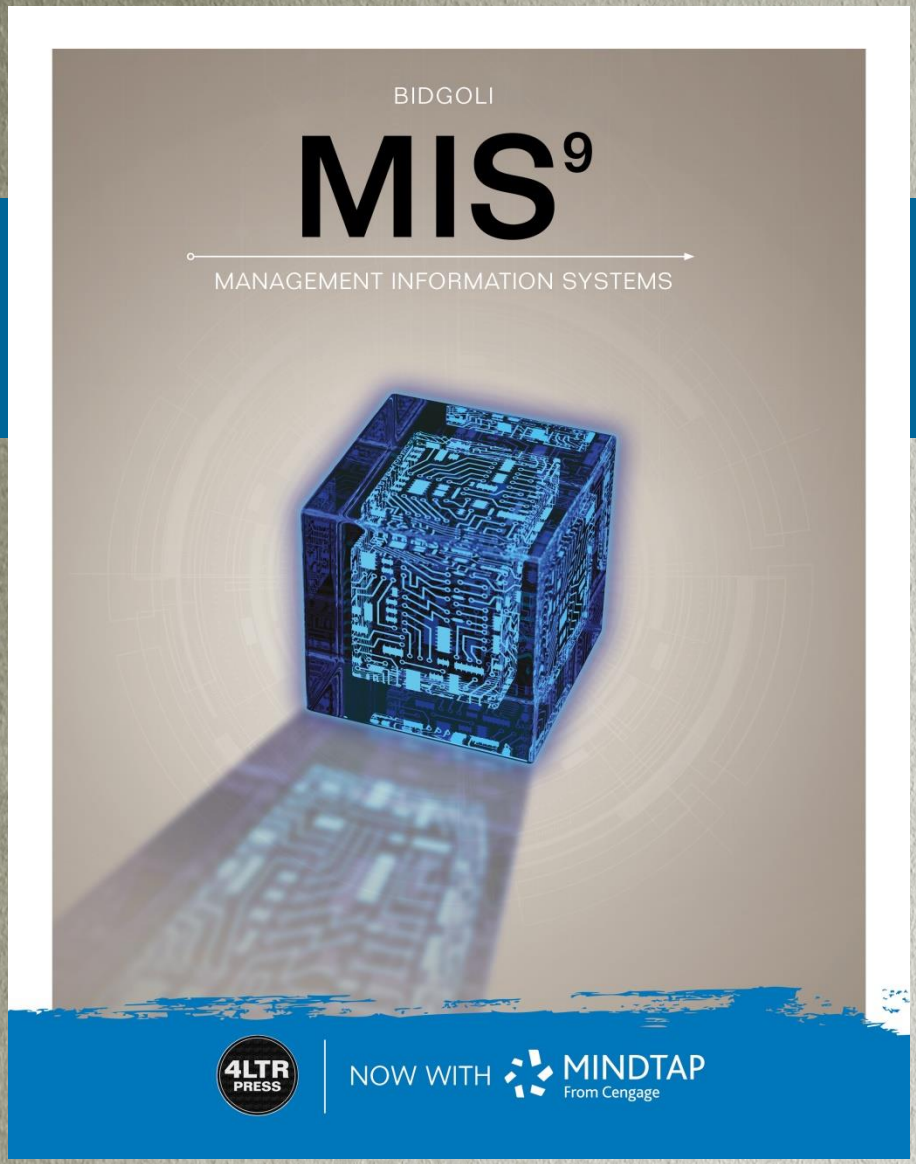


3

Database Systems, Data Warehouses, and Data Marts

Learning Objectives (1 of 2)

- Define a database and a database management system
- Explain logical database design and the relational database model
- Define the components of a database management system
- Summarize recent trends in database design and use

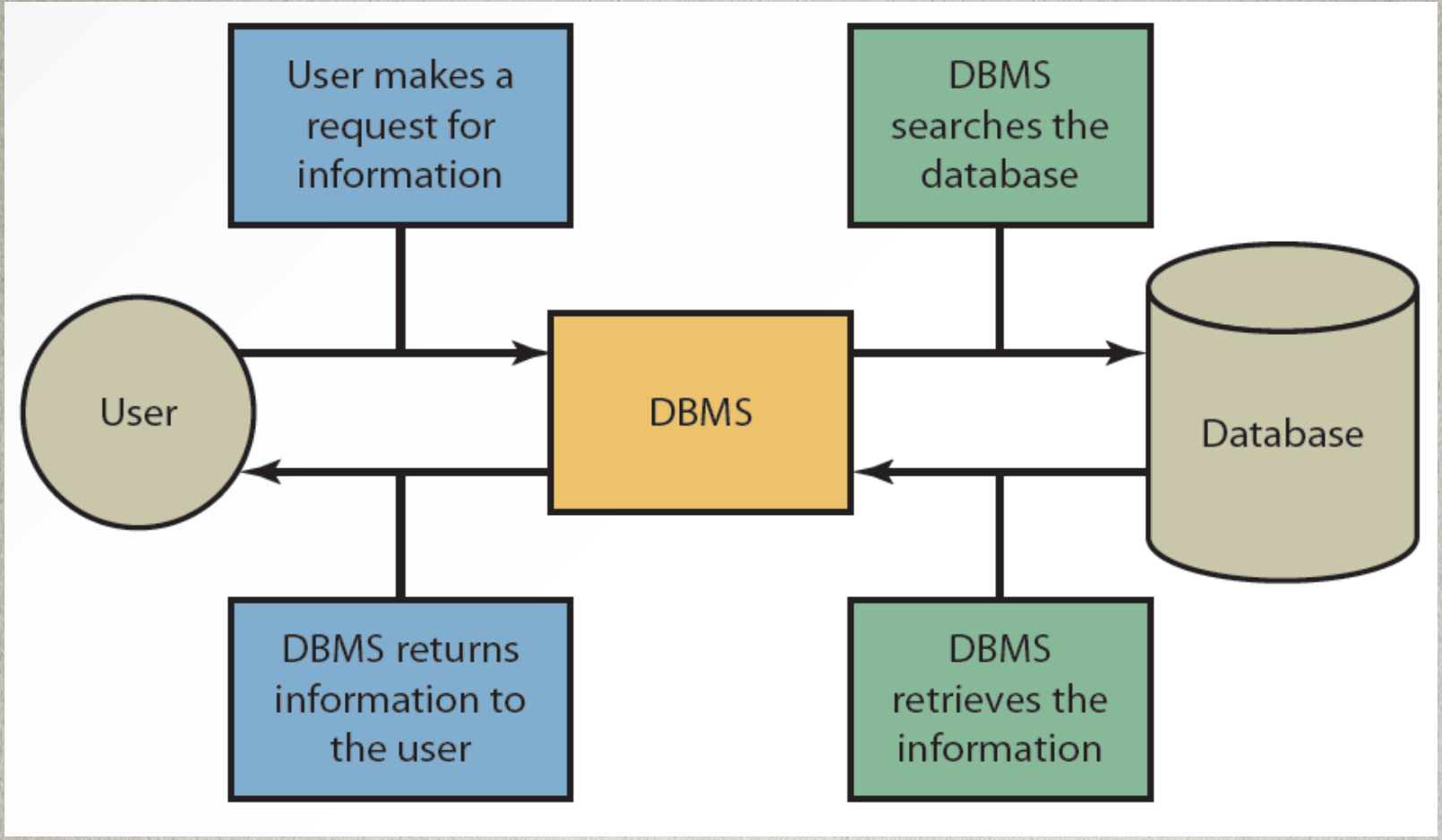
Learning Objectives (2 of 2)

- Explain the components and functions of a data warehouse
- Describe the functions of a data mart
- Define business analytics and describe its role in the decision-making process
- Explain big data and its business applications
- 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
- Database management system (DBMS)
  - Software for creating, storing, maintaining, and accessing database files
  - Makes using databases more efficient

Exhibit 3.2 Interaction between the User, DBMS, and Database



Types of Data in a Database

- Internal data
  - Collected from within an organization
  - Stored in the organization's internal databases and can be used by functional information systems
- External data
  - Comes from a variety of sources
  - Stored in a data warehouse

Methods for Accessing Files (1 of 3)

- Sequential access file structure
  - Records in files are organized and processed in numerical or sequential order
  - Records are organized based on a primary key (e.g., Social Security numbers or account numbers)
  - Used for backup and archive files because they rarely need updating

Methods for Accessing Files (2 of 3)

- 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

Methods for Accessing Files (3 of 3)

- Indexed sequential access method (ISAM)
  - Records accessed sequentially or randomly, depending on the number accessed
    - Random access: used for a small number
    - Sequential access: used for a large number
    - Uses an index structure and has two parts
      - Indexed value
      - Pointer to the disk location of the record matching the indexed value

Logical Database Design (1 of 3)

- Information is viewed in a database in two ways
  - Physical view: how data is stored on and retrieved from storage media
  - Logical view: how information appears to users and how it can be organized and retrieved
    - Depending on the user, there can be more than one logical view of data

Logical Database Design (2 of 3)

- Data model determines how data is created, represented, organized, and maintained
  - Includes:
    - Data structure
    - Operations
    - Integrity rules

Logical Database Design (3 of 3)

- Hierarchical model
  - Relationships between records form a treelike structure
  - Records are called nodes, and relationships between records are called branches
- Network model
  - Similar to the hierarchical model but records are organized differently
  - Each record can have multiple parent and child records

Exhibit 3.3 A Hierarchical Model

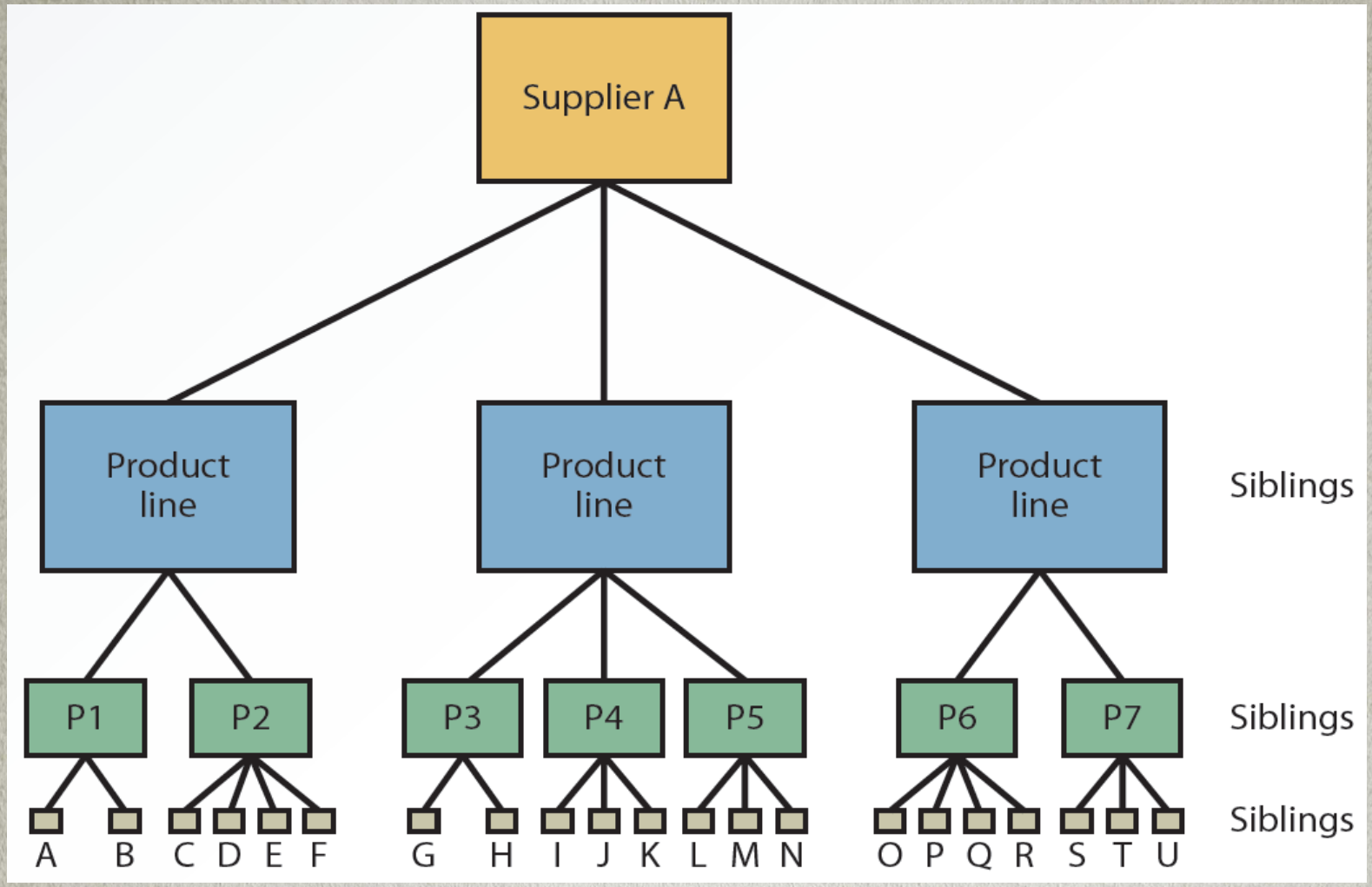
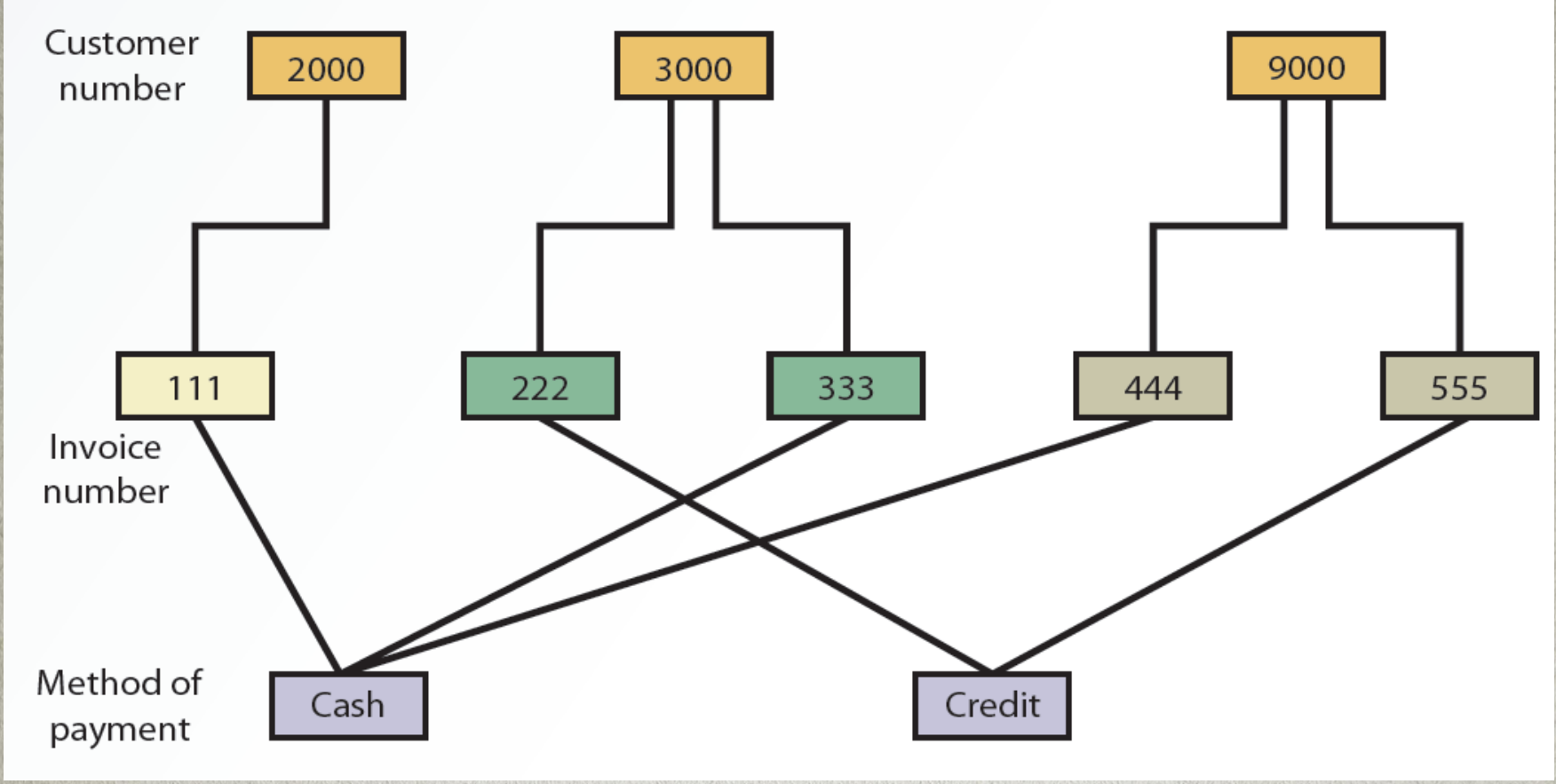


Exhibit 3.4 A Network Model



The Relational Model (1 of 4)

- 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, such as data types for fields, default values, and validation rules for data in each field

The Relational Model (2 of 4)

- Primary key
  - Uniquely identifies every record in a relational database
- Foreign key
  - Field in a relational table that matches the primary key column of another table
  - Used to cross-reference tables

The Relational Model (3 of 4)

- Normalization
  - Used to improve database efficiency
    - Eliminates redundant data
    - Ensures only related data is stored in a table
  - Goes through different stages, from first normal form (1NF) to fifth normal form (5NF)

The Relational Model (4 of 4)

- Operations
  - Help retrieve data from tables
  - Common operations: select, project, join, intersect, union, and difference

Components of a DBMS (1 of 4)

- DBMS software components
  - Database engine
  - Data definition
  - Data manipulation
  - Application generation
  - Data administration

Components of a DBMS (2 of 4)

- Database engine
  - Responsible for data storage, manipulation, and retrieval
  - Interacts with other components of the DBMS to convert logical requests from users into their physical equivalents
- Data definition
  - Used to create and maintain the data dictionary and define database file structure
  - Makes changes to a database's structure

Components of a DBMS (3 of 4)

- Data manipulation
  - Used to add, delete, modify, and retrieve records from a database
  - Uses a query language, such as Structured Query Language (SQL)
- Application generation
  - Designs elements of an application using a database
  - Used by IT professionals and database administrators

Components of a DBMS (4 of 4)

- Data administration
  - Used for tasks such as backup and recovery, security, and change management
  - Used to determine who has permission to perform certain functions, summarized as create, read, update, and delete (CRUD)
- Database administrators (DBAs)
  - Handle database design and management

Recent Trends in Database Design and Use

- Include:
  - Data-driven Web sites
  - Natural language processing
  - Distributed databases
  - Object-oriented databases
  - Advances in artificial intelligence

Data-Driven Web Sites

- Acts as an interface to a database
  - Retrieves data and allows users to enter data in the database
- Improves access to information
  - Reduces support and overhead needed to maintain static Web sites
  - Gives users more current information from a variety of data sources

Distributed Databases (1 of 2)

- 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
    - Cost advantage
    - Not limited by physical location of the data

Distributed Databases (2 of 2)

- Approaches to setting up a DDBMS
  - Fragmentation: addresses how tables are divided among multiple locations
  - Replication: each site stores a copy of the data in the organization's database
  - Allocation: combines fragmentation and replication

Object-Oriented Databases (1 of 2)

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

Object-Oriented Databases (2 of 2)

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

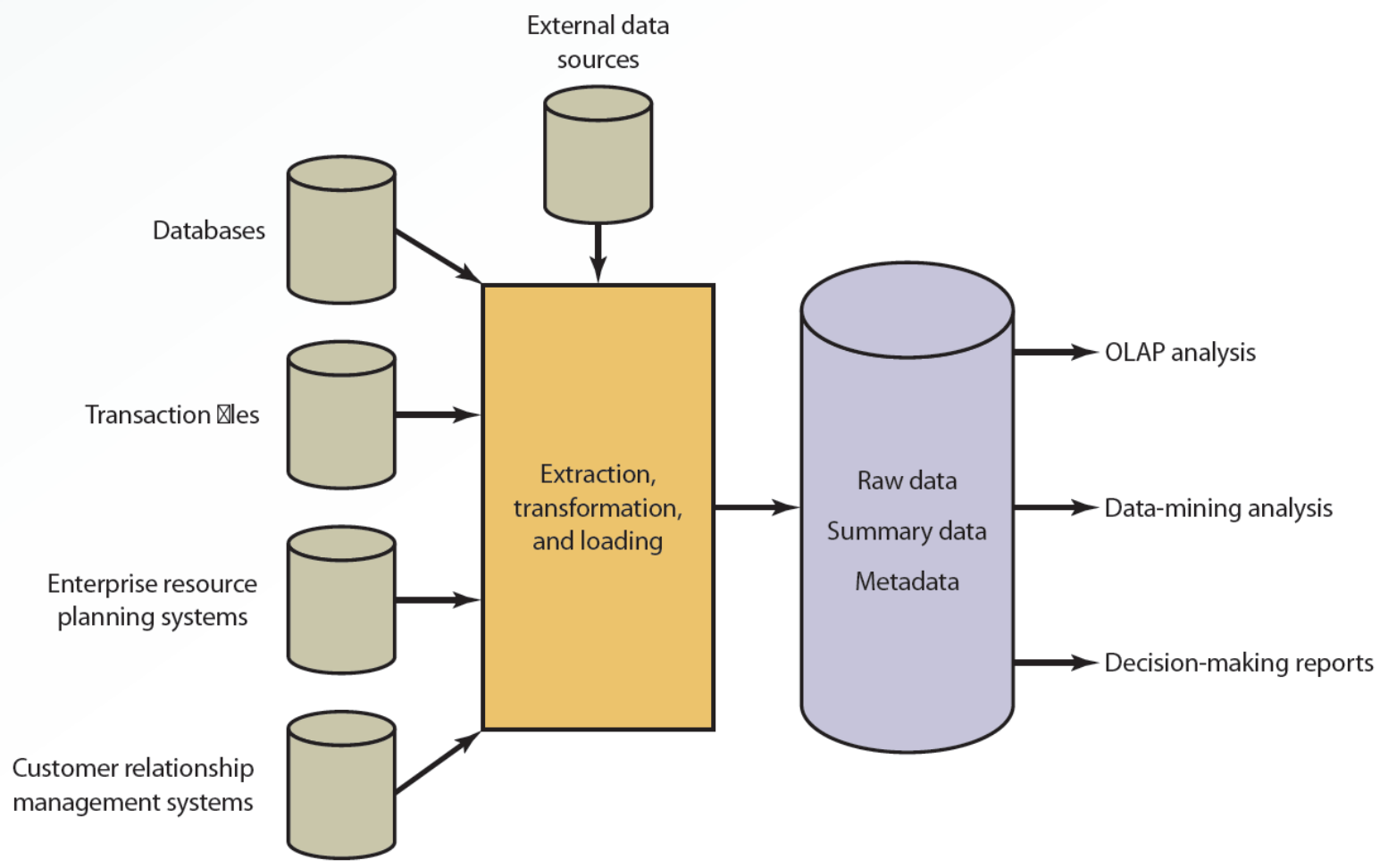
Data Warehouses (1 of 2)

- Collection of data from a variety of sources
  - Support decision-making applications
  - Generate business intelligence
- Called hypercubes because they store multidimensional data

Data Warehouses (2 of 2)

- Characteristics of data in a data warehouse
  - Subject oriented
  - Comes from a variety of sources
  - Categorized based on time
  - Captures aggregated data
  - Used for analytical purposes

Exhibit 3.6 A Data Warehouse Configuration



Input

- Different data sources provide the input for a data warehouse to perform analyses and generate reports
  - External data sources, databases, and transaction files
  - Enterprise resource planning (ERP) systems
  - Customer relationship management (CRM) systems

ETL

- Extraction, transformation, and loading
  - Processes used in a data warehouse
    - Collecting data from a variety of sources
    - Converting data into a format that can be used in transformation processing
    - Loading data into the data warehouse

Storage

- Collected information is organized in a data warehouse as:
  - Raw data: information in its original form
  - Summary data: gives users subtotals of various categories
  - Metadata: information about data's content, quality, condition, origin, and other characteristics

Output (1 of 3)

- Data warehouses use the following to generate reports:
  - Online analytical processing (OLAP)
    - Uses multiple sources of information and provides multidimensional analysis
    - Generates business intelligence
  - Data-mining analysis
    - Used to discover patterns and relationships

Output (2 of 3)

- Benefits of data warehouses
  - Cross-reference segments of an organization's operations for comparison
  - Generate complex queries and reports faster than when using databases
  - Generate reports efficiently using data from a variety of sources
  - Find patterns and trends that cannot be found with databases

Output (3 of 3)

- Analyze large amounts of historical data quickly
- Assist management in making well-informed business decisions
- Manage a high demand for information from many users with different needs and decision-making styles

Data Marts (1 of 2)

- Smaller version of a data warehouse, used by a single department or function
  - Advantages over data warehouses
    - Faster access to data owing to its smaller size
    - Improved response time for users
    - Easier to create because of its size and simplicity
    - Less expensive
    - Effective targeting of users

Data Marts (2 of 2)

- Disadvantages
  - Limited scope
  - Difficulty in consolidating information from different departments or functional areas

Business Analytics (1 of 3)

- Uses data and statistical methods
  - Gains insight into the data
  - Provides decision makers with information to act on
- Methods
  - Descriptive
  - Predictive
  - Prescriptive

Business Analytics (2 of 3)

- Descriptive analytics
  - Reactive strategy
  - Reviews past events, analyzes the data, and provides a report indicating:
    - What happened in a given period of time
    - How to prepare for the future

Business Analytics (3 of 3)

- Predictive analytics
  - Proactive strategy
  - Prepares decision makers for future events
- Prescriptive analytics
  - Recommends a course of action that decision makers should follow
  - Shows the likely outcome of each decision

The Big Data Era (1 of 3)

- Voluminous data
  - Conventional computing methods are unable to efficiently process and manage it
- Involves five dimensions
  - Volume
  - Variety
  - Velocity
  - Veracity
  - Value

The Big Data Era (2 of 3)

- Provides competitive advantage in many areas
  - Retail, financial services, advertising and public relations, government, manufacturing, healthcare, etc.
- Many technologies and applications have contributed to growth and popularity
  - Mobile and wireless technology, the popularity of social networks, etc.

The Big Data Era (3 of 3)

- Executives should guard against privacy risks
  - Discrimination, privacy breaches and embarrassments, unethical actions based on interpretations, loss of anonymity, etc.

Database Marketing (1 of 3)

- Uses an organization's database of customers and potential customers to promote products or services
  - Main goal: use information within the database to implement marketing strategies
    - Increase profits
    - Enhance competitiveness

Database Marketing (2 of 3)

- Transforms marketing from a reactive to a proactive process
  - Multivariate analysis
  - Data segmentation
  - Automated tools

Database Marketing (3 of 3)

- Tasks performed by successful database marketing campaigns
  - Calculating customer lifetime value (CLTV)
  - Conducting recency, frequency, and monetary analysis (RFM)
  - Using different techniques to communicate effectively with customers
  - Using different techniques to monitor customer behavior across a number of retail channels

Summary (1 of 2)

- All files are integrated in a database, which enables faster retrieval of data
- Components of a DBMS are database engine, data definition, data manipulation, application generation, and data administration
- Recent trends in database design are data-driven Web sites, natural language processing, and distributed and object-oriented databases

Summary (2 of 2)

- Data warehouse is a collection of data from a variety of sources
- Data marts focus on business functions for a specific user group in an organization
- Industries gain a competitive advantage from big data analytics