

MODULE 2

CHAPTER 2

Data and Knowledge Management

University Prescribed Syllabus

Data and Knowledge Management : Database Approach, Big Data, Data warehouse and Data Marts, Knowledge Management.

Business Intelligence (BI) : Managers and Decision Making, BI for Data analysis and Presenting Results

2.1	Introduction.....	2-3
	UQ. Explain the importance of data in today's environment with an example. (MU - Q. 2(b), Jan 21, 5 Marks)	2-3
2.2	Data Governance	2-4
2.3	Database Approach.....	2-5
2.3.1	The Data Hierarchy	2-6
2.3.2	Designing the Database	2-6
2.3.3	The Relational Database Model	2-7
2.4	Big Data.....	2-7
	UQ. Define Big Data and discuss its basic characteristics? (MU - Q. 2(B), Dec. 19, 5 Marks)	2-7
2.4.1	Concept	2-7
2.4.2	Characteristics of Big Data	2-8
2.4.3	Benefits of Big Data.....	2-8
2.4.4	Issues with Big Data	2-9
2.5	Datawarehouses and Datamarts.....	2-9
2.5.1	Concept and Definition	2-9
2.5.2	Architecture of a Datawarehouse	2-10

UQ.	Explain the architecture of Data mart and Data warehouse in an organization. (MU - Q. 3(A), Dec. 19, 10 Marks).....	2-10
2.5.3	Benefits and Drawbacks of Datawarehousing.....	2-12
2.6	Knowledge Management Systems.....	2-12
2.6.1	Concept.....	2-12
UQ.	Describe what is meant by knowledge management. What factors have led to its development. (MU - Q. 2(A), Jan. 21, 5 Marks).....	2-12
2.6.2	Factors Leading to Development of knowledge Management Systems.....	2-13
2.6.3	The KMS Cycle.....	2-13
2.7	Managers and Decision Making	2-14
UQ.	Discuss the Impact of BI on Decision Making. (MU - Q. 3(B), Dec. 19, 10 Marks)	2-14
2.8	Business Intelligence (BI)	2-15
2.8.1	Concept	2-15
2.8.2	Business Intelligence Applications for Data Analysis	2-16
2.8.3	Data Mining	2-17
2.8.4	Decision Support Systems	2-18
2.8.5	Business Intelligence Applications for Presenting Results	2-19
2.9	Multiple Choice Questions.....	2-20
•	Chapter Ends	2-24

► 2.1 INTRODUCTION

UQ. Explain the importance of data in today's environment with an example.

MU - Q. 2(b), Jan 21, 5 Marks

- Information systems play a very vital role in managing data within organizations. Managing data involves accessing, processing and storing huge amounts and variety of data generated within organizations due to the daily transactions.
- These transactions could be anything from financial transactions, maintaining confidential information, reviews and feedback on social media, intellectual property.
- Data is plain facts and values that are generated during these day to day transactions.
- The data generated due to these transactions is really huge and cost involved to manage this data is even higher.
- But managing and maintaining this data is very critical for the business organization as data analysts will analyze and use this data as per their requirements to generate useful information.
- Senior managers can then apply their experience to use this information to solve critical business problems. This activity of addressing business problems will further lead to generation of knowledge.
- This knowledge will be captured and stored by knowledge management systems and can be accessed by anyone within the organization who is authorized to do so thereby creating flexible and powerful learning organizations.
- Hence this implies that data and knowledge management is very vital to modern business organizations.
- Managing data means maintaining the quality of data at high levels. The data should be accurate, complete, timely, consistent, accessible, relevant and concise.
- Say for example, Customer data would be the values or facts that are generated due to customer interactions like placing an order, making a booking or giving feedback etc. As this data generated is raw, it can be put up into a meaningful context to convert it into information.
- As a result of which proper interpretations can be made and will help the business organizations and management in making informed decisions.
- In the context of above examples related to customer data, useful metrics like negative customer reviews would help in making decisions on how to devise better ways to satisfy the customer.
- The value of the information lies in the actions that arise from the information. For example, if the information alerts you to poor customer satisfaction, it is useful only if this creates a change in the way the business deals with customers. Hence the information process should form part of a wider review process within the business to gain the best outcomes.
- But managing and maintaining the quality of data is not an easy task. The process is becoming really challenging with such high volumes and variety of data being generated.

- Few of the difficulties in managing data are discussed below:
 - o Most of the data needs to be kept for a long time as historical records for analysis purpose. New data keeps on adding exponentially with time.
 - o The data is collected, processed and distributed from and to different locations within the organization and using different devices and methods .
 - o Different sources of information include internal sources like company documents and corporate databases, external sources like corporate websites or government reports, personal feedbacks and reviews, blogs, digital footprints etc.
 - o Further the data are frequently stored on heterogeneous computing systems, databases and that too in different formats. These days the data that is generated is highly unstructured.
 - o Also data changes or even degrades over time, say for example customers change their names and addresses, new products are added or some products do not exist anymore.
 - o Data duplication and redundancy is another major issue. Like it happens in organizations where same data is maintained at several places like marketing, billing and even customer service function.
 - o Data is subjected to data rot over time due to temperature, humidity, and exposure to light. It can damage the storage media and make it difficult to retrieve the data.
 - o Legal constraints regarding data and information across countries differs which is also a major concern.
 - o Maintaining data integrity, ensuring its security is yet another issue to be considered while managing data.

► 2.2 DATA GOVERNANCE

- To address these numerous problems related to managing data, organizations are turning towards Data Governance.
- **Data Governance** is an approach to managing information across an organization. It involves people, business processes, policies and tools that are designed to ensure that data is handled in a certain well-defined fashion.
- There should be proper well defined procedures within organizations for acquiring, processing ,and storing data. Rules for handling and protecting data should be clearly mentioned and followed within the organization.
- The objective is to make the information available to only those people who are authorized to access it, from the moment it enters an organization until it is outdated and deleted.
- Data governance can be defined as a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

- Data governance clearly states who can take what action, upon what data, in what situations, using what methods.
- A well defined Data Governance policy will cover all strategic, tactical and operational level decisions and also justify how your business benefits from consistent business processes and responsibilities.
- One strategy for implementing data governance is Master Data Management.
 - o **Master data management (MDM)** deals with creating a single master record for all business critical data gathered from various sources either internal or external to the organization.
 - o It helps to create and maintain a single version of data which is available throughout all the functional units of the organization.
 - o The data is synchronized and consistent throughout various business processes within the organization.
 - o Master data management involves removing duplicates, standardizing data and incorporating rules to eliminate incorrect data from entering the system.

► 2.3 DATABASE APPROACH

- Until the early seventies, business organizations managed their data using file management environment.
- The applications were developed independently and as a result they maintained their own version of data independently in separate data files. These data files were collection of logically related records.
- This was fine until applications were few. As number of applications started increasing, each application maintaining its own version of data became chaotic. Same data was duplicated at many places. Some places the data was updated and at some places the older version of data was in use.
- Say for example, the sales department, marketing department and the customer support department were maintaining three different copies of the same customer data. Further it could happen that customer support department had new updated address whereas sales department had old postal address in their data file.
- These problems could be very well minimized using a Database Approach.
- In this approach, Database Management systems which are a collection of software programs are used to store, access and manage data.
- These systems minimize the problems of data redundancy, data inconsistency, data security, data integrity, data isolation etc.

2.3.1 The Data Hierarchy

- Data is always organized in an hierarchical fashion.
- To begin with we have the smallest unit of data which is a **binary digit or bit** represented in terms of 0 and 1.
- A group of 8 bits is one **byte**. A byte can be used for symbols, letters or characters.
- Group of characters form words and logical grouping of characters or words forms a **field**. Fields could be text, numbers, images etc. For example roll no, name, address, phone number etc.
- A logical grouping of related fields, such as the student's roll no., name, subject marks, make up a **record**.
- A logical grouping of related records is called a **table**. For example, a grouping of the records of student roll no, name, marks etc. would form a table of student details.
- And at the highest level a logical grouping of related files constitutes a **database**.

2.3.2 Designing the Database

- While designing a database care should be taken that data should be organized properly so that it can be accessed and understood with ease.
- A key tool for designing an effective database is the data model. A **data model** is used to represent entities and their relationships within the database. One such model is **Entity-Relationship diagram**.
- An **entity** is a person, place, thing, or event - such as a customer, an employee, or a product - about which information is maintained in the database. A record generally describes an entity.
- An **instance** of an entity is a specific, unique representation of that entity. For example, an instance of the entity EMPLOYEE would be details of a specific employee.
- Each characteristic or quality of a particular entity is called an **attribute**. For example, for the entity employee, attributes would include employee name, employee code, designation etc .
- Every record in a file must contain at least one field that uniquely identifies that record so that it can be retrieved, updated, and sorted.
- This field is called the **primary key**. For example, employee code is unique for every employee and it can be considered as primary key.
- At times for identifying a record in a database we also need additional fields. A **secondary key** is another field that has some identifying information but typically does not identify the record with complete accuracy always. For example, the employee's address can be considered as secondary key for identifying the employee records.
- As mentioned earlier, A **database management system (DBMS)** is a set of programs that provide users with tools to add, delete, access, modify, and analyze data stored in a database.
- An organization can access the data by using query and reporting tools that are a part of the DBMS.

- There are a number of different database architectures, but most commonly used are relational database model due to ease of use. Popular examples of relational databases are Microsoft Access, Oracle and SQL Server.

2.3.3 The Relational Database Model

- The **relational database model** is based on the concept of two-dimensional tables. Data is organized into simple tables made up of rows and columns. These models help in accessing and retrieving data quickly by finding intersection of rows and columns. Usually, a relational database is a collection of number of related tables. Every row in the table indicates a record and column represents attributes. Each of these tables contains records (listed in rows) and attributes (listed in columns).
- **Structured Query Language** is the most popular language used for requesting, searching or accessing information from the database.
- However, the relational databases have their limitations, like for large-scale enterprise wide databases may have many interrelated tables because of which the overall design can become complex and lead to slow search and access times.
- **Data Dictionary** is another important term used in relational database model. It is like metadata i.e. data about data. The data dictionary provides all information about each attribute, such as its name, its type, whether it is a key etc. Data dictionaries provide names and standard definitions for all attributes.

2.4 BIG DATA

UQ. Define Big Data and discuss its basic characteristics?

(MU - Q.2(B), Dec. 19, 5 Marks)

2.4.1 Concept

- Big Data implies data that is very huge, rather superabundance of data available today is Big Data.
- And it's just not huge but also growing exponentially with time as well as is of diverse types.
- So handling such variety of large data is highly impossible using simple database management tools but requires advanced tools to capture, process, transform and analyse this data.
- Big Data could be best explained with few examples like
 - o The New York Stock Exchange generates about *one terabyte* of new trade data per day.
 - o Around *500 + terabytes* of data in the form of photos, videos, messages, reviews etc. enters into the databases of social media sites like Facebook, every day.
 - o A single Jet engine can generate *so many terabytes* of data in *few minutes* of flight time collected from its various sensors.



2.4.2 Characteristics of Big Data

The most common characteristics of Big Data are **Volume, Variety, Velocity and Value**.

- i) **Volume** : Volume implies size of Big Data which is enormous. The size of data helps in deciding the value of data. As just discussed above one terabyte of data per day generated by New York Stock Exchange, many terabytes of data generated on social media daily etc., these are all examples of high volumes of data.
- ii) **Variety** : The next characteristic of Big Data is its Variety. It means that Big data is diverse in nature. It comes from various sources internal and external to the organization and is heterogenous in nature. It can be structured as in case of records stored in relational tables or unstructured as in multimedia data on websites. The data generated can be of any type from PDF documents , emails, to photos and videos,
- iii) **Velocity** : The pace with which Big data is generated is very fast. The speed with which data is generated and processed is termed as Velocity. Velocity can be best explained with example of a single Jet engine that can generate *so many terabytes* of data in *30 minutes* of flight time collected from its various sensors like temperature, pressure, fuel monitoring systems. The data generated on social media sites, from mobile devices and sensors is really extensive and continues. Google alone processes on an average more than 40,000 search queries every second.
- iv) **Value** : The next characteristic of Big data is the actual Value of this enormous data that is collected. Just volume is not sufficient but the data should be meaningful and of help to whichever organization that is collecting it. For example the customer feedback data collected should help in understanding the customers, their needs and grievances and accordingly change business policies to satisfy their needs.

2.4.3 Benefits of Big Data

Below are certain benefits of Big Data.

1. As just explained above the customer feedback data collected can be analyzed to help organizations in understanding the trends in customer needs and thereby build new products. It could also deal with pricing of the product i.e. giving better products at affordable prices.
2. There are certain tools of Big Data like Hadoop that are very cost effective when it comes to storing large amounts of data.
3. Also analysis of data collected from various internal and external sources helps in getting a better understanding of current market conditions.
4. The analysis of data and further decision making becomes very fast with the help of Big Data tools like Hadoop.
5. The organization can work on maintaining their brand image and reputation by considering the feedback and reviews posted by customers and thereby taking necessary actions.

The use of Big Data is becoming common these days by the companies to outperform their peers. In most industries, existing competitors and new entrants alike will use the strategies resulting from the analyzed data to compete, innovate and capture value.

❖ 2.4.4 Issues with Big Data

1. **Veracity** : Big Data comes from various sources of which all may not be trusted sources. Hence the authenticity and reliability of data and analysis performed on such data is questionable. For example, how authentic are the reviews posted on a shopping site by customers. Some negative feedbacks could be posted by competitors only to defame the business organization.
2. **Validity** : Big Data can be dirty and of poor quality. For example incomplete survey forms filled by customers. The data may not be accurate and complete. Such data cannot serve as a base for data analysis.
3. **Security** : Big data breaches are possible and can have disastrous impact. Around 87 million Facebook users had their profiles exposed by Facebook to Cambridge Analytica, a political consulting firm for some election campaign.
4. **Variable** : Big Data also changes due to heterogenous data sources and types. These inconsistencies lead to difficulty in handling and managing data.

► 2.5 DATAWAREHOUSES AND DATAMARTS

❖ 2.5.1 Concept and Definition

- Any organization needs a database to manage its data. Sometimes it could be more than one databases. But the problem with databases is that they process data in real time or near real time. So if any analysis has to be performed it is not advisable to work on real time databases as they would keep changing as new data enters in real time. To avoid this problem, Data warehouses have been developed to allow users to access data for analysis and decision making.
- Consider an example, The profit margin on second hand cars sold by company like Cars21 can be easily found using SQL. But now suppose the trend in profit margin on second hand cars over last 15 years in a particular geographic location say any city needs to be found, then writing the query would be very complex. So, for this reason building Data warehouses and datamarts is becoming essential.
- For accessing, analyzing and processing data efficiently within Data warehouses and datamarts various Online Analytical Processing tools, Data Mining tools are available
- Data warehouses and datamarts are also used to support number of business intelligence applications within organizations for better decision making.
 - o **Business intelligence (BI)** is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions.

- A **datawarehouse** is a repository of historical data that are organized by business subject/dimension to support decision makers in the organization.
- As maintaining Data warehouses is very expensive, they can be afforded majorly by large companies.
- For smaller organizations or strategic business units a datamart is sufficient
- A **datamart** is a low-cost, scaled-down version of a Data warehouse that is designed for the end-user needs in a strategic business unit (SBU) or an individual department.
- Wherever there are very few BI applications, datamart are sufficient, rather than a Data warehouse.
- The basic characteristics of Data warehouses and datamarts are:
 - In contrast to transactional databases which are organized by business process, Data warehouses and datamarts are organized by business subject or dimension. Usually in business organizations customer, product, vendor, geographic location can be considered as business subjects.
 - As opposed to databases that use Online Transaction Processing (OLTP), Data warehouses and datamarts are designed for Online Analytical Processing (OLAP) that involves analysis of accumulated data for decision making.
 - Data about the business subject is integrated from multiple internal and external sources and accumulated at one place.
 - Unlike transactional databases that contain only recent data(of about days ,weeks or months), Data warehouses and datamarts store very old historical data(of about years)which is used for analysis purpose.
 - The data within Data warehouses and datamarts does not change frequently i.e.it is nonvolatile. Users cannot update the warehouses very often. Whenever they need updation it is done through IT-controlled load processes rather than by users directly.
 - Data warehouses or marts are multidimensional unlike databases which are two dimensional.

2.5.2 Architecture of a Data warehouse

UQ. Explain the architecture of Data mart and Data warehouse in an organization.

(MU - Q. 3(A), Dec. 19, 10 Marks)

The basic block diagram of a Data warehouse and datamart architecture is shown in Fig. 2.5.2. It includes:

- The source systems that provide data to the data warehouse or datamart.
- The data-integration technology and processes that prepare the data for use.
- Different architectures for storing data in an organization's Data warehouse or datamarts.
- Different tools and applications for the variety of users.

- Metadata, data quality and governance processes that ensure that the warehouse or mart meets its purpose.

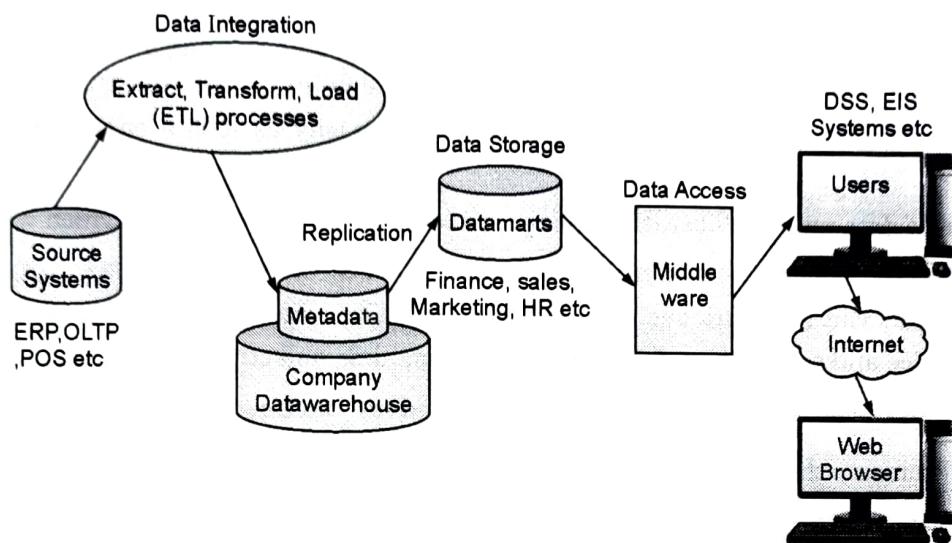


Fig. 2.5.2 : Architecture of a Data warehouse

As shown in the Fig. 2.5.2, let us have a look at the components and their functions in a Data warehouse architecture.

- **Data Extraction :** In this step, data is collected from various heterogenous sources both external and internal to the organization.
- **Data Cleaning :** In data cleaning ,the errors in data are found and simultaneously corrected.
- **Data Transformation :** In this step data is converted into a format to be stored in Data warehouses, which was previously stored in a format suitable for legacy systems.
- **Data Loading :** Data Loading involves various functions performed such as sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing :** This step deals with updating data from data sources to Data warehouse.

There are different types of datawarehouse applications such as

- **Information Processing :** Various data processing functions like querying, statistical analysis, reporting using tables, charts, or graphs is supported by Data warehouses.
- **Analytical Processing :** The data within warehouses can be analyzed by means of basic OLAP operations, such as slice-and-dice, drill down, drill up, and pivoting.
- **Data Mining :** Data mining can also be performed on Data warehouses. It supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction.

2.5.3 Benefits and Drawbacks of Data Warehousing

The benefits of Data warehouses are listed below

- End users can obtain a consolidated view of organizational data.
- The data for analysis can be accessed more faster and easily as it is accumulated at a central location.
- Broad range of data analysis operations are possible with Data warehouses, which was previously not possible.
- These operations can help improve business knowledge, provide competitive advantage, enhance customer service and satisfaction, facilitate decision making, and streamline business processes.

Inspite of their benefits, Data warehouses have few drawbacks

- First and foremost, they are very expensive to build and to maintain hence only very large enterprises where huge amounts of data are generated can afford it.
- Secondly, integrating data from obsolete mainframe systems can be very difficult and expensive.
- Thirdly, different departments within an enterprise might be hesitant to share their data with other departments.

2.6 KNOWLEDGE MANAGEMENT SYSTEMS

2.6.1 Concept

UQ. Describe what is meant by knowledge management. What factors have led to its development.

(MU - Q. 2(A), Jan. 21, 5 Marks)

- **Knowledge Management (KM)** is a process that helps organizations manipulate important knowledge that comprises part of the organization's memory, usually in an unstructured format.
- For an organization to be successful, knowledge must be effectively disseminated among people of different departments. Also knowledge must keep growing for the organization to grow. It should not remain stagnant.
- Knowledge is different from data and information. As explained earlier data are a collection of facts, measurements, and statistics; information is organized or processed data that is timely and accurate.
- Knowledge is information that is contextual, relevant, and useful. Knowledge is also termed as Intellectual asset.
- To understand the concept of knowledge let us take an example. The list of percentage marks is data. When these marks are associated with a particular student (mapped with his roll no. and name), it becomes information. Now when these percentages are used to decide a cutoff on how many students can go for campus placement, say all students scoring above 60% marks are eligible for interview then it becomes knowledge.

- **Knowledge Management Systems** aim to help an organization make the most optimum use of the knowledge that has been aggregated.
- **Knowledge Management Systems (KMSs)** deal with the use of modern information technologies - the Internet, intranets, extranets, databases - to systematize, enhance, and accelerate intrafirm and interfirm knowledge management.

2.6.2 Factors Leading to Development of knowledge Management Systems

Below are few factors that are leading to increased use of knowledge management systems:

- KMSs are designed to help an organization manage rapid changes and turnover by using the information assets available.
- The best practices and methods to improve the overall performance of the business organization is readily available to the employees. For example, sales managers can now make available their knowledge about how to best handle their customers. The organization can then utilize this knowledge when it trains new customer sales representatives.
- Due to the knowledge gathered organizations can improve their customer service drastically.
- At the same time products can be developed in more efficient manner satisfying customer requirements even better.
- Also improved employee retention is motivating increased use of knowledge management systems. The automated business processes are leading to higher employee satisfaction.
- Knowledge management systems encourage the employees by rewarding those who have added their skills to enhance the knowledge base.
- Thus, Knowledge Management Systems help in making most productive use of accumulated knowledge.
- Now let us have a look at the Knowledge Management System(KMS) Cycle.

2.6.3 The KMS Cycle

- The KMS cycle consists of six steps as shown in Fig. 2.6.3. The system is cyclic because the environment and the knowledge is ever changing and growing. It is continuously refined.

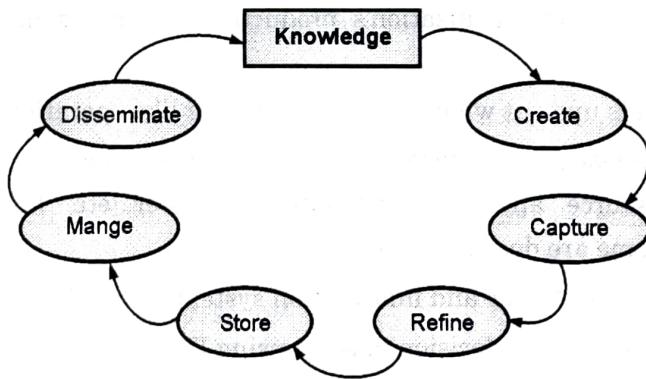


Fig. 2.6.3 : Knowledge Management Systems

- **Create** : Create involves knowledge creation when people discover new things as the data is continuously accumulated from various internal and external sources.
- **Capture** : New knowledge must be identified and gathered and represented in a standard way.
- **Refine** : The new knowledge captured must be refined and brought in a form that is useful and can be put into action.
- **Store** : The useful knowledge captured must then be stored in a proper format in a knowledge repository so that people in the organization can access it.
- **Manage** : The knowledge in the repository must be continuously managed and updated. It must always have current knowledge.
- **Disseminate** : Knowledge must be spread and made available in a useful format to anyone in the organization who needs it, anywhere and anytime.

Some examples of knowledge management systems are:

- The company employees might be working on some project. So, all the files and information related to the project is shared amongst the team members. This will help every member to be updated what the other one is doing.
- The customer and employee feedback can be stored at a single location and can be shared with people who are authenticated to access it. Based on the knowledge shared, necessary actions can be taken.
- A company's research data on developing new products and services can be shared among the employees of research and development team, so that healthy discussions and appropriate decisions for progress of the business can be taken.

2.7 MANAGERS AND DECISION MAKING

UQ. Discuss the Impact of BI on Decision Making.

(MU - Q. 3(B), Dec. 19, 10 Marks)

- **Management** is a process by which an organization achieves its goals through the use of resources (people, money, materials, and information). Managers must strive for optimizing this process. They must try to improve the organization's productivity by using minimum resources to get maximum output.
- The manager's job depends upon at what level he is working like operational, tactical or strategic.
- The manager works as a leader, supervisor, middleman between lower level employees and upper level management, resource allocator, negotiator, analyzer etc. Some of these roles are interpersonal whereas some are decision making.
- The managers take the support of IT and information systems to perform these roles.
- Here, in this section our focus is on decision making using IT.
- A **decision** refers to a choice among two or more alternatives that individuals and groups make.



- Economist Herbert Simon (1977) described decision making as composed of three major phases: *intelligence, design, and choice*. Once the choice is made, the decision is implemented.
- The decision-making process begins with the *intelligence phase*, in which managers scrutinize a situation and identify the problem or opportunity.
- In the *design phase*, decision makers construct a model for the situation. They make assumptions, identify relations between various parameters. The model is validated with test data.
- Finally in the *choice phase* a decision is made i.e. selecting the best solution to solve the problem.
- Computer-based decision support systems assists managers in the decision making process.
- Managers alone cannot take up the decisions. They need IT support to take up decisions because:
 - o The number of alternative solutions is very vast due to innovations in technology, global markets, use of internet to do business etc. Therefore selecting the best from these various options is a great challenge. For this IT support is needed to assist in making optimal searches and comparisons.
 - o The ever increasing information processing to take up decision should be done very fast which is not possible manually but only with the support of IT tools.
 - o Due to uncertainty in the decision environment, taking up decisions is becoming quiet complex. It is usually necessary to conduct a sophisticated analysis in order to make a good decision.
 - o The decision making process should be done rapidly and inexpensively ,which is again a major challenge.
- Here comes into picture the concept of Business Intelligence.

2.8 BUSINESS INTELLIGENCE (BI)

2.8.1 Concept

- To provide users with access to corporate data, many organizations are implementing data warehouses and data marts, which we have seen in the earlier sections. Users analyze the data in warehouses and marts using a wide variety of BI tools.
- Many vendors offer bundled packages of various tools by the name *Business Intelligence* (BI) software.
- Major BI software vendors include SAS, Hyperion, Business Objects, Information Builders, SPSS, and Cognos.
 - o BI is vital to modern decision making and organizational performance. The term *Business Intelligence* was introduced by Business and IT analyst Howard Dresner in 1989 while he was an analyst at Gartner, a market research firm.
 - o BI covers all decision-support applications. It includes both “getting data in” (to a data mart or warehouse) and “getting data out” (through BI applications).

- Previously BI was used by only managers for decision support but these days all other stakeholders like customers, customer support executives, suppliers, etc. are using BI to achieve current updated information.
- BI tools can be as simple as using Excel sheets for analysis by smaller organizations or sophisticated data mining, predictive analysis and data visualization tools for bigger organizations. BI tools are used within organizations for making decisions on market campaigns, sponsorships and fundings needed, what could be the impact and benefit of these campaigns to the organization etc.
- BI is used to fundamentally transform the ways in which a company competes in the marketplace. BI supports a new business model, and it enables the business strategy.
- Just consider an example of an entertainment company named Sarah's amusement that achieved its business target using BI.
- Sarah's amusement developed a customer loyalty program, known as *Total Benefits*. To implement the program, Sarah's created a data warehouse that collected data from nearby hotels, bars and casinos about their customers through different access points like internet, slot machines etc..
- Sarah's used this data to reward loyal customers and reach out to them in personal and attractive ways, such as through promotional offers. As a result, the company became a leader in the gaming industry.
- It used analytics to predict customer behavior.

Until now we have seen the basics of data warehouses and data marts. We have also seen the use of data warehouses and data marts in BI. Now, we will go ahead with how the user community can analyze the data in data warehouses and data marts, how the results of these analyses are presented to users, and how organizations can use the results of these analyses.

2.8.2 Business Intelligence Applications for Data Analysis

- Firstly we need to understand how users use BI to analyze the data, how they present the result of analyses and how users in the form of managers and executives implement these results.
- A variety of BI applications for analyzing data are available. They include multidimensional analysis (also called *online analytical processing*, or *OLAP*), data mining, and decision support systems. Now let us have a look at these applications.

Multidimensional Analysis or Online Analytical Processing (OLAP)

- Certain BI applications include online analytical processing (OLAP), also known as multidimensional analysis capabilities.
- OLAP involves operations such as “slicing and dicing” data stored in a multi dimensional format, drilling down in the data to greater detail, and aggregating the data.

- Let us have a look at the multi dimensional data cube as shown in Fig. 2.8.2 which represents the clothing product on the x-axis, geographical location on the y-axis, and year on the z-axis.

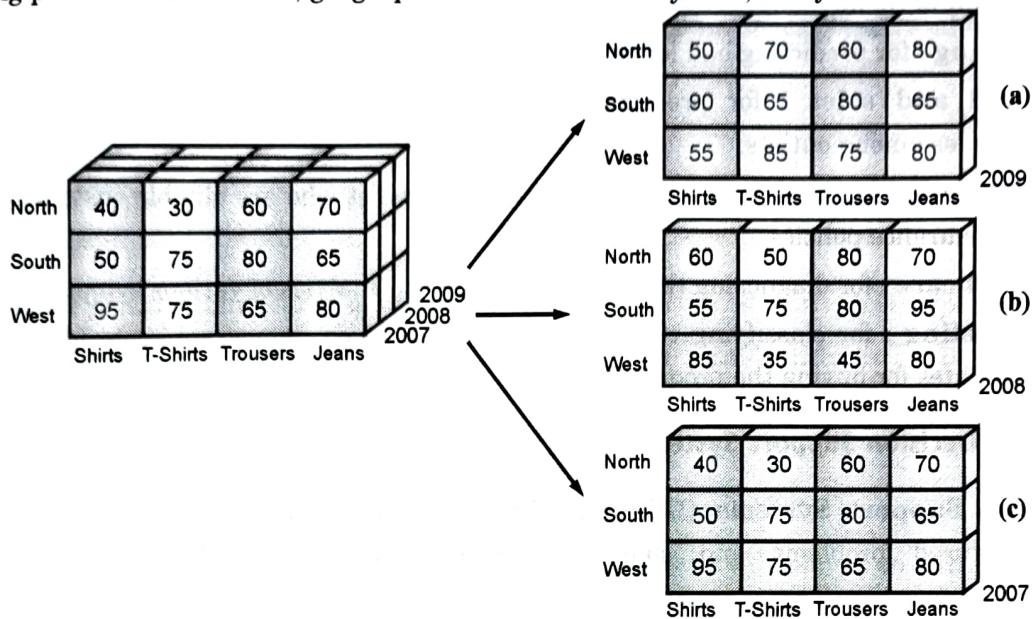


Fig. 2.8.2 : Representation of a multi dimensional cube

- Just imagine you want to know how many tshirts were sold by the company in the North region in the year 2008. A slice and dice operation will be performed on the cube, using *tshirts* as the specific measure for clothing product, *North* as the measure for geographical location, and year as 2008.
- Further, if you wanted to find how many tshirts were sold during 2007-2009, aggregation function, also called as “rollup” would be used.

2.8.3 Data Mining

- **Data mining** refers to the process of searching for valuable business information in a large database, data warehouse, or data mart.
- Two basic operations are performed in data mining :
 - Predicting trends and behaviours** which means finding predictive information in large databases. Say, for example a bank loan manager can perform analysis of certain parameters like past credit history of customer, assets, income level etc. and based on that make a decision whether to sanction the loan or not.
 - Identifying previously unknown patterns** : Like it is seen that in departmental stores usually people buy products in a particular pattern. If they purchase bread, they tend to buy butter and jam. So, data mining tools can identify such patterns and help make informed decisions like whether to give discounts when customer buys these products together. Another example could be of credit card usage patterns. If your card is stolen and used fraudulently, the usage pattern will surely vary visibly from your regular pattern. Data mining tools can distinguish this difference and bring the issue to your notice.

- Data mining has been used in several industries to achieve a competitive advantage. Let us have a look at a few of them.
 - o **Banking** : for predicting bad loans and which customers are eligible for issue of loan.
 - o **Retail and sales** : for predicting sales, determining inventory levels and distribution schedules among outlets.
 - o **Insurance** : for forecasting claim amounts and predict who are eligible customers for buying new insurance policies.
 - o **Healthcare** : for diagnosing the illness based on the analysis of various symptoms.
 - o **Marketing** : for classifying customer segment based on demographics to identify proficient candidates for buying the products.

2.8.4 Decision Support Systems

- **Decision Support Systems (DSSs)** combine models and data with the intent to analyze semistructured problems and some unstructured problems that involve extensive user involvement.
- **Models** are simplified representations, or abstractions, of reality. DSSs allow data analysts to access data, manipulate it and perform analysis on it. DSSs perform sensitivity analysis, what-if analysis, and goal-seeking analysis.
- Let us see what is each one of these.
 - o **Sensitivity Analysis** : *Sensitivity analysis* is the study of the impact that changes in one or more input parameters have on the output parameters.
 - o There are two types of input variables: decision variables and environmental variables. "What is our threshold for reordering raw materials?" is a decision variable. "What is the GST percentage?" is an environmental variable. Based on the analysis of the input variables, output variable is determined. As in this case the total cost of raw materials is the output desired. Sensitivity analysis is extremely important because it enables the system to adapt to changing environmental conditions and to the varying requirements of different decision-making situations.
 - o **What-if Analysis** : *What-if analysis* helps in predicting the impact of change in one or more input variables on the proposed solution. For example, what will happen to the total inventory cost if the originally assumed cost of carrying inventories is 15 percent rather than 12 percent? BI systems help data managers to ask such questions to the DSSs and get the responses appropriately.
 - o **Goal-Seeking Analysis** : *Goal-seeking analysis* tries to calculate the value of the inputs necessary to achieve a desired level of output. Say, for example the BI analysis initially predicted a profit of Rs.1 crore for the company. The manager might want to know how much sales quantity should be increased to get a profit of Rs.2 crores. For achieving this, various parameters need to be adjusted like increase in funding, lower product costs, enhance advertising efforts etc.

2.8.5 Business Intelligence Applications for Presenting Results

Dashboards and Data Visualization techniques are used for presenting the results of the various types of analysis performed.

Dashboard

- They are designed for use by top level executives, employees, business partners and even customers.
- A dashboard provides simple way to access timely information and managerial reports.
- It is user friendly, it has graphical interfaces, and, most significantly, it enables managers to examine exception reports and drill down into detailed data. This ability helps the executive to go deep down into levels of details.
- Helps to access real time data for the most critical factors responsible for success of business.
- Reports showing trend analysis and focusing on deviations can be displayed on dashboards.
- One wonderful example of a dashboard is the Bloomberg Terminal. Bloomberg LP a privately held company, provides a subscription service that sells financial data, software to analyze this data and trading tools.

Data Visualization Technologies

- After the data has been analyzed, it can be presented in different visual formats such as text, tables, charts, graphics etc.
- This process, known as *data visualization*, makes IT applications more attractive and understandable to users.
- A lot of data visualization softwares that support decision making are available.
- Two well-known ones are **geographic information systems** and **reality mining**.
 - o **Geographic Information Systems** : A geographic information system (GIS) is a computer based system for capturing, integrating, manipulating, and displaying data using digitized maps. Its most unique feature is that every record or digital object has an identified geographical location. This process, called *geocoding*, enables users to generate information for planning, problem solving, and decision making. In addition, the graphical format makes it easy for managers to visualize the data.
 - o **Reality Mining** : It is an integration of GISs and global positioning systems. **Reality mining** lets analysts to extract information from the usage patterns of mobile phones and other wireless devices. With this information they are able to provide a more accurate picture of what people do, where they go, and with whom they communicate.

► 2.9 MULTIPLE CHOICE QUESTIONS

- Q.1** The term Field in a data represents _____. (Jan 2021, 2 Marks)
- (a) Integrated collection of logically related data
 - (b) A group of related records
 - (c) Logical structure
 - (d) Data attribute ✓Ans. : (d)
- Q.2** Functions of a DBMS includes _____. (Jan. 2021, 2 Marks)
- (a) Database
 - (b) Datamart
 - (c) Data Warehouse
 - (d) Manipulation of records in a table ✓Ans. : (d)
- Q.3** Data Mart is a subset of _____. (Jan 2021, 2 Marks)
- (a) Data
 - (b) Data mining
 - (c) Data Warehouse
 - (d) Database ✓Ans. : (c)
- Q.4** Data mining is not used for _____. (Jan 2021, 2 Marks)
- (a) Day to Day operations
 - (b) Market analysis
 - (c) Customer retention
 - (d) Discover new correlations ✓Ans. : (a)
- Q.5** Data scrubbing is which of the following ? (Jan 2021, 2 Marks)
- (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 ✓Ans. : (c)
- Q.6** The data is nothing but _____ that are generated during day to day transactions. ✓Ans. : (a)
- (a) facts and figures
 - (b) fields
 - (c) attributes
 - (d) records
- Q.7** Data represented in a meaningful context is termed as ✓Ans. : (b)
- (a) Record
 - (b) Information
 - (c) Attribute
 - (d) Field
- Q.8** Which of the following is a barrier in managing data ✓Ans. : (c)
- (a) data is stored on homogenous computing environments
 - (b) data is simple facts and figures
 - (c) data is stored on heterogenous computing environments
 - (d) None of above
- Q.9** _____ can be defined as a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals ✓Ans. : (a)
- (a) Data governance
 - (b) Data cleansing
 - (c) Data management
 - (d) Data mining

- Q.33** In this step data is converted into a format to be stored in Data warehouses, which was previously stored in a format suitable for legacy systems.
- (a) Data transformation (b) Data cleansing
 (c) Data extraction (d) Data integration
- ✓Ans. : (a)
- Q.34** Data Refreshing involves
- (a) involves functions such as sorting, summarizing, consolidatingetc.
 (b) involves updating data from data sources to datawarehouse
 (c) involves conversion of data formats
 (d) None of above
- ✓Ans. : (b)
- Q.35** Which of these is not a benefit of datawarehousing
- (a) non volatile (b) faster analysis
 (c) competitive advantage (d) expensive
- ✓Ans. : (d)
- Q.36** _____ is a process that helps organizations manipulate important knowledge that comprises part of the organization's memory, usually in an unstructured format.
- (a) Knowledge management (b) Data management
 (c) Information management (d) All of above
- ✓Ans. : (a)
- Q.37** Knowledge is information that is
- (a) contextual (b) relevant (c) useful. (d) all of above
- ✓Ans. : (d)
- Q.38** _____ deal with the use of modern information technologies to systematize, enhance, and expedite intrafirm and interfirm knowledge management.
- (a) Knowledge management systems (b) Data Management Systems
 (c) Information Management Systems (d) Database Management Systems
- ✓Ans. : (a)
- Q.39** Basic steps in KMS cycle are
- (a) identify-gather-modify-manage-distribute
 (b) create-capture-refine-store-manage-disseminate
 (c) create-capture-store-manage-disseminate
 (d) identify-gather-refine-modify-manage-distribute
- ✓Ans. : (b)
- Q.40** What are the different roles played by a manager?
- (a) negotiator (b) leader (c) analyst (d) all of above
- ✓Ans. : (d)
- Q.41** Decision making process is composed of three major phases. They are:
- (a) intelligence, design and choice (b) intelligence, analysis and design
 (c) intelligence, analysis and choice (d) intellect, design and choice
- ✓Ans. : (a)
- Q.42** _____ refers to the process of searching for valuable business information in a large database, data warehouse, or data mart.
- (a) Data mining (b) Data hunting
 (c) Data retrieval (d) Data searching
- ✓Ans. : (a)

Q.43 Data mining can perform operations such as

- (a) identifying previously unknown patterns.
- (b) predicting trends and behaviours
- (c) All of above
- (d) None of above

✓Ans. : (c)

Q.44 Decision support systems (DSSs) perform various types of analyses such as

- (a) Sensitivity analysis
- (b) What-if analysis
- (c) Goal seeking analysis
- (d) All of above

✓Ans. : (d)

Q.45 _____ is the study of the impact that changes in one or more parameters have on the output parameters

- (a) Sensitivity Analysis
- (b) What-if analysis
- (c) Goal seeking analysis
- (d) None of above

✓Ans. : (a)

Q.46 _____ tries to calculate the value of the inputs necessary to achieve a desired level of output.

- (a) Sensitivity Analysis
- (b) What-if analysis
- (c) Goal seeking analysis
- (d) All of above

✓Ans. : (c)

Q.47 _____ are used for presenting the results of the various types of analysis performed.

- (a) Dashboards
- (b) Widgets
- (c) Graphical Interfaces
- (d) None of above

✓Ans. : (a)

Q.48 Data visualization techniques involve

- (a) representing the impact of change in one or more input variables on the proposed solution.
- (b) representing data in visual formats such as text, tables, charts, graphics after analysis.
- (c) representing the impact that changes in one or more input parameters have on the output parameters.
- (d) discovery and representation of new correlations.

✓Ans. : (b)

Q.49 A _____ is a computer based system for capturing, integrating, manipulating, and displaying data using digitized maps.

- (a) geographic information system
- (b) global positioning system
- (c) geographic positioning system
- (d) global information system

✓Ans. : (a)

Q.50 _____ is an integration of geographic information systems and global positioning systems.

- (a) Data mining
- (b) Reality Mining
- (c) Datawarehousing
- (d) Satellite mining

✓Ans. : (b)