allows you to answer questions about the demographics or characteristics of your customers, products, stores, etc., and answer questions about the performance of your business across a number of different dimensions	Predictive analytics allow organizations to go beyond the answers generated by BI by providing more predictive answers and recommendations to many of the same questions
Example: ➤ when did customer X last visit the store. ➤ How much revenue did store X generate last Christmas?	Example: ➤ How many customers are likely to visit the store next week. ➤ How much revenue will be generated by store X next Christmas?

**6. List the Top Business Intelligence Software Products?**

- Sisense
- Corporater EPM Suite
- SAP BusinessObjects
- Hyperion
- PowerCenter
- Cognos
- Active Intelligence
- Teradata Database

**7. Define BI search?**

A BI Search interface promises to change the way users' access information. Picture a Google interface to BI. Without any training in a BI tool, users can enter a phrase such as "Recent sales for customer A" and then be presented with either a list of predefined reports or, in some cases, a newly generated query.

**8. What are the Potential Benefits of BI Search?**

- Self service for report consumers is the best reason for BI search.
- BI search is mostly about finding pre-built reports, not creating new ones.
- BI search uncovers more facts for decision making.
- Users associate ease of use with BI search

**9. What is text analytics?**

Text analytics is closely related to search in that unstructured information or text can be transformed into quantitative data. For example, it allows for searching of information in a comment field to see how many times a customer praised a particular product. Text analytics is the numerical analysis of textual information.

**10. What is the difference between structured and unstructured data?**

Structured data (quantitative)	Unstructured data(textual)
Structured data refers to the numerical values typically captured in the operational systems and subsequently stored in a data warehouse	Unstructured content refers to information stored in textual comment fields, documents, annual reports, websites, and so on.
Example: ➤ RDMS data	Example: ➤ Textual data ➤ Social network data

**11. Write notes on advanced visualization?**

Advanced visualization goes beyond a simple chart such as a bar or line chart to includes:

- Spark lines, a highly condensed trend line the size of a word.
- Bullet graphs, a construct by Stephen Few that includes a target indicator within the bar chart.
- Small multiples, which are series of small, similar graphics or charts.
- Heat maps that display two variables as different intensifying colors.
- Decomposition trees, a visualization that displays each drill-down akin to an ever-expanding organization chart.
- Geographic maps that display things such as sales figures in a map form, using color to highlight sales performance. By mousing over a particular country, region, or state, you can see the individual data values.

**12. What is Rich reportlets?**

- Rich reportlets are powered by Web 2.0 technologies to create rich Internet applications (RIA). When BI suites were first re-architected for the Web, report consumers could only view a static page.
- With rich reportlets, someone accesses a report over the Web but in a much more interactive and appealing way.
- At a simple click, data can be re-sorted, filtered, or graphed, without having to launch a complicated report editor.

**13. What are the Future Trends of Business Intelligence?**

Looking beyond the obvious trends (Social, Mobile, Cloud), would like to focus on three trends brewing at the tectonic levels of the BI industry.

Business intelligence tools need to be,

- (i)Simplified (ii) Specialized (iii) Personalized

**14. Define the term Decision management?**

Decision management is the term industry experts and vendors use to describe the integration of analytics with business rules and process management systems to achieve a predictive and proactive posture in a real-time world.

**15. Write the Scope of BI Search?**

The list of files, documents, reports, and systems indexed by a BI search implementation constitute its scope. The scope of the implementation determines many things, including what's visible through the search index (all else is invisible) and what elements of the indexed source are included in the index.

**16. What are the machine learning strategies?**

Machine learning can be done by applying specific learning strategies, such as:

- A supervised strategy to map the data inputs and model them against desired outputs.
- An unsupervised strategy, to map the inputs and model them to find new trends.

**17. What is Machine Learning?**

In simple terms, machine learning is a branch of the larger discipline of Artificial Intelligence, which involves the design and construction of computer applications or systems that are able to learn based on their data inputs and/or outputs. Basically, a machine learning system learns by experience; that is, based on specific training, the system will be able to make generalizations based on its exposition to a number of cases and then be able to perform actions after new or unforeseen events.

**18. What are the three components of learning algorithms?**

Learning = Representation + Evaluation + Optimization

- **Representation** means the use of a classifier element represented in a formal language that a computer can handle and interpret;
- **Evaluation** consists of a function needed to distinguish or evaluate the good and bad classifiers;
- **Optimization** represents the method used to search among these classifiers within the language to find the highest scoring ones.

**19. Difference between artificial and business intelligence?**

Artificial intelligence	Business intelligence
This subject deals with Logic, Reasoning, Graph traversing/Mining etc. It deals with automatic ways of reasoning and reaching to a conclusion by computers.	This field is even more business goal focused than Data Mining.
Some of the algorithms of this domain are BFS, DFS, A*, Djikstra, Best First, Backtracking etc.	There is no algorithm for this rather it is a skill of converting raw data to useful information for the business.
Search and Optimization are two big use cases of AI.	Data sources could be internet search, internal business sources, web click data, customer feedback data etc.
Robot Navigation, Automatic Clinical Decision System, Knowledge representation	Business dashboards, Reports, SWOT Analysis

**20. What is Event-Streaming?**

Event-streamed data is the term for Big Data generated from sensors, instruments, and other monitoring systems that constantly produce data. Such data is frequently related to weather, global positioning, video monitoring, etc. Although sensor data is already used in industry specific applications such as oil mining or in retail stores via smartphones, the analytics and BI industry is preparing to make tools for such data more commonplace.

**21. List the application areas of text mining. (APRIL/MAY 2017)**

- Cybercrime prevention
- Customer care service
- Fraud detection
- Contextual Advertising

**22. What are emerging technologies in BI?**

- data visualizations,
- BI on clouds or SaaS,
- Hadoop,
- Apache Spark and Shark,
- mobile BI,
- social media,
- Internet of things,
- solid-state drives,
- streaming data

**23. Define discriminant analysis. (APRIL/MAY 2017)**

Discriminant analysis is a statistical tool with an objective to assess the adequacy of a classification. It is often used to complement the findings of cluster analysis and principal component analysis.

**24. Define information extraction. (NOV/DEC 2017)**

Extraction of information from a text in the form of text strings and processed text strings which are placed into slots labeled to indicate the kind of information that can fill them. Convert the benefits of powerful query tools such as SQL. This method of getting meaning from text is called Information Extraction.

**25. What is morphological analysis? (NOV/DEC 2017)**

Morphological analysis is a method developed by Fritz Zwicky for exploring all the possible solutions to a multi-dimensional, non-quantified complex problem. Consider a complex, real-world problem, like those of marketing or making policies for a nation, where there are many governing factors, and most of them cannot be expressed as numerical time series data, as one would like to have for building mathematical models

**PART-B**

1. Explain in detail the emerging technologies in future of business intelligence.
2. Explain in detail the framework of BI technology evaluation.
3. Explain in detail predicting the future using data mining algorithms with business intelligence technology.
4. Discuss in detail BI search and Text Analytics. (APRIL/MAY 2017)( NOV/ DEC '17)
5. Explain in detail Advanced Visualization with suitable examples. (APRIL/MAY 2017)
6. (i) Discuss in detail on Rich Reportlets.  
(ii) Discuss the future Beyond Technology.
7. Discuss the pros and cons of business intelligence technologies.
8. How artificial intelligence makes business more productively in the varying trends of business intelligence.
9. Explain in detail on machine learning algorithms.
10. Explain Internet of Things (IOT) in terms of business intelligence.
11. Explain the four steps of case based reasoning(CBR) process and Write the various benefits of case based Reasoning. (APRIL/MAY 2017)( NOV/ DEC '17)