By Prof.Harsha Devendra Zope

02	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 analysisand Presenting Results

DATA

- What does the number 29061997 mean?

- Is it
- A bank account number?
- A birthday?(29th June 1997)
- A telephone number?
- Without processing or more information this data is meaningless

DATA	INFORMATION	KNOWLEDGE
29061997	A birthdate(29th June 1997)	16 YEARS OLD

DATA

- ‘Data are the main raw material of Information systems. Data constitute valuable organizational resources that must be managed effectively to benefit all stakeholders in an organization. The increasing importance of data as a valuable organizational resource has brought a variety of changes in the organizations everywhere.
- Eg. Customer ID,First Name,Last Name,Contact no etc.

- Data that are generated as a result of various transactions are now stored, processed, and analyzed using software applications like Database Management System (DBMS).

- Define Data

•**Data** are raw facts, figures or observations, particularly about
physical activities or business transactions. For example, the sale of an automobile generates a lot of data describing the process. Data are identified by its types and attributes. For example: people, places, things, events, etc. they all are data.

DATA CAN TAKE MANY FORMS

- ALPHANUMERIC DATA (COMBINATION OF NUMBERS AND LETTERS)
- TEXT DATA (SENTENCES & PARAGRAPHS USED IN WRITTEN COMMUNICATION)
- IMAGE DATA (GRAPHICS, SHAPES, FIGURES ETC)
- AUDIO (HUMAN VOICE & OTHER SOUNDS)

Describe the importance of data in today's business environment.

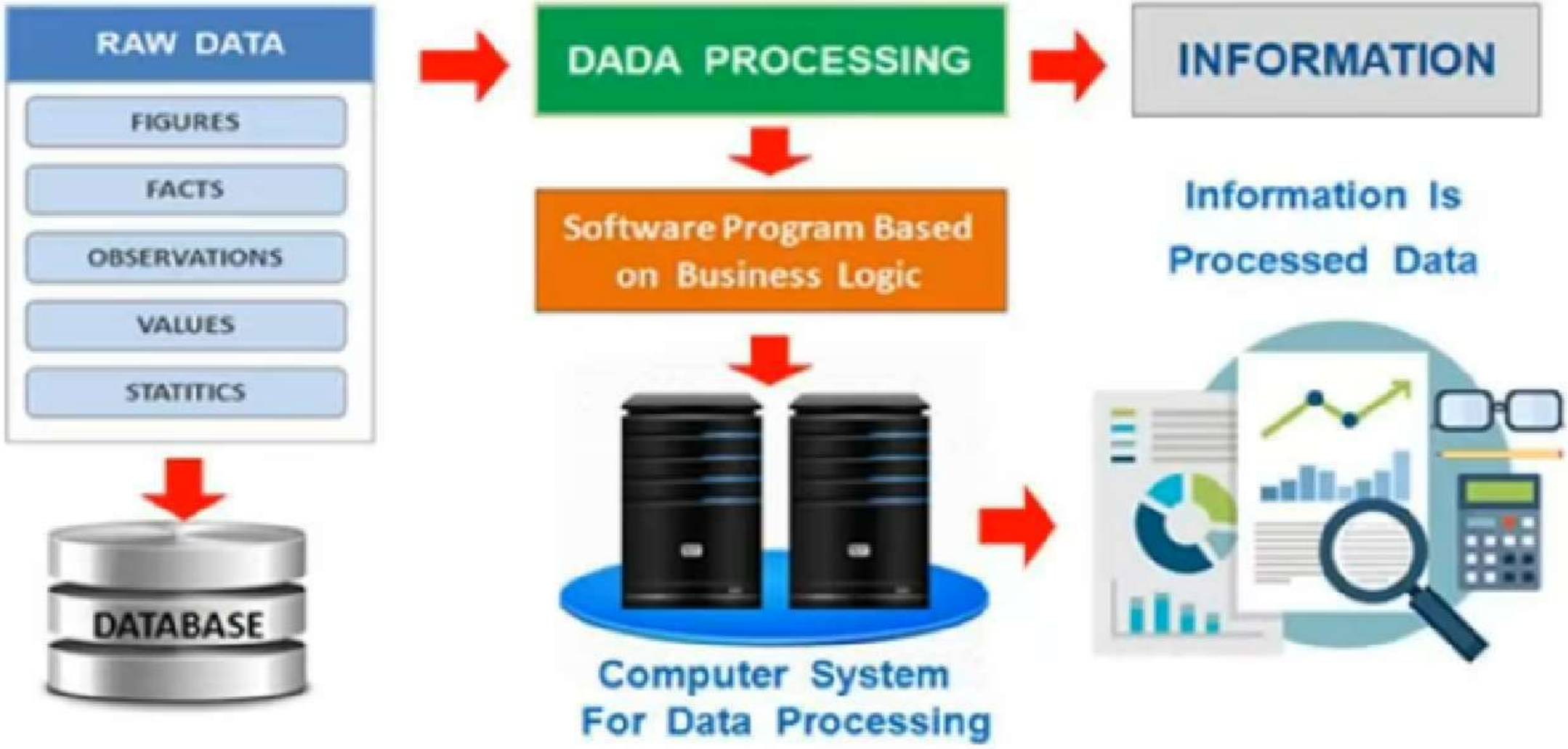
- Data are the main raw material of information systems. Data constitute valuable organizational resources that must be managed effectively to benefit all stakeholders in an organization. The increasing importance of data as a valuable organizational resource has brought a variety of changes in the organizations everywhere.

- Data that are generated as a result of various transactions are now stored, processed, and analysed using software applications like Database Management system (DBMS).

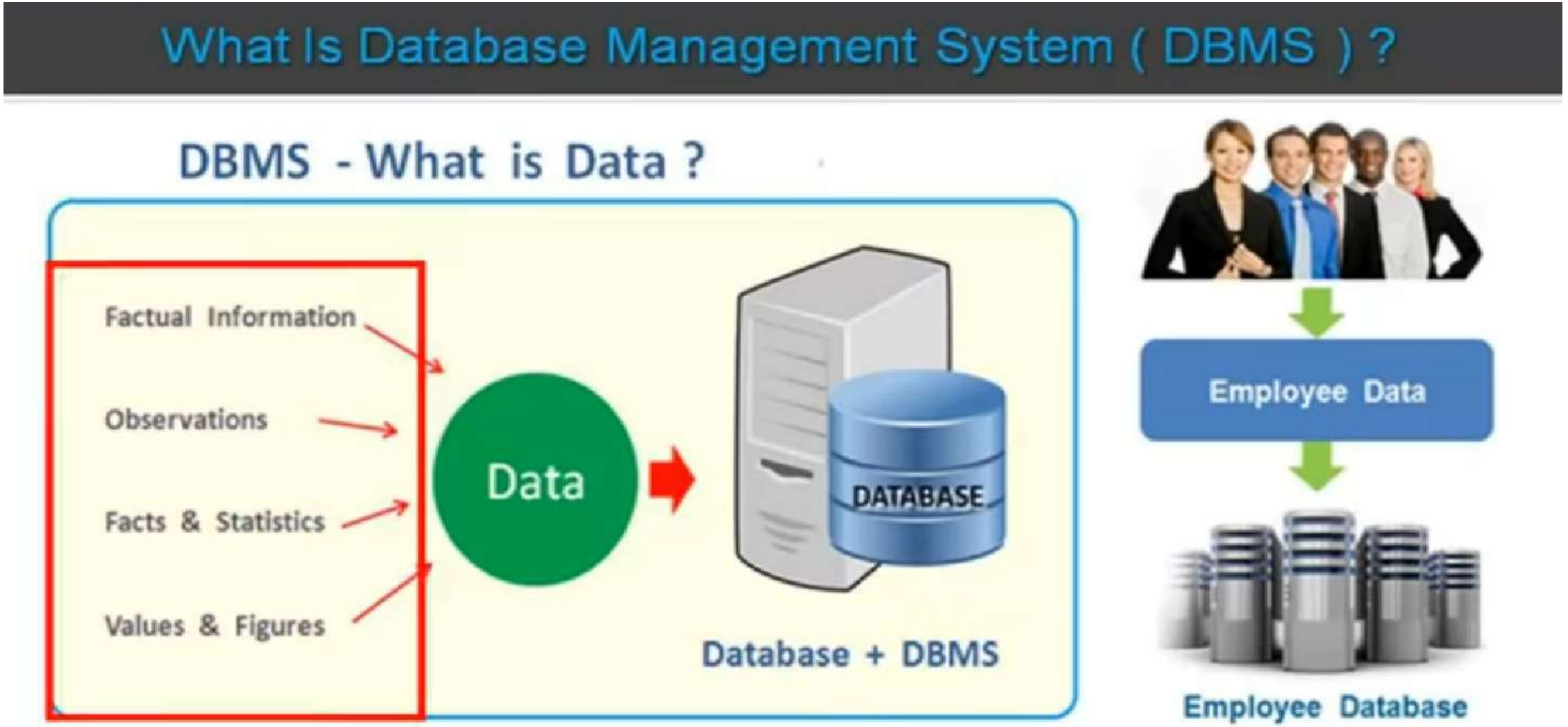
- They show relationships among various organizational entities such as sales, customers, competitors, and markets. In today's globalized world running on communication network, data are protected with the same energy as the cash in a bank vault.

- Data have become the lifeblood of today's organizations, and the effective and efficient management of data is considered an integral part of organizational strategy.

What Is Database Management System (DBMS) ?



What is Data?


What Is Database Management System (DBMS) ?
DBMS - What is Data ?
Factual Information
Observations
Facts & Statistics
Values & Figures
Data
DATABASE
Database + DBMS
Employee Data
Employee Database

What is meaning of Data management?

- **Organization need to practice data management.**
- **Data management** is a managerial activity that applies information systems technologies like database management system, data warehousing, data mining and other data management tools to manage an organization's data resource, so that they can meet the information needs of their businesses.

- Managing Data All IT applications require data. These data should be of high quality, meaning that they should be accurate, complete, timely, consistent, accessible, relevant, and concise. Unfortunately, the process of acquiring, keeping, and managing data is becoming increasingly

- The Difficulties of Managing Data Because data are processed in several stages and often in multiple locations, they are frequently subject to problems and difficulties. Managing data in organizations is diffi cult for many reasons. First, the amount of data increases exponentially with time. Much historical data must be kept for a long time, and new data are added rapidly

Why is data important for an organization?

- Data constitute valuable organizational resources. Data have become the lifeblood of today's organizations, and the effective and efficient management of data is considered an integral part of organizational strategy.

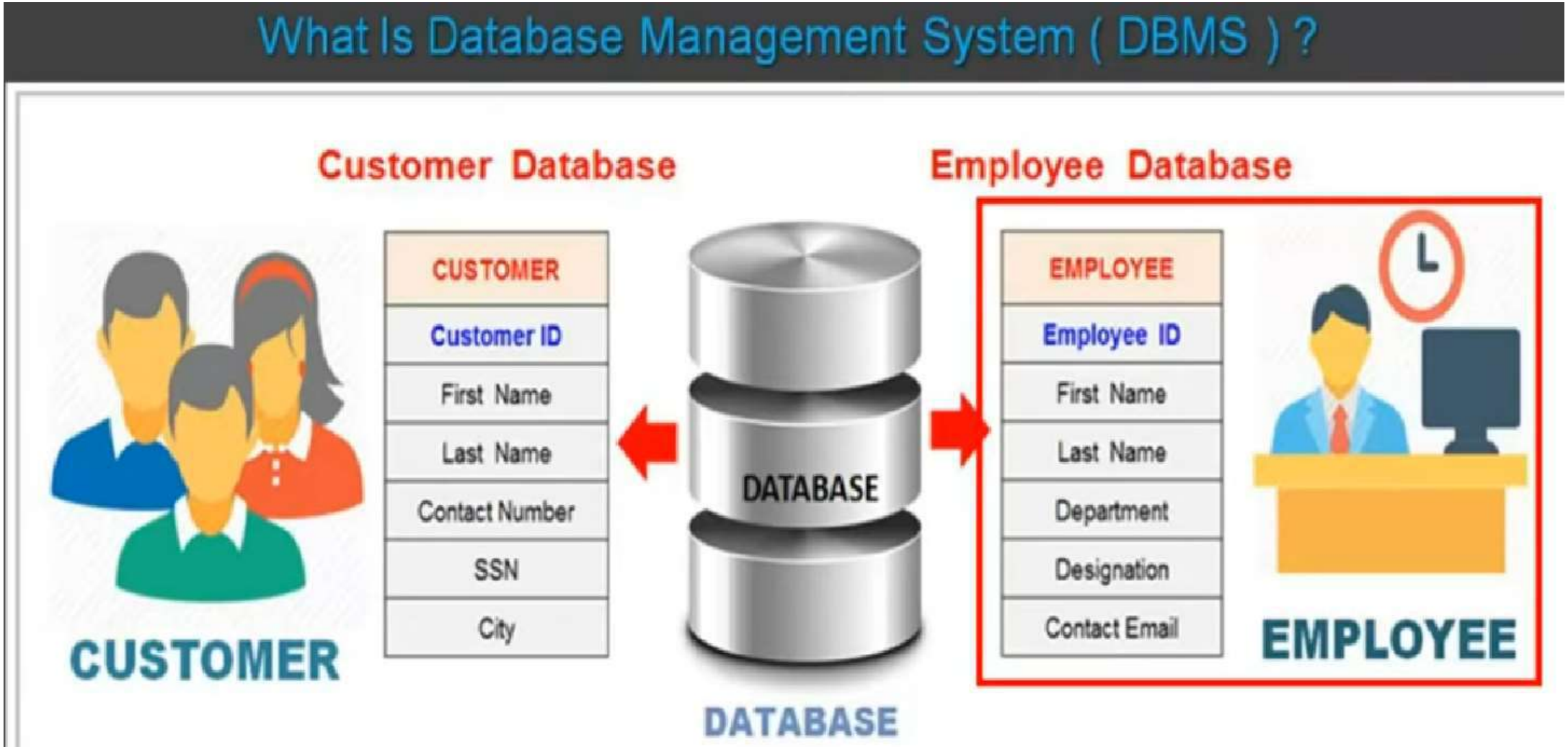
•What is the difference between data and information?

•People often use the terms data and information interchangeably.
•However, there is a difference. Data is a raw fact or figure that after processing results into information.
• Data by itself does not provide any understanding of the context in which it was recorded. In contrast, for information, both the context of the data and the motive of the person accessing the data become essential.

Describe database

A **database** is an integrated collection of logically related data elements. A database consolidates records previously stored in separate files into a common pool of data elements that provides data for many applications. The data stored in a database are independent of the application programs using them and of the type of storage devices on which they are stored. Thus, databases contain data elements describing entities and relationships among them.

What is database?



- Applications (ERP Application,WEB application,Dynamic Application,Website /Blog) uses database with the help of DBMS software.

Explain different type of databases in detail.

•Over the last several decades there has been continuous development
•in Information technology and its business applications which have resulted in the evolution of several major types of databases. Some of them are given below:

- **Operational databases** are used to store data which are needed in details to support different business processes and operations of a company. Some of the examples are a customer database, human resource database, Inventory database, and similar other databases that contain data generated by different business operations.

- **Distributed database** is another type of database which resides on network servers on the World Wide Web, on corporate intranets or extranets, or on other company networks.

- Distributed databases contain copies of operational or analytical databases or hypermedia databases, or any other type of database. Replication and distribution of databases improve database performance at users end.

- **Web-based database**: The rapid growth of websites on the internet, intranets and extranets has rapidly increased the demand and use of web-based databases. A website stores information in the form of a hypermedia database. Such a database is a collection of hyperlinked pages of multimedia-based information (text, graphic, images, video clips, audio segments, and so on)

What is Database Management System (DBMS)? Example

A **DBMS** is a set of software programs that controls the organization, storage, management, and retrieval of data in a database. DBMS are categorized according to their data structures or types. They are used to store, update and retrieve a database. The DBMS accepts requests for data from the application program and instructs the operating system to transfer the appropriate data. For example: MS-Access, Oracle, etc.

ORACLE

Functions of DBMS.

- **Creation** / **modifications** / **deletion** of tables, which may be physically located at one computer or different networked computers
- **Manipulation** of records in tables by sorting / filtering them
- **Searching** of desired records & **updating** / **deleting** them

What is a query language?

•A database **query language** is a part of DBMS which allows users
•to interactively interrogate the database, analyse its data and update it according to the users' privileges on data. It also controls the security of the database

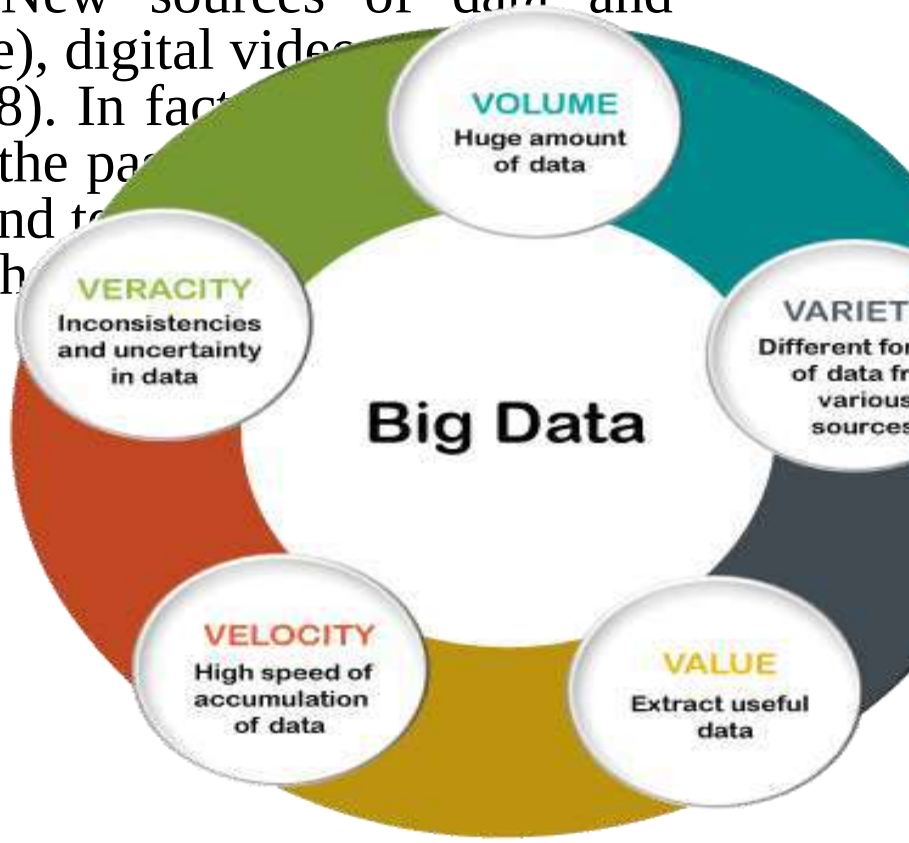
- A query language, also known as data query language or database query language (DQL), is a computer language used to make queries in databases and information systems. In database systems, query languages rely on strict theory to retrieve information. A well known example is the Structured Query Language (SQL).
-

Define data structure

A **data structure** is a way of storing data in a computer so that it can be used efficiently. It is an organization of mathematical and logical concepts of data. A well-designed data structure allows a variety of critical operations to be performed using few resources reducing both execution time and memory space.

Define Big Data and its basic characteristics.

- **Big Data** is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.
- We are accumulating data and information at an increasingly rapid pace from such diverse sources as company documents, e-mails, Web pages, credit card swipes, phone messages, stock trades, memos, address books, and radiology scans. New sources of data and information include blogs, podcasts, videocasts (think of YouTube), digital vide and RFID tags and other wireless sensors (discussed in Chapter 8). In fac capturing data about almost all events—including events that, in the pa think of as data at all, such as a person's location, the vibrations and t or the stress at numerous points on a bridge—and then analyzing th
- Big data can be described by the following characteristics:
  - Volume
  - Variety
  - Velocity
  - Variability



Define Big Data and its basic characteristics.

1. Volume:

- The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data.
- Hence, 'Volume' is one characteristic which needs to be considered while dealing with Big Data solution

2. Variety:

- Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.
- During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices,

• PDFs, audio, etc. are also being considered in the analysis applications.

- This variety of unstructured data poses certain issues for storage,mining and analysing data.

•3. Velocity:

• - The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.

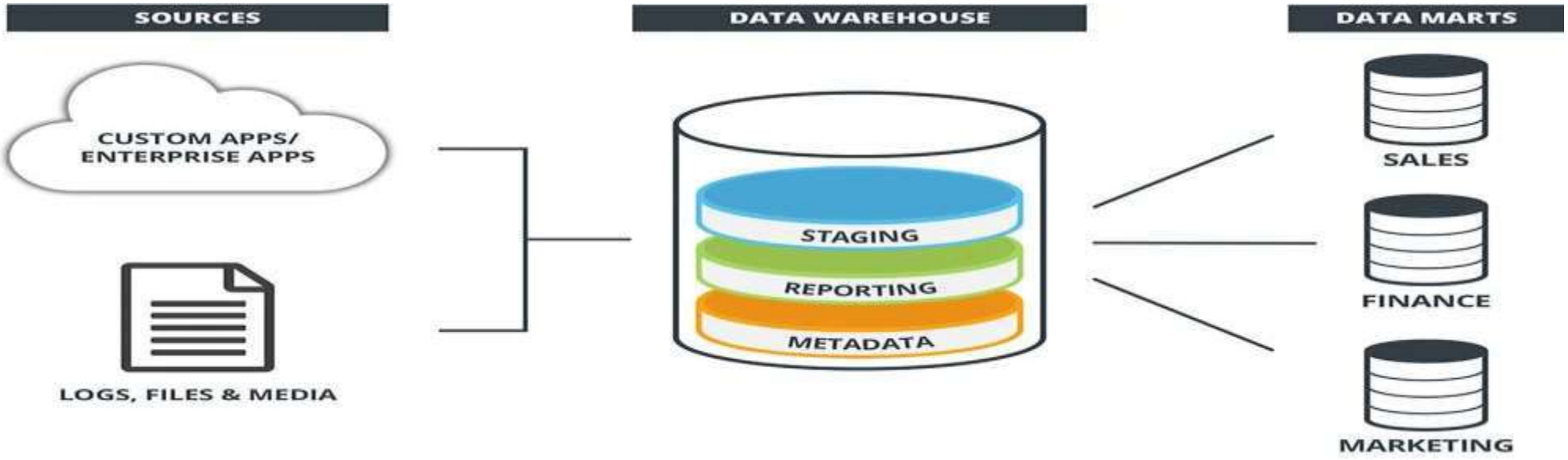
• - Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors, Mobile devices, etc. The flow of data is massive and continuous.

•4. Variability:

• - This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to

Data warehouse?

•**Data warehousing** is defined as a technique for collecting and managing data from various sources to provide meaningful business insights. It stores data that have been extracted from the various operational, external, and other databases of an organization.



- Today, the most successful companies are those that can respond quickly and flexibly to market changes and opportunities. A key to this response is the effective and efficient use of data and information by analysts and managers
- Let's look at an example. If the manager of a local bookstore wanted to know the profi t margin on used books at her store, she could obtain that information from her database, using SQL. . However, if she needed to know the trend in the profit margins on used books over the past 10 years, she would have to construct a very complicated SQL query.
- This example illustrates several reasons why organizations are building data warehouses and/or data marts. First, the bookstore's databases contain the necessary information to answer the manager's query, but this information is not organized in a way that makes it easy for her to fi nd what she needs. Second, the organization's databases are designed to process millions of transactions every day. Therefore, complicated queries might take a long time to answer, and they also might degrade the performance of the databases. Third, transactional databases are designed to be updated. This update process requires extra processing. Data warehouses and data marts are read-only, and the extra processing is eliminated because data already contained in the data warehouse are not updated. Fourth, transactional databases are designed to access a single record at a time. Data warehouses are designed to access large groups of related records. As a result of these problems, companies are using a variety of tools with data warehouses and data marts to make it easier and faster for users to access, analyze, and query data.

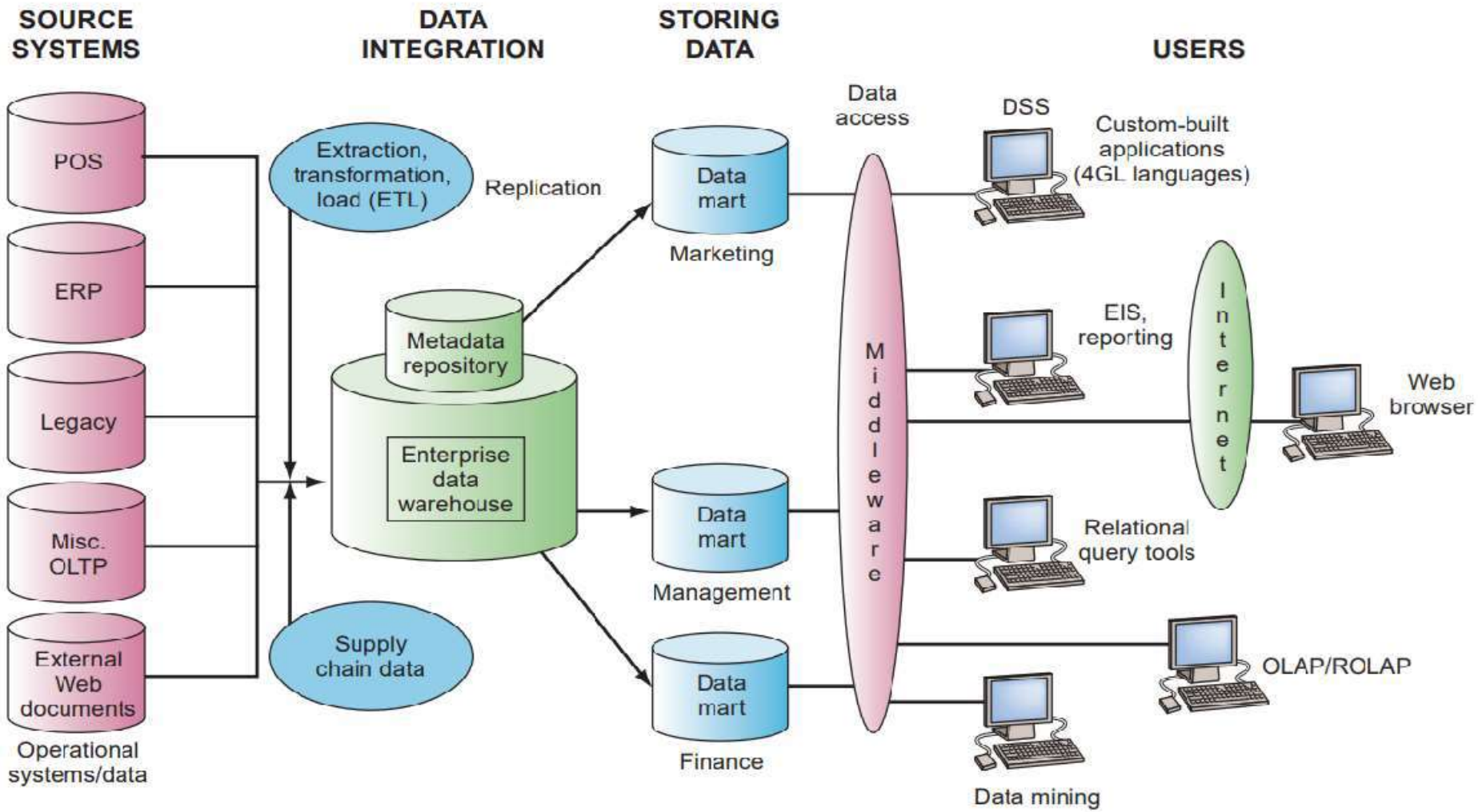


FIGURE 5.9 Data warehouse

A Generic Data Warehouse Environment

- A Generic Data Warehouse Environment
- The environment for data warehouses and marts includes the following: •
- Source systems that provide data to the warehouse or mart
- • Data-integration technology and processes that prepare the data for use
- • Different architectures for storing data in an organization's data warehouse or data marts
- • Different tools and applications for the variety of users.
- • Metadata, data-quality, and governance processes that ensure that the warehouse or mart meets its purposes

- Source Systems.
- Modern organizations can select from a variety of source systems: operational/transactional systems, enterprise resource planning (ERP) systems, Web site data, third-party data (e.g., customer demographic data), and more. The trend is to include more types of data (e.g., sensing data from RFID tags). These source systems often use different software packages (e.g., IBM, Oracle) and store data in different formats (e.g., relational, hierarchical). A common source for the data in data warehouses is the company's operational databases
- Data Integration.
- In addition to storing data in their source systems, organizations need to extract the data, transform them, and then load them into a data mart or warehouse. This process is often called ETL, but the term data integration is increasingly being used to refl ect the growing number of ways that source system data can be handled. For example, in some cases, data are extracted, loaded into a mart or warehouse, and then transformed

- Data extraction can be performed either by handwritten code (e.g., SQL queries) or by commercial data-integration software. Most companies employ commercial software. This software makes it relatively easy to specify the tables and attributes in the source systems that are to be used, map and schedule the movement of the data to the target (e.g., a data mart or warehouse), make the required transformations, and ultimately load the data.

- Data warehouse is a central source of the data that have been cleaned , transformed and catalogued so that they can be used by managers and other business professionals for mining online analytical processing and other forms of business analysis , market research and decision support.
- Data warehouse is core for business intelligence system which is built for data analysis and reporting.
- How it workds:

A data warehouse works as a central depository where information arrives from one or more data sources and merges into comprehensive database.

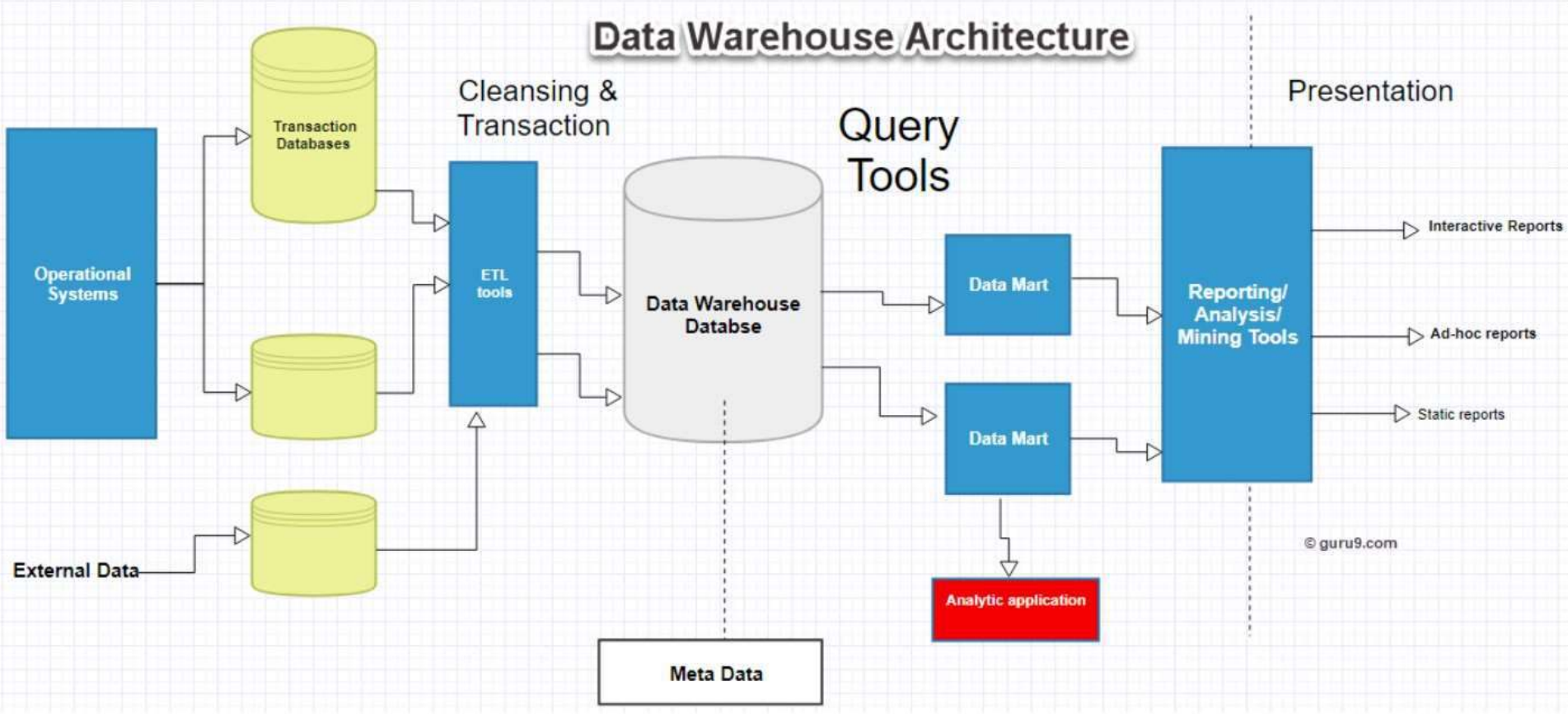
Data Mart

- It is a subset of the data warehouse designed specifically to cater to a particular line of business such as sales, finance, operations. Here data are directly collected from the sources. It focuses on the specific aspects of a business by offering specific Information.

Architecture of Data mart and Data warehouse in an organization

- A **Data warehouse** is an information system that contains
- historical and commutative data from single or multiple sources. Data Warehouse Concepts simplify the reporting and analysis process of organizations.
- **Data Warehouse Architecture** is complex as it's an information
- system that contains historical and commutative data from multiple sources. There are 3 approaches for constructing Data Warehouse layers: Single Tier, Two tier and Three tier.

Data Warehouse Architecture



Data Warehouse Architecture

- There are mainly 5 components of Data Warehouse Architecture: Database, ETL Tools, Meta Data, Query Tools and DataMart.
- The **central database** is the foundation of the data warehousing environment.
- The data sourcing, transformation, and migration tools are used for performing all the conversions, summarizations, and all the changes needed to transform data into a unified format in the data warehouse. They are also called Extract, Transform and Load (**ETL**) Tools.
- **Metadata** is data about data which defines the data warehouse. It is used for building, maintaining and managing the data warehouse.
- (you are sending package to friend,address to from is visible ,same for message sending)
- One of the primary objects of data warehousing is to provide information to businesses to make strategic decisions. **Query tools** allow users to interact with the data warehouse system.
- A **Data Mart** is focused on a single functional area of an organization and contains a subset of data stored in a Data Warehouse. A Data Mart is a condensed version of DataWarehouse and is designed for use by a specific department, unit or set of users in an organization. E.g., Marketing, Sales, HR or finance. It is often controlled by a single department in an organization.imagine

Application areas of data warehouse

•Following are the most common areas where data warehouses are

•extensively used these days:

- **Airline**: In the airline system, it Is used for operation purpose like crew assignment, analysis of route profitability, frequent flyer program promotions, etc.
- **Banking**: It is widely used in the banking sector to manage effectively the resources available on desk. Few banks also use it for the market research, performance analysis of the product and operations.

- Healthcare: Healthcare sector use dat warehouse to make strategies and predict outcomes,treatment reports,share reports
- Public sector:in public sector data warehouse is used for intelligent gathering,it help government agencies to maintain and analyse tax records,health policy records for every individual

Advantages of Data Warehouse

- Data warehouse allows business users to quickly access critical data from sources all in one place
- Data warehouses helps to integrate many sourses of data to reduce stress on the production system
- Data warehouse store large amount of historical data . This helps users to analyse different time periods and trends to make future prediction

Define knowledge. How is it different from information? What

Is its significance for a business firm?

- Knowledge is a set of information which provides capability to understand different situations, anticipate results and judge their effects, and offer solutions or clues to handle the situation.

- Data when organized in the right context becomes information. Information when shared, utilized and applied at the right time and right place to sort out problems becomes knowledge.

- Whether it is a person or an organization, knowledge plays a vital role in its growth and development. In today's challenging, tough and competitive business environment, having knowledge and the capability to use and manage it gives an edge. In a business, knowledge plays a key role today. Regardless of industry, all organizations today rely on the power of knowledge. They aim to build processes to store, grow and share knowledge to increase knowledge base of the overall workforce.

- Today businesses are also affected by the pace of the change. The change is rapid, innovative and path breaking. The product life cycles are shorter and organizations have to improve products and services to remain competitive.

- To manage this shift, knowledge is the key resource of the organization and workplace. Due to this critical importance of knowledge, business economy is today termed as knowledge economy.

Knowledge Management Systems

- The goal of knowledge management is to help an organization make the most productive use of the knowledge it has accumulated. Historically, management information systems have focused on capturing, storing, managing, and reporting explicit knowledge. Organizations now realize they need to integrate explicit and tacit knowledge into formal information systems. Knowledge management systems (KMSs) refer to the use of modern information technologies—the Internet, intranets, extranets, databases—to systematize, enhance, and expedite intrafi rm and interfi rm knowledge management. KMSs are intended to help an organization cope with turnover, rapid change, and downsizing by making the expertise of the organization's human capital widely accessible.
- Enhanced access to best-practice knowledge improves overall organizational performance. For example, account managers can now make available their tacit knowledge about how best to manage large accounts. The organization can then utilize this knowledge when it trains new account managers. Other benefi ts include improved customer service, more effi cient product development, and improved employee morale and retention.

Knowledge management cycle

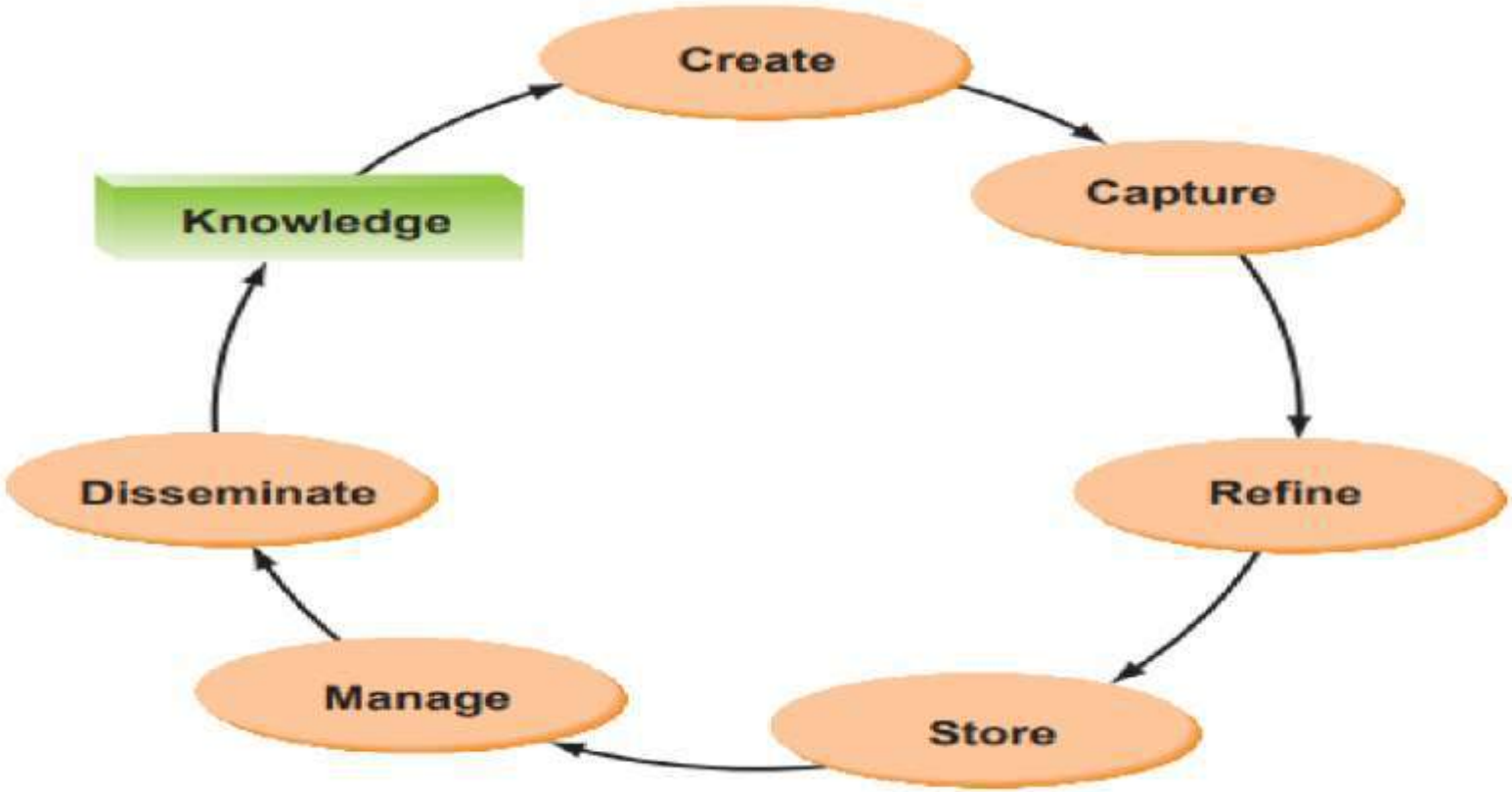


FIGURE 5.13 The knowledge management system cycle.

Knowledge management cycle

- A functioning KMS follows a cycle that consists of six step
- The reason the system is cyclical is that knowledge is dynamically refined over time.
- The cycle works as follows:
- 1. Create knowledge.
- Knowledge is created as people determine new ways of doing things or develop know-how. Sometimes external knowledge is brought in.
- 2. Capture knowledge. New knowledge must be identifi ed as valuable and be represented in a reasonable way.
- 3. Refine knowledge. New knowledge must be placed in context so that it is actionable. This is where tacit qualities (human insights) must be captured along with explicit facts.
- 4. Store knowledge. Useful knowledge must then be stored in a reasonable format in a knowledge repository so that others in the organization can access it.
- 5. Manage knowledge. Like a library, the knowledge must be kept current.
- It must be reviewed regularly to verify that it is relevant and accurate. 6. Disseminate knowledge. Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and anytime.

Types of knowledge

Types of Knowledge

- Explicit Knowledge: knowledge that is easy to write & share. For example knowledge found in books, databases, files etc.
- Implicit Knowledge: knowledge that is explains how best to implement explicit knowledge. For example Coworker coaching etc.
- Tacit Knowledge: Knowledge gained from personal experience, it is less easy to document and share with others. For example, leadership skills, body language etc.

What Is knowledge management? What factors have led to its development?

•Knowledge management is the planned and systematic management of knowledge-related activities, practices, programs and policies within an organization to create a big knowledge database and then share it with others and improve its contents and quality continuously.

•In today's tough and challenging business world, KM is a necessity due to various forces which drive KM. They are both external and internal.

•**External forces**

•The forces in external environment are the following:

- **Globalization of business**: With loosening of trade barriers and advanced Internet technologies, businesses today operate beyond the local and national boundaries. It has a bigger market to tap and more sources and resources to bank upon. But there is a stiff competition everywhere. Overcoming them is a pre-requisite for success.

- **Changing customers and their demands**: Customers today easily access information about product and services, and are now more knowledgeable to demand more value at least cost. They drive your business by demanding better quality, new features, quick response and delivery.

- **Innovative competitors**: Competition is no longer limited to quality and cost but extended to providing value added services using technologies and best practices. This puts a heavy demand on organizations to compete with the same force.

•**Internal forces:**

•The forces in external environment are the following:

- **Effectiveness**: Organization's effectiveness in handling operations, seizing opportunities play an important role in their success. Effectiveness is anticipating the change in market and environment requiring pro-active actions to deal with it.

- **Technological capability**: Businesses need technology to bring in efficiency and effectiveness. Businesses must operate through collaborative work,high end information management.All this put together defines technology capability.
- - **Effectiveness of human resource**: People and organization behaviour affects effectiveness of the business enterprise. Knowledge about its human resource in terms of understanding their mental models and associations which affect them and the decision making is essential. KM initiative is the result of this requirement.

challenges faced by KM

- **Creating a flexible culture and collaboration**: This is one of the most significant challenges of KM. Organizations struggle to implement new policies, because people by nature tend to resist change. Employees generally want to protect their skills and knowledge, or they are reluctant to learn from their peers.

- **Security**: This is another challenge as it is important not just to protect sensitive information but also the intellectual capital. KM is based on sharing and storing of knowledge by individual employees.

- **Measuring knowledge**: At times it is difficult to define a system to measure the knowledge within your organization, especially for tacit knowledge that cannot be easily quantified.

- **Document storage and management**: Knowledge will have to be stored and organized in some form which is never easy. Document management is a challenge for many companies. Documents must be well organized otherwise; it will be impossible to locate and use the knowledge you have stored.

- **Disseminating knowledge**: You'll need to devise a process where once you store knowledge other team members can access it. This is complicated both theoretically and tactically. So many organizations opt for a software system designed specifically for this purpose.

Business Intelligence

What Is Business Intelligence?

The process of collecting, analyzing, and transforming raw data into actionable insights for business purposes is known as Business Intelligence.