Instructor: Mohammad Moniruzzaman Bhuiya E-mail: shagorx@gmail.com
Management Information Systems (MIS)

CHAPTER 06: FOUNDATION OF BUSINESS INTELLIGENCE DATABASES AND INFORMATION MANAGEMENT

- **What are the problems of managing data resources in a traditional file environment and how are they solved by a database management system?**
- **What are the major capabilities of DBMS and why is a relational DBMS so powerful?**
- **What are some important database design principles?**
- **What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?**
- **Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?**

- **What are the problems of managing data resources in a traditional file environment and how are they solved by a database management system?**

  Traditional file management techniques make it difficult for organizations to keep track of all of the pieces of data they use in a systematic way and to organize these data so that they can be easily accessed. Different functional areas and groups were allowed to develop their own files independently. Over time, this traditional file management environment creates problems such as data redundancy and inconsistency, program-data dependence, inflexibility, poor security, and lack of data sharing and availability. A database management system (DBMS) solves these problems with software that permits centralization of data and data management so that businesses have a single consistent source for all their data needs. Using a DBMS minimizes redundant and inconsistent files.

  - **Problems with the traditional file environment (files maintained separately by different departments)**
    - **Data redundancy**
      - Presence of duplicate data in multiple files
    - **Data inconsistency**
      - Same attribute has different values
    - **Program-data dependence**
      - When changes in program requires changes to data accessed by program
    - **Lack of flexibility**
    - **Poor security**
    - **Lack of data sharing and availability**

  - **The Database Approach to Data Management**
    - **Database**
      - Serves many applications by centralizing data and controlling redundant data
    - **Database management system (DBMS)**
      - Interfaces between applications and physical data files
      - Separates logical and physical views of data
      - Solves problems of traditional file environment
        - Controls redundancy
        - Eliminates inconsistency

        - Uncouples programs and data
        - Enables organization to centrally manage data and data security
    - **Relational DBMS**
        - Represent data as two-dimensional tables called relations or files
        - Each table contains data on entity and attributes
        - Table: grid of columns and rows
            - Rows (tuples): Records for different entities
            - Fields (columns): Represents attribute for entity
            - Key field: Field used to uniquely identify each record
            - Primary key: Field in table used for key fields
            - Foreign key: Primary key used in second table as look-up field to identify records from original table
        - Operations of a Relational DBMS
            - Three basic operations used to develop useful sets of data
                - SELECT: Creates subset of data of all records that meet stated criteria
                - JOIN: Combines relational tables to provide user with more information than available in individual tables
                - PROJECT: Creates subset of columns in table, creating tables with only the information specified
    - **Object-Oriented DBMS (OODBMS)**
        - Stores data and procedures as objects
        - Objects can be graphics, multimedia, Java applets
        - Relatively slow compared with relational DBMS for processing large numbers of transactions
        - Hybrid object-relational DBMS: Provide capabilities of both OODBMS and relational DBMS
    - **Databases in the cloud**
        - Typically less functionality than on-premises DBs
        - Amazon Web Services, Microsoft SQL Azure

- **What are the major capabilities of DBMS and why is a relational DBMS so powerful?**

  The principal capabilities of a DBMS includes a data definition capability, a data dictionary capability, and a data manipulation language. The data definition capability specifies the structure and content of the database. The data dictionary is an automated or manual file that stores information about the data in the database, including names, definitions, formats, and descriptions of data elements. The data manipulation language, such as SQL, is a specialized language for accessing and manipulating the data in the database.

  The relational database is the primary method for organizing and maintaining data today in information systems because it is so flexible and accessible. It organizes data in two-dimensional tables called relations with rows and columns. Each table contains data about an entity and its attributes. Each row represents a record and each column represents an attribute or field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation. Relational database tables can be combined easily to deliver data required by users, provided that any two tables share a common data element.

  - **Capabilities of Database Management Systems**

    - Data definition capability: Specifies structure of database content, used to create tables and define characteristics of fields
    - Data dictionary: Automated or manual file storing definitions of data elements and their characteristics
    - Data manipulation language: Used to add, change, delete, retrieve data from database
        - Structured Query Language (SQL)
        - Microsoft Access user tools for generation SQL
    - Many DBMS have report generation capabilities for creating polished reports (Crystal Reports)

The relational database is the primary method for organizing and maintaining data today in information systems because it is so flexible and accessible. It organizes data in two-dimensional tables called relations with rows and columns. Each table contains data about an entity and its attributes. Each row represents a record and each column represents an attribute or field. Each table also contains a key field to uniquely identify each record for retrieval or manipulation. Relational database tables can be combined easily to deliver data required by users, provided that any two tables share a common data element

- **What are some important database design principles?**

Designing a database requires both a logical design and a physical design. The logical design models the database from a business perspective. The organization's data model should reflect its key business processes and decision-making requirements. The process of creating small, stable, flexible, and adaptive data structures from complex groups of data when designing a relational database is termed normalization. A well-designed relational database will not have many-to-many relationships, and all attributes for a specific entity will only apply to that entity. It will try to enforce referential integrity rules to ensure that relationships between coupled tables remain consistent. An entity-relationship diagram graphically depicts the relationship between entities (tables) in a relational database.

  - **Designing Databases**
    - Conceptual (logical) design: Abstract model from business perspective
    - Physical design: How database is arranged on direct-access storage devices
  - **Design process identifies**
    - Relationships among data elements, redundant database elements
    - Most efficient way to group data elements to meet business requirements, needs of application programs
  - **Normalization**
    - Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships
  - **Entity-relationship diagram**
    - Used by database designers to document the data model
    - Illustrates relationships between entities (tables) in a RDBMS
  - **Distributing databases: Storing database in more than one place**
    - Partitioned: Separate locations store different parts of database
    - Replicated: Central database duplicated in entirety at different locations

- **What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?**

Powerful tools are available to analyze and access the information in databases. A data warehouse consolidates current and historical data from many different operational systems in a central database designed for reporting and analysis. Data warehouses support multidimensional data analysis, also known as online analytical processing (OLAP). OLAP represents relationships among data as a multidimensional structure, which can be visualized as cubes of data and cubes within cubes of data, enabling more sophisticated data analysis. Data mining analyzes large pools of data, including the contents of data warehouses, to find patterns and rules that can be used to predict future behavior and guide decision making. Text mining tools help businesses analyze large unstructured data sets consisting of text. Web mining tools focus on analysis of useful patterns and information from the World Wide Web, examining the structure of Web sites and activities of Web site users as well as the contents of Web pages. Conventional databases can be linked via middleware to the Web or a Web interface to facilitate user access to an organization's internal data.

  - **Using Databases to Improve Business Performance and Decision Making**
    - Very large databases and systems require special capabilities, tools

- To analyze large quantities of data
- To access data from multiple systems
- Three key techniques
  - Data warehousing
  - Data mining
  - Tools for accessing internal databases through the Web

- **Data warehouse**
  - Stores current and historical data from many core operational transaction systems
  - Consolidates and standardizes information for use across enterprise, but data cannot be altered
  - Data warehouse system will provide query, analysis, and reporting tools
- **Data marts**
  - Subset of data warehouse
  - Summarized or highly focused portion of firm's data for use by specific population of users
  - Typically focuses on single subject or line of business
- **Business Intelligence**
  - Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - E.g., Harrah's Entertainment analyzes customers to develop gambling profiles and identify most profitable customers
  - Principle tools include:
    - Software for database query and reporting
    - Online analytical processing (OLAP)
    - Data mining
- **Online analytical processing (OLAP)**
  - Supports multidimensional data analysis
    - Viewing data using multiple dimensions
    - Each aspect of information (product, pricing, cost, region, time period) is different dimension
    - E.g., how many washers sold in the East in June compared with other regions?
  - OLAP enables rapid, online answers to ad hoc queries
- **Data mining**
  - More discovery driven than OLAP
  - Finds hidden patterns, relationships in large databases and infers rules to predict future behavior
  - E.g., Finding patterns in customer data for one-to-one marketing campaigns or to identify profitable customers.
  - Types of information obtainable from data mining
    - Associations
    - Sequences
    - Classification
    - Clustering
    - Forecasting
- **Predictive analysis**
  - Uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events
  - E.g., Probability a customer will respond to an offer
- **Text mining**
  - Extracts key elements from large unstructured data sets (e.g., stored e-mails)
- **Web mining**
  - Discovery and analysis of useful patterns and information from WWW
    - E.g., to understand customer behavior, evaluate effectiveness of Web site, etc.
  - Web content mining
    - Knowledge extracted from content of Web pages
  - Web structure mining

        - E.g., links to and from Web page
    - Web usage mining
        - User interaction data recorded by Web server
- **Databases and the Web**
    - Many companies use Web to make some internal databases available to customers or partners
    - Typical configuration includes:
        - Web server
        - Application server/middleware/CGI scripts
        - Database server (hosting DBM)
    - Advantages of using Web for database access:
        - Ease of use of browser software
        - Web interface requires few or no changes to database
        - Inexpensive to add Web interface to system

- **Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?**

  Developing a database environment requires policies and procedures for managing organizational data as well as a good data model and database technology. A formal information policy governs the maintenance, distribution, and use of information in the organization. In large corporations, a formal data administration function is responsible for information policy, as well as for data planning, data dictionary development, and monitoring data usage in the firm.

  Data that are inaccurate, incomplete, or inconsistent create serious operational and financial problems for businesses because they may create inaccuracies in product pricing, customer accounts, and inventory data, and lead to inaccurate decisions about the actions that should be taken by the firm. Firms must take special steps to make sure they have a high level of data quality. These include using enterprise-wide data standards, databases designed to minimize inconsistent and redundant data, data quality audits, and data cleansing software.

  - **Establishing an information policy**
    - Firm's rules, procedures, roles for sharing, managing, standardizing data
    - Data administration:
        - Firm function responsible for specific policies and procedures to manage data
    - Data governance:
        - Policies and processes for managing availability, usability, integrity, and security of enterprise data, especially as it relates to government regulations
    - Database administration:
        - Defining, organizing, implementing, maintaining database; performed by database design and management group

  - **Ensuring data quality**
    - More than 25% of critical data in Fortune 1000 company databases are inaccurate or incomplete
    - Most data quality problems stem from faulty input
    - Before new database in place, need to:
        - Identify and correct faulty data
        - Establish better routines for editing data once database is in operation.