

## **Chap-03, Organizing Data & Information**

### **Chapter Overview**

Chapter 3 gives the business student insight into how an organization uses data and transforms it into meaningful information. Most businesses suffer from too much data rather than not enough.

A database helps extract significance from these numbers. Traditionally, data was maintained in separate files, which were created by separate applications. The database approach to data management offers many advantages over the traditional approach to data management, such as reduced data dependency, easier update and modification, reduced redundancy, standardization of data access, and more efficient program development. Use of the database approach has required computer specialists to give extra care to content and access, logical structure, and physical organization of data. The result can be a well-designed and well-managed database that is an extremely valuable tool in supporting decision making. Most frequently, this is done through the use of a database management system (DBMS).

#### **Data Management: The Hierarchy of Data**

- **Bit** (a binary digit): a circuit that is either on or off
- **Byte**: 8 bits
- **Character**: each byte represents a character; the basic building block of information
- **Field**: name, number, or characters that describe an aspect of a business object or activity
- **Record**: a collection of related data fields
- **File**: a collection of related records
- **Database**: a collection of integrated and related files

#### **Data Entities, Attributes, and Keys**

- **Entity**: a generalized class of people, places, or things (objects) for which data is collected, stored, and maintained. [ e.g.: Students]
- **Attribute**: a characteristic of an entity [Student No., Lname, Fname, Class, Address, Major, ...]
- **Data item**: a value of an attribute [0012345, Smith, Amanda, 2008, 34 Main St, FM, ...]
- **Key**: field(s) that identify a record [Student No., Class, Major, ...]
- **Primary key**: field(s) that **uniquely** identify a record [Student No.]

#### **The Traditional Approach Versus the Database Approach**

- **Traditional approach**: separate data files are created for each application Results in data redundancy (duplication) which conflicts with data integrity
- **Database approach**: pool of files with related data. Significant advantages over traditional approach.

#### **Data Modeling and the Relational Database Model**

- When building a database, consider:
  - **Content**: What data should be collected, at what cost?
  - **Access**: What data should be provided to which users, and when?
  - **Logical structure**: How should data be arranged to make sense to a given user?
  - **Physical organization**: Where should data be physically located?

#### **Data Modeling**

- Building a database requires two types of design
- **Logical design**
  - Shows an abstract model of how data should be structured and arranged to meet an organization's information needs
- **Physical design**

- Fine-tunes the logical database design for performance and cost considerations
- Data model: a diagram of data entities and their relationships
  - Entity-relationship (ER) diagrams: data models that use basic graphical symbols to show the organization of and relationships between data
- The Relational Database Model {contains 2 or more files with related fields}
  - all data elements are placed in two-dimensional tables (relations), which are the equivalent of files
- In the relational model:
  - Each row of a table is a data entity [ One student to a row]
  - Each column is an attribute [ separate column for Student No., Lname, Fname, Class, ...]
  - Domain: the allowable values for data attributes [e.g.: Abbreviations for Majors are specified ]

### **Manipulating Data**

- Selecting: select rows that meet the criteria, [eg: all rows that have 'FM' in the column for Major]
- Projecting: view only certain columns in a table [eg: May just want to view Student No. and Major]
- Joining: combines two or more tables.
- Linking: relates or links two or more tables using common data attributes

### **Database Management Systems (DBMS)**

- DBMS is the Interface between
  - Database and application programs
  - Database and the user
  - Database types
- Flat file – NOT a Database – one file that does not have fields that relate to another file.
- Flat Files & Databases can be Single user or a Multiple user.

### **Creating and Modifying the Database**

- Data definition language (DDL)
- Collection of instructions/commands that define and describe data and data relationships in a database
- Allows database creator to describe the data and the data relationships that are to be contained in the schema and the subschemas
- Data dictionary: a detailed description of all the data used in the database

### **Storing and Retrieving Data**

- When an application requests data from the DBMS, the application follows a logical access path
- When the DBMS goes to a storage device to retrieve the requested data, it follows a path to the physical location (physical access path) where the data is stored

### **Manipulating Data and Generating Reports**

- Query-By-Example (QBE): a visual approach to developing database queries or requests
- Data manipulation language (DML): commands that manipulate the data in a database
- Structured Query Language (SQL): ANSI standard query language for relational databases
- Database programs can produce reports, documents, and other outputs

### **Database Administration**

- Database administrator (DBA): directs or performs all activities to maintain a database environment
- Designing, implementing, and maintaining the database system and the DBMS
- Establishing policies and procedures
- Training employees

### **Popular Database Management Systems**

- Popular DBMSs for end users: Microsoft Access and Corel Paradox
- The complete database management software market includes databases by IBM, Oracle, and Microsoft
- Examples of open-source database systems: PostgreSQL and MySQL

- Many traditional database programs are now available on open-source operating systems

### Special-Purpose Database Systems

- Summation and Concordance
- CaseMap
- LiveNote
- Scottish Intelligence Database (SID)
- GlobalSpec

### Selecting a Database Management System

- Important characteristics of databases to consider:
  - Size of the database
  - Number of concurrent users
  - Performance
  - Ability to be integrated with other systems
  - Features of the DBMS
  - Vendor considerations
  - Cost of the system

### Using Databases with Other Software

- Database management systems are often used with other software packages or the Internet
- A database management system can act as a front-end application or a back-end application
  - Front-end application: interacts with users
  - Back-end application: interacts with applications

### Database Applications: Linking the Company Database to the Internet

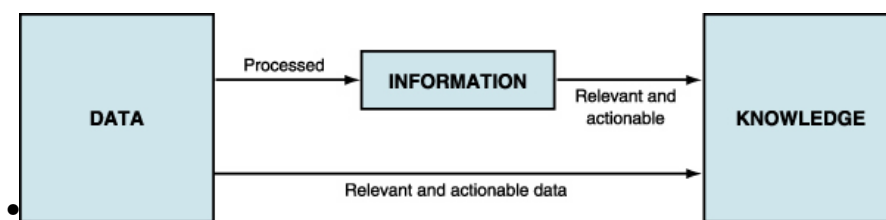
- Corporate databases can be accessed by customers, suppliers, and employees through:
  - The Internet
  - Intranets
  - Extranets
- Semantic Web: Developing a seamless integration of traditional databases with the Internet

### Data Warehouses, Data Marts, and Data Mining

- Data warehouse: subject-oriented historical data, organized to enable analytical processing.
- Data mart: a *small data warehouse* designed for a *strategic business unit* or *department*.
- Data mining: an information-analysis tool for discovering patterns and relationships in a data warehouse, data mart or database. Can predict trends and discover previously undiscovered patterns

### Business Intelligence

- Business intelligence (BI): gathering the right information in a timely manner and usable form and analyzing it to have a positive impact on business. Can make better business & strategic decisions.
- Knowledge management: A process that helps organizations identify, select, organize, disseminate, transfer, and apply information and expertise that are part of the organization's memory and that typically reside within the organization in an unstructured manner



### **Distributed database**

- Data may be spread across several smaller databases connected via telecommunications devices
- Corporations get more flexibility in how databases are organized and used
- Replicated database: Holds a duplicate set of frequently used data

### **Online Analytical Processing (OLAP)**

- Software that allows analytical processing of data when a transaction occurs instead of historically

### **Visual, Audio, and Other Database Systems**

- Visual database systems: Visual presentation of data using Graphics, multidimensional tables and charts, videos and animation. Geographical: displaying data using digitized maps.
- Audio database systems
- Virtual database systems
- Spatial data technology

### **Summary**

- Hierarchy of data: bits, characters, fields, records, files, and databases
- An entity is a generalized class of things (objects) for which data is collected, stored, and maintained
- Attribute: characteristic of an entity
- Data model: diagram of entities and relationships
- Relational model: describes data in two-dimensional tables with relationships among  $\geq 2$  tables
- Selecting eliminates rows according to criteria
- Projecting eliminates columns in a table
- A database management system (DBMS) is a group of programs used as an interface between:
  - The database and application programs
  - The database and the user
- Data dictionary: detailed description of all the data used in the database
- Data warehouse: database that collects historical business information from all aspects of a company's processes, products, and customers
- Data mining: an information-analysis tool for discovering patterns and relationships in a data warehouse

## Key Terms

- **Attribute:** A characteristic of an entity
- **Business intelligence (BI):** The process of gathering enough of the right information in a timely manner and usable form and analyzing it to have a positive impact on business strategy, tactics, or operations
- **Character:** A basic building block of information, consisting of uppercase letters, lowercase letters, numeric digits, or special symbols
- **Competitive intelligence:** A continuous process involving the legal and ethical collection of information about competitors, its analysis, and controlled dissemination of information to decision makers
- **Concurrency control:** A method of dealing with a situation in which two or more people need to access the same record in a database at the same time
- **Counterintelligence:** The steps an organization takes to protect information sought by “hostile” intelligence gatherers
- **Data definition language (DDL):** A collection of instructions and commands used to define and describe data and data relationships in a specific database
- **Data dictionary:** A detailed description of all data used in the database
- **Data integrity:** The degree to which the data in any one file is accurate
- **Data item:** The specific value of an attribute
- **Data manipulation language (DML):** The commands that are used to manipulate the data in a database
- **Data mart:** A subset of a data warehouse
- **Data mining:** An information-analysis tool that involves the automated discovery of patterns and relationships in a data warehouse or data mart
- **Data model:** A diagram of entities and their relationships
- **Data redundancy:** A duplication of data in separate files
- **Data warehouse:** A database that collects business information from many sources in the enterprise, covering all aspects of the company’s processes, products, and customers
- **Database administrator (DBA):** A skilled IS professional who directs all activities related to an organization’s database
- **Database approach to data management:** An approach whereby a pool of related data is shared by multiple application programs
- **Database management system (DBMS):** A group of programs that manipulate the database and provide an interface between the database and its users and other application programs
- **Distributed database:** A database in which the data can be spread across several smaller databases connected via telecommunications devices
- **Domain:** The allowable values for data attributes
- **Enterprise data modeling:** The data modeling done at the level of the entire enterprise
- **Entity:** A generalized class of people, places, or things for which data is collected, stored, and maintained
- **Entity-relationship (ER) diagram:** The data models that use basic graphical symbols to show the organization of and relationships between data
- **Field:** Typically a name, number, or combination of characters that describes an aspect of a business object or activity
- **File:** A collection of related records
- **Hierarchy of data:** Bits, characters, fields, records, files, and databases
- **Joining:** The data manipulation that combines two or more tables
- **Key:** A field or set of fields in a record that is used to identify the record

- **Knowledge management:** The process of capturing a company's collective expertise wherever it resides—in computers, on paper, in people's heads—and distributing it wherever it can help produce the biggest payoff
- **Linking:** The data manipulation that relates or links two or more tables using common data attributes
- **Object-oriented database:** The database that stores both data and its processing instructions
- **Object-oriented database management system (OODBMS):** A group of programs that manipulate an object-oriented database and provide a user interface and connections to other application programs
- **Object-relational database management system (ORDBMS):** A DBMS capable of manipulating audio, video, and graphical data
- **Online analytical processing (OLAP):** The software that allows users to explore data from a number of different perspectives
- **Planned data redundancy:** A way of organizing data in which the logical database design is altered so that certain data entities are combined, summary totals are carried in the data records rather than calculated from elemental data, and some data attributes are repeated in more than one data entity to improve database performance
- **Predictive analysis:** A form of data mining that combines historical data with assumptions about future conditions to predict outcomes of events, such as future product sales or the probability that a customer will default on a loan
- **Primary key:** A field or set of fields that uniquely identifies the record
- **Projecting:** The data manipulation that eliminates columns in a table
- **Record:** A collection of related data fields
- **Relational model:** A database model that describes data in which all data elements are placed in two-dimensional tables, called relations, which are the logical equivalent of files
- **Replicated database:** A database that holds a duplicate set of frequently used data
- **Schema:** A description of the entire database
- **Selecting:** The data manipulation that eliminates rows according to certain criteria
- **Subschema:** A file that contains a description of a subset of the database and identifies which users can view and modify the data items in the subset
- **Traditional approach to data management:** An approach whereby separate data files are created and stored for each application program