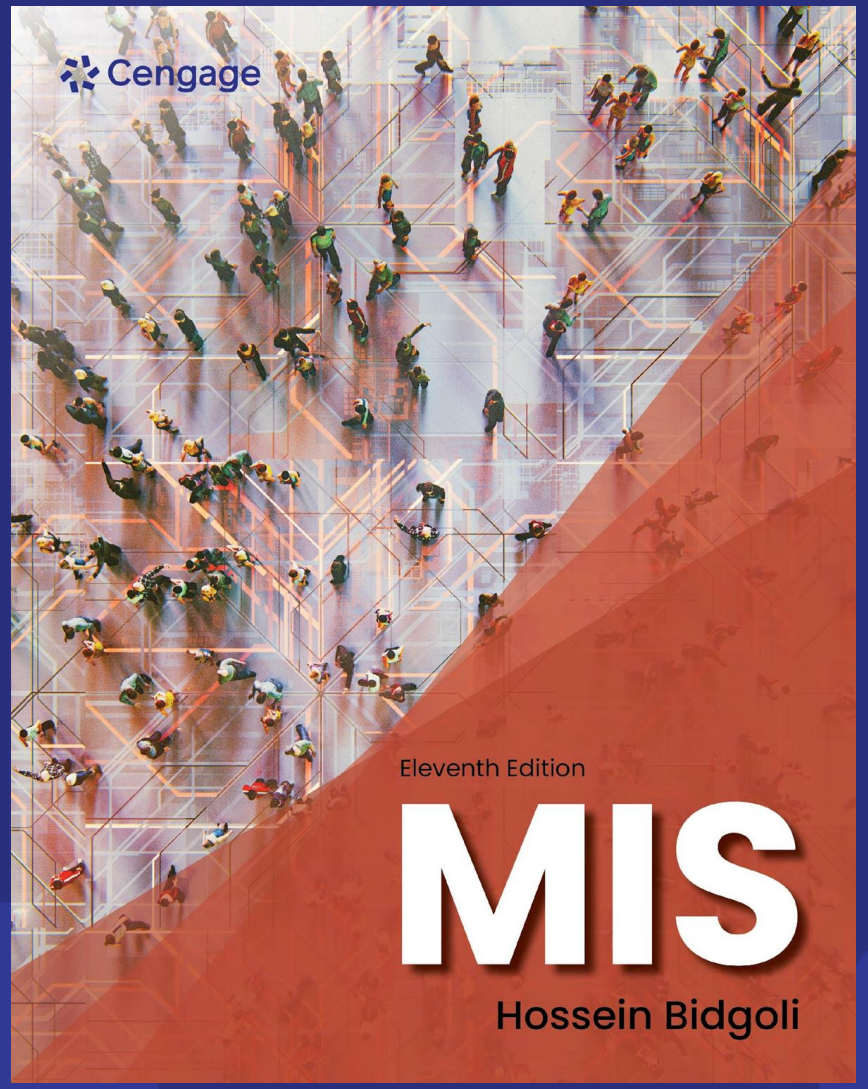


MIS, 11e

Module 3: Data and Business Intelligence

Module Objectives

By the end of this module, you should be able to:

- 3.1 Define a database and a database management system.
- 3.2 Explain logical database design and the relational database model.
- 3.3 Define the five components of a database management system.
- 3.4 Summarize three recent trends in database design and use.
- 3.5 Analyze the four major components and functions of a data warehouse and their use for business.
- 3.6 Describe the functions of a data mart.
- 3.7 Compare and contrast data lakes with data warehouses.
- 3.8 Describe the role of business analytics in the decision-making process.
- 3.9 Examine the advantages and challenges of big data and predictive analytics for a business.
- 3.10 Explain database marketing and its business applications.

Databases

- **Database** – Collection of related data that is stored in a central location or in multiple locations
- **Data hierarchy** – Structure and organization of data, which involves fields, records, and files (Ref to Exhibit 3.1 in the textbook)
- **Database management system (DBMS)** – Software for creating, storing, maintaining, and accessing database files

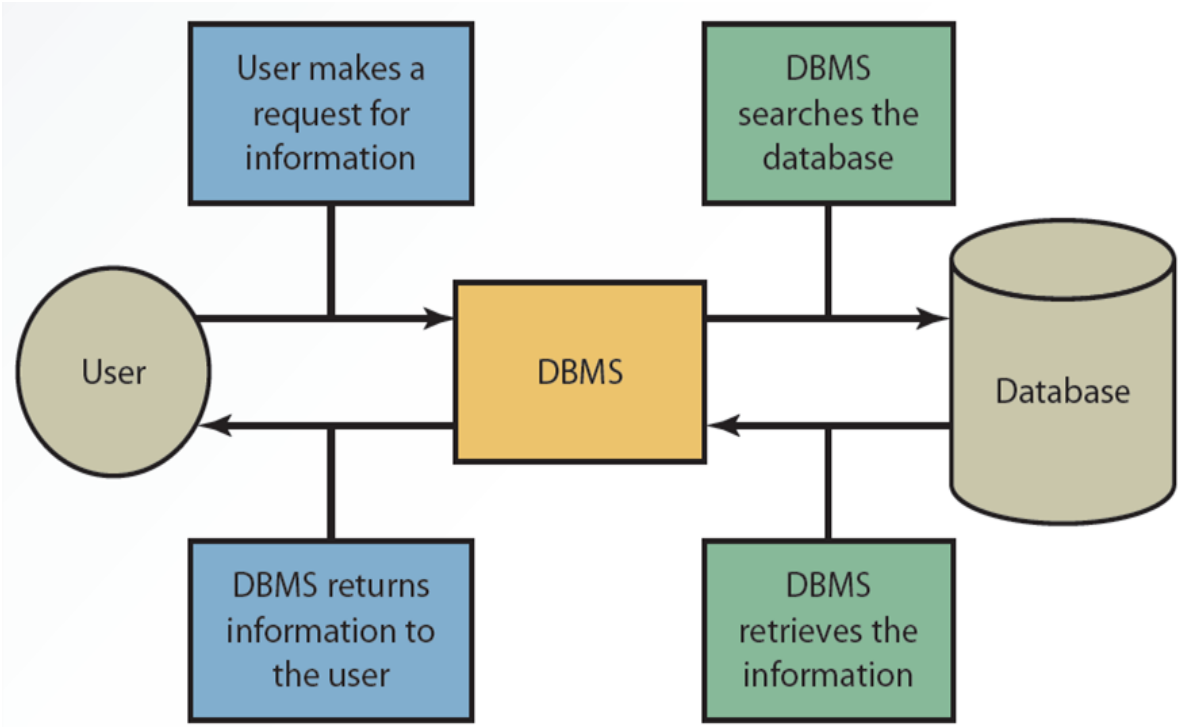


Exhibit 3.2 - Interaction between the User, DBMS, and Database

Methods for Accessing Files

Sequential access file structure

- Records are organized and processed in numerical or sequential order
- Records are organized based on a primary key
- Used for backup and archive files
- Typically stored on magnetic tape.

Random access file structure

- Records can be accessed in any order, regardless of their physical locations in storage media
- Fast and very effective when a small number of records need to be processed daily or weekly
- Records are stored on magnetic disks to achieve speed

Indexed sequential access method (ISAM)

- Records accessed sequentially or randomly, depending on the amount
  - Random access: small number of records
  - Sequential access: large number of records
- Uses an index structure with two parts:
  - Indexed value
  - Pointer to the disk location of the record matching the indexed value

Database Design

Information is viewed in a database in two ways

- Physical view – how data is stored on and retrieved from storage media
    - Hard disks or magnetic tapes
- Logical view – how information appears, and is organized and retrieved
    - Depending on the user, there can be more than one logical view

Data model – determines how data is created, represented, organized, and maintained

Includes:

- Data structure - organization
- Operations – methods, calculations, etc.

Hierarchical model

- Relationships (or branches) between records (or nodes)

Network model

- Similar to hierarchical model but organized differently
- Each record can have multiple parent and child records

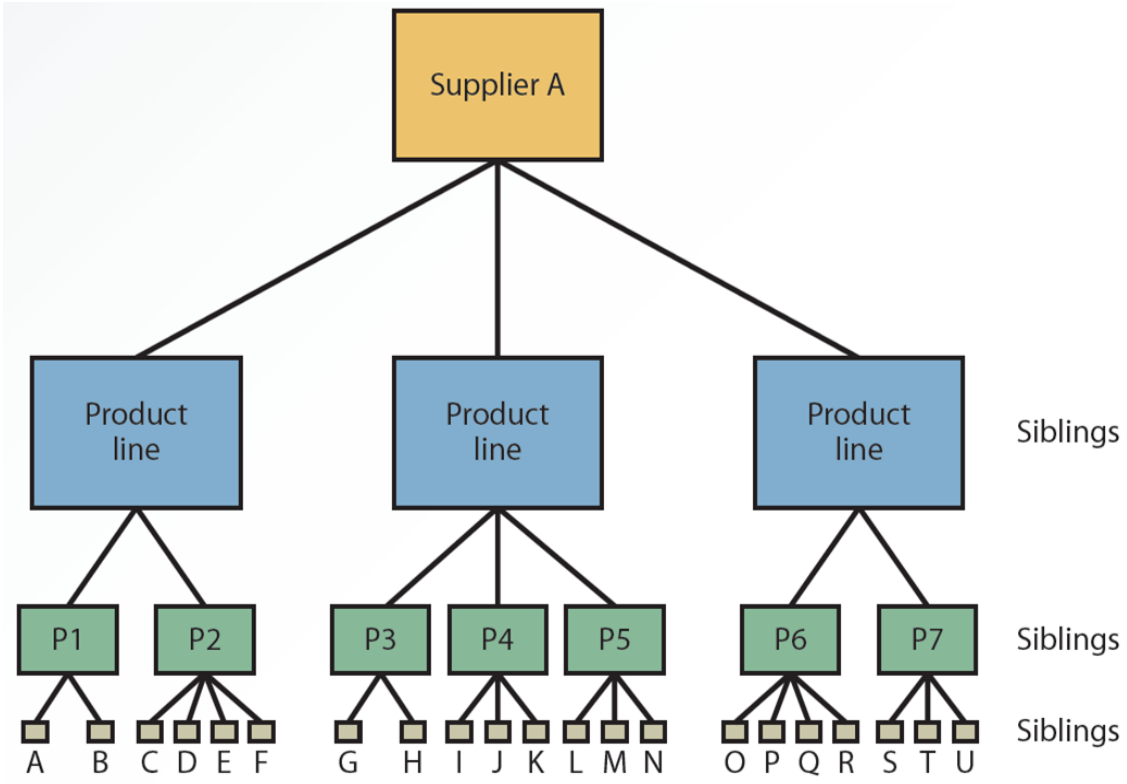


Exhibit 3.3 - A Hierarchical Model

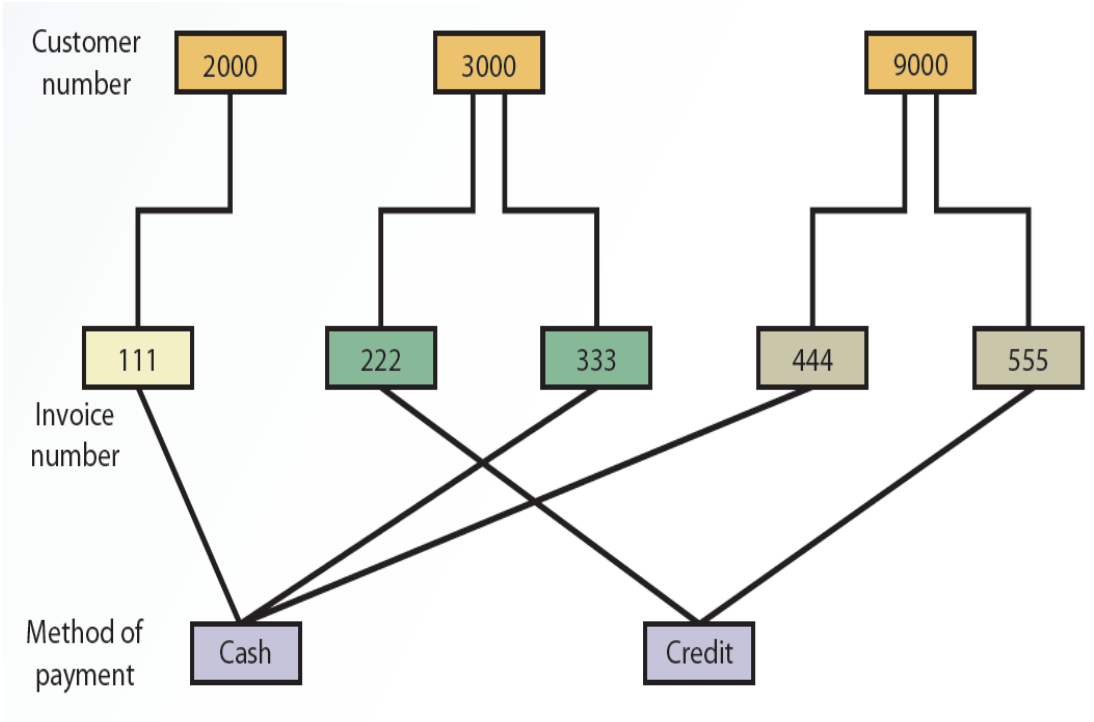


Exhibit 3.4 - A Network Model

The Relational Model (1 of 2)

Relation Model

- Uses a two-dimensional table of rows and columns of data
    - Rows are records (i.e., tuples)
    - Columns are fields (i.e., attributes)

Data dictionary

- Stores definitions
- Examples: Field name, field data type, default value, validation rule

Primary key

- Uniquely identifies every record

Foreign key

- A key of a child table that links to primary key of parent table
- Used to establish relationship between tables

The Relational Model (2 of 2)

Normalization – eliminates redundant data (Ref to the supplement doc)

- Ensures only related data is stored in a table
- First normal form (1NF) to fifth normal form (5NF)

Operations – process in which data is retrieved from tables

- Common operations: select, project, join, intersect, union, and difference

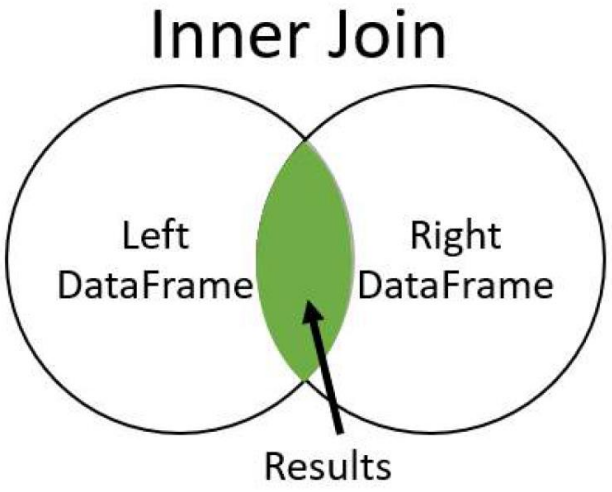
Basic Select Operation

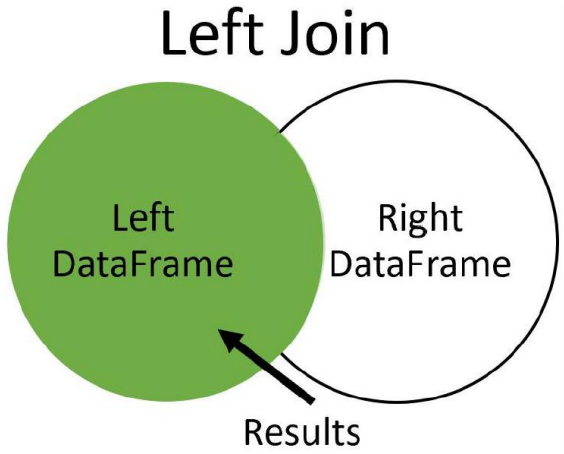
SELECT field **FROM** table or file **WHERE** conditions

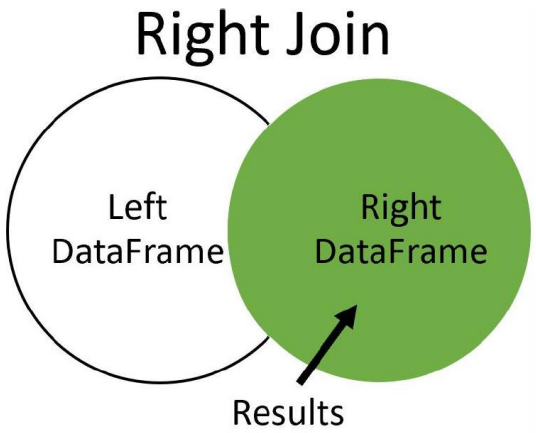
Example:

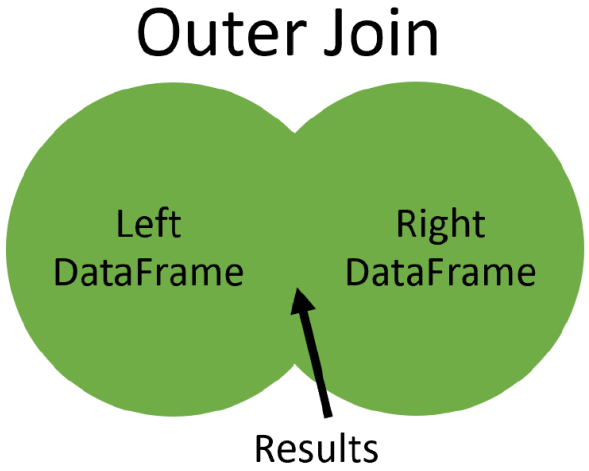
SELECT NAME, SSN, TITLE, GENDER, SALARY **FROM** EMPLOYEE, PAYROLL **WHERE** EMPLOYEE.SSN = PAYROLL.SSN AND TITLE = “ENGINEER”

Various types of Joins









Left Join Example

Left join

Left Table

A	B	C
A2	B2	C2
A3	B3	C3
A4	B4	C4

Right Table

C	D
C1	D1
C2	D2
C4	D4
C5	D5

=

Result Table

A	B	C	D
A2	B2	C2	D2
A3	B3	C3	
A4	B4	C4	D4

Right Join Example

Right join

Left Table

A	B	C
A2	B2	C2
A3	B3	C3
A4	B4	C4

Right Table

C	D
C1	D1
C2	D2
C4	D4
C5	D5

=

Result Table

A	B	C	D
		C1	D1
A2	B2	C2	D2
A4	B4	C4	D4
		C5	D5

Outer Join Example

Outer join

Left Table

A	B	C
A2	B2	C2
A3	B3	C3
A4	B4	C4

Right Table

C	D
C1	D1
C2	D2
C4	D4
C5	D5

=

Result Table

A	B	C	D
		C1	D1
A2	B2	C2	D2
A3	B3	C3	
A4	B4	C4	D4
		C5	D5

Knowledge Check Activity 3-1

In a relational model, rows in a table are used for which of the following?

a. Primary key

b. Records

c. Columns

d. Attributes

Knowledge Check Activity 3-1: Answer

In a relational model, rows in a table are used for which of the following?

Answer: Records

Each row in a table is a record. For example, a record for a person's contact information may include first name, last name, street address, city, state, and zip code.

Components of a DBMS (1 of 5)

Database engine

- Responsible for data storage, manipulation, and retrieval

Data definition

- Creates and maintains the data dictionary
- Defines the structure of files in a database

Data manipulation

- Used to add, delete, modify, and retrieve records
- Query languages like Structured Query Language (SQL)

Application generation

- Used to create designs elements of an application
- Example: create a menu system on an application

Data administration

- Used for backup and recovery, security, and change management
- Used to determine permissions to create, read, update, and delete (CRUD)

Database administrators (DBAs)

- Handle database design and management
- Establish security measures
- Develop recovery procedures
- Evaluate database performance
- Add and fine-tune database functions

Data-Driven Web Sites

Act as an interface to a database

- Retrieve data and allow users to enter data

Improve access to information

- Reduce support and overhead needed to maintain static Websites
- Give users more current information from a variety of data sources

Distributed Databases

Distributed Database Management System (DDBMS)

- Stores data on multiple servers throughout an organization
- Several advantages
  - Design better reflects the firm's structure
  - Local data storage reduces response time
  - Minimizes effects of computer failure
  - Not limited by data's physical location

Approaches to setting up a DDBMS

- Fragmentation: how tables are divided among multiple locations
- Replication: each site stores a copy of the data
- Allocation: combines fragmentation and replication

Object-Oriented Databases

Data and their relationships are contained in a single object

- An object consists of attributes and methods that can be performed on the object's data
  - Encapsulation: grouping objects with their attributes and methods into a class
  - Inheritance: new objects can be created faster and more easily by entering new data in attributes

Advantages

- Supports more complex data management
- Handles storing and manipulating all types of multimedia as well as numbers and characters

Knowledge Check Activity 3-2

Which of the following can act as an interface for a database?

a. Website

b. Database engine

c. DBA

d. SQL

Knowledge Check Activity 3-2: Answer

Which of the following can act as an interface for a database?

Answer: Website

A website can act as an interface to retrieve data from a database and allow users to enter data into the database.

Data Warehouses

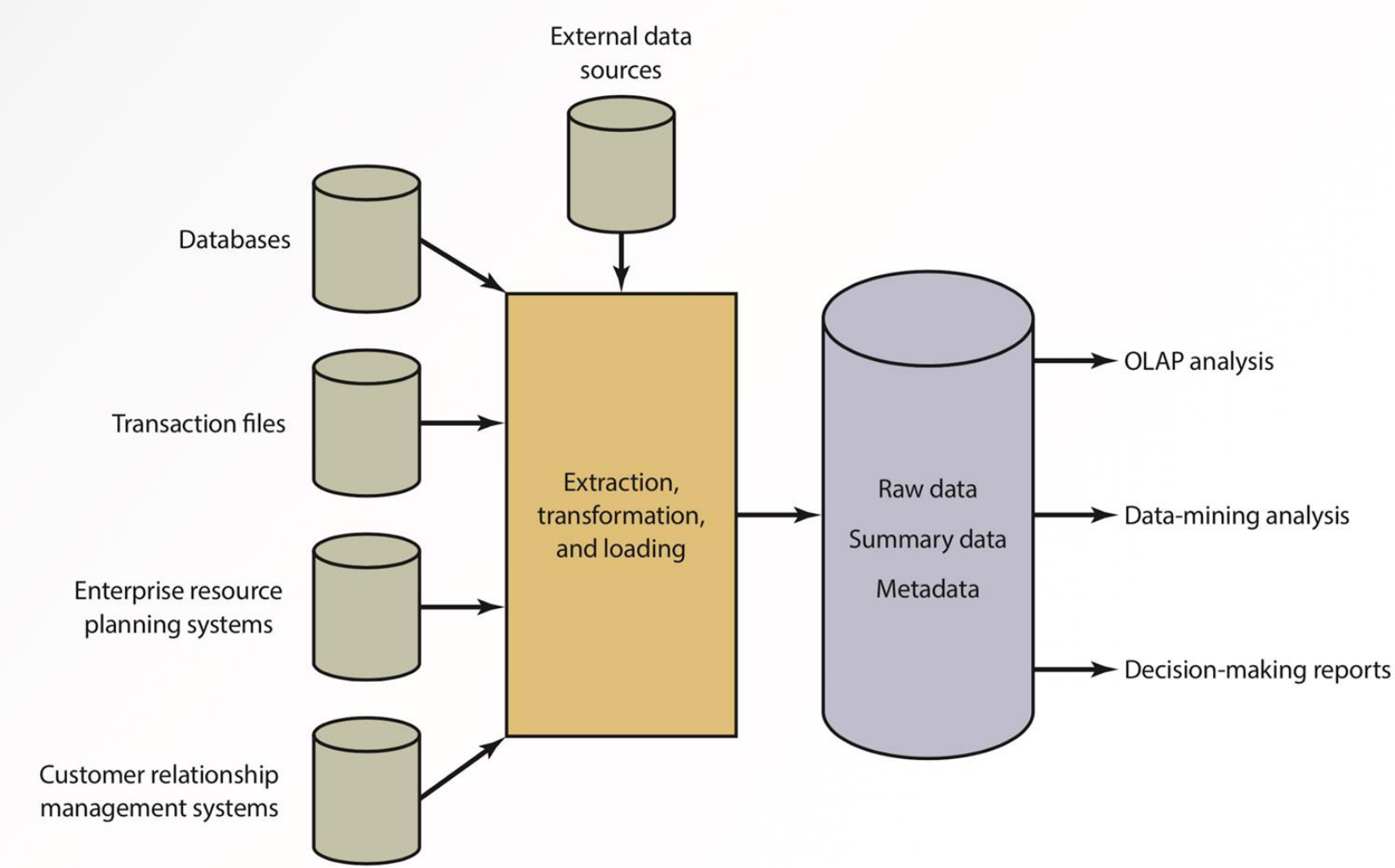
Data Warehouse – Collection of data from a variety of sources

- Used to support decision-making applications
- Used to generate business intelligence
- Also known as hypercubes because they store multidimensional data

Characteristics of data in a data warehouse

- Subject oriented – focused on a specific area
- Integrated - comes from a variety of sources
- Time variant - categorized based on time
- Type of data - captures aggregated data
- Purpose - used for analytical purposes

Exhibit 3.6 – A Data Warehouse Configuration



Input

A variety of data sources provide the input for a data warehouse to perform analyses and generate reports.

These sources can include:

- External data sources, databases, and transaction files
- Enterprise resource planning (ERP) systems
- Customer relationship management (CRM) systems

ETL

Extraction, transformation, and loading

- Extracting data from outside sources
- Transform data to fit operational needs
- Loading into the end target (database or data warehouse)

Storage

Collection of information stored as:

- Raw data - information in its original form
- Summary data - gives users subtotals of various categories
- Metadata: information about data's
  - Examples: content, quality, condition, origin, and others

Output (1 of 2)

- **Online transaction processing (OLTP)**
  - Generates reports for decision making
- **Online analytical processing (OLAP)**
  - Quickly answers multidimensional analytical queries
  - Generates business intelligence

Output (2 of 2)

Data-mining analysis

- Used to discover patterns and relationships

Text-mining (or text data-mining) analysis

- Used to analyze vast amounts of textual information
- Capture key concepts, trends, and hidden relationships

Data Warehouse Benefits

- Cross-reference segments of an organization's operations for comparison
- Generate complex queries and reports faster than databases
- Generate reports efficiently using data from a variety of sources
- Find patterns and trends that cannot be found with databases
- Analyze large amounts of historical data quickly.

Data Mart – Smaller version of a data warehouse

- Advantages over data warehouses
  - Faster access to data due to their smaller size
  - Improved response time for users
- Disadvantages over data warehouses
  - Limited scope
  - Difficulty consolidating information

Data Lake – gathers and stores data in its original format in a central location

- Collected data can be structured as well as unstructured
- Suitable for big data analytics and machine learning applications

Knowledge Check Activity 3-3

How would you summarize the purpose the data-mining analysis process?

a. Discover patterns

b. Quick access to data

c. Gather data in a central location

d. Store metadata information

Knowledge Check Activity 3-3: Answer

How would you summarize the purpose the data-mining analysis process?

Answer: Discover patterns

Data-mining analysis discovers patterns and relationship in a data warehouse.

Business Analytics

Uses data and statistical methods

- Gains insight into the data
- Provides decision makers with information to act on

Methods

- Descriptive - Review past events, analyze data, provide a report
- Predictive - Prepare decision makers for future events
- Prescriptive - Shows likely outcome of each decision

The Big Data Era

Big Data is voluminous data

Five dimensions (the 5 Vs): Volume, Variety, Velocity, Veracity, Value

- Commonly used platform: Apache Hadoop
- Privacy risks: Discrimination, privacy breaches, loss of anonymity
- Integration with IoT: Reveal trends and find unseen patterns

Database Marketing

- Uses an organization's database of customers and potential customers to promote products or services
- Transforms marketing into a proactive process
- Successful task examples:
  - Calculating customer lifetime value (CLTV)
  - Recency, frequency, and monetary analysis (RFM)
  - Customer communications
  - Analytical software

Summary

Now that the lesson has ended, you should be able to:

- 3.1 Define a database and a database management system.
- 3.2 Explain logical database design and the relational database model.
- 3.3 Define the five components of a database management system.
- 3.4 Summarize three recent trends in database design and use.
- 3.5 Analyze the four major components and functions of a data warehouse and their use for business.
- 3.6 Describe the functions of a data mart.
- 3.7 Compare and contrast data lakes with data warehouses.
- 3.8 Describe the role of business analytics in the decision-making process.
- 3.9 Examine the advantages and challenges of big data and predictive analytics for a business.
- 3.10 Explain database marketing and its business applications.