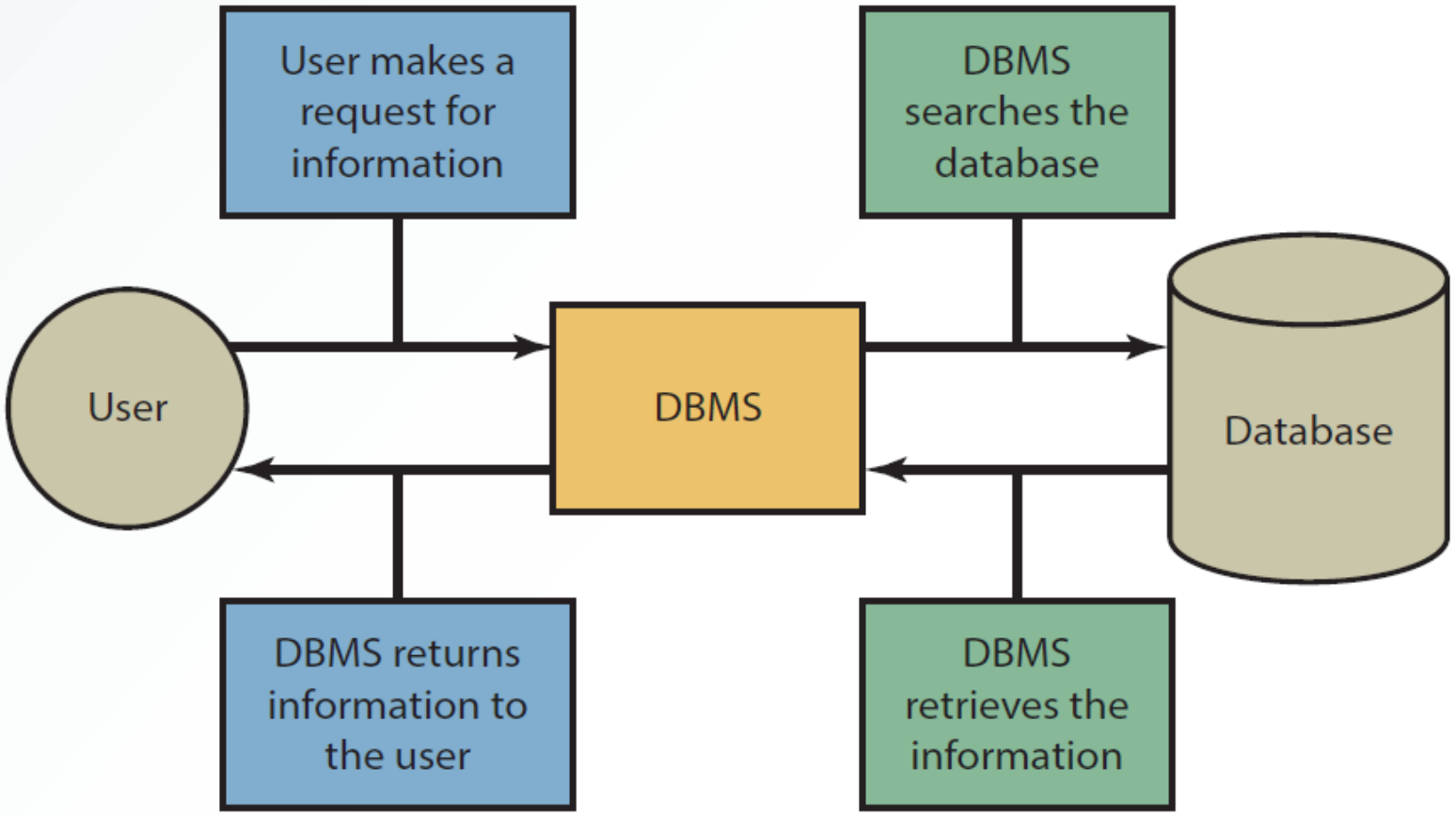
DEVELOPING DATABASES
&
COMPUTER FILES

Databases

- **Database**
  - *Collection of related data that is stored in a central location or in multiple locations*
- **Data hierarchy**: Structure and organization of data involving fields, records, and files
- **Database management system (DBMS)**
  - *Software for creating, storing, maintaining, and accessing database files*
  - *Makes using databases more efficient*

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*
  - *Organized based on a primary key*
    - Social Security numbers or account numbers
  - *Used for backup and archive files as they rarely need updating*

Types of Data in a Database

- Internal
  - *Collected from within an organization*
  - *Stored in the organization's internal databases*
- External
  - *Comes from a variety of resources*
  - *Stored in a data warehouse*

Methods for Accessing Files

- **Random access file structure**
  - *Records can be accessed in any order irrespective of the 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 tapes*

Methods for Accessing Files

- **Indexed sequential access method (ISAM)**
  - *Records are accessed sequentially or randomly depending on the number being accessed*
    - Random access is used for a small number
    - Sequential access is 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

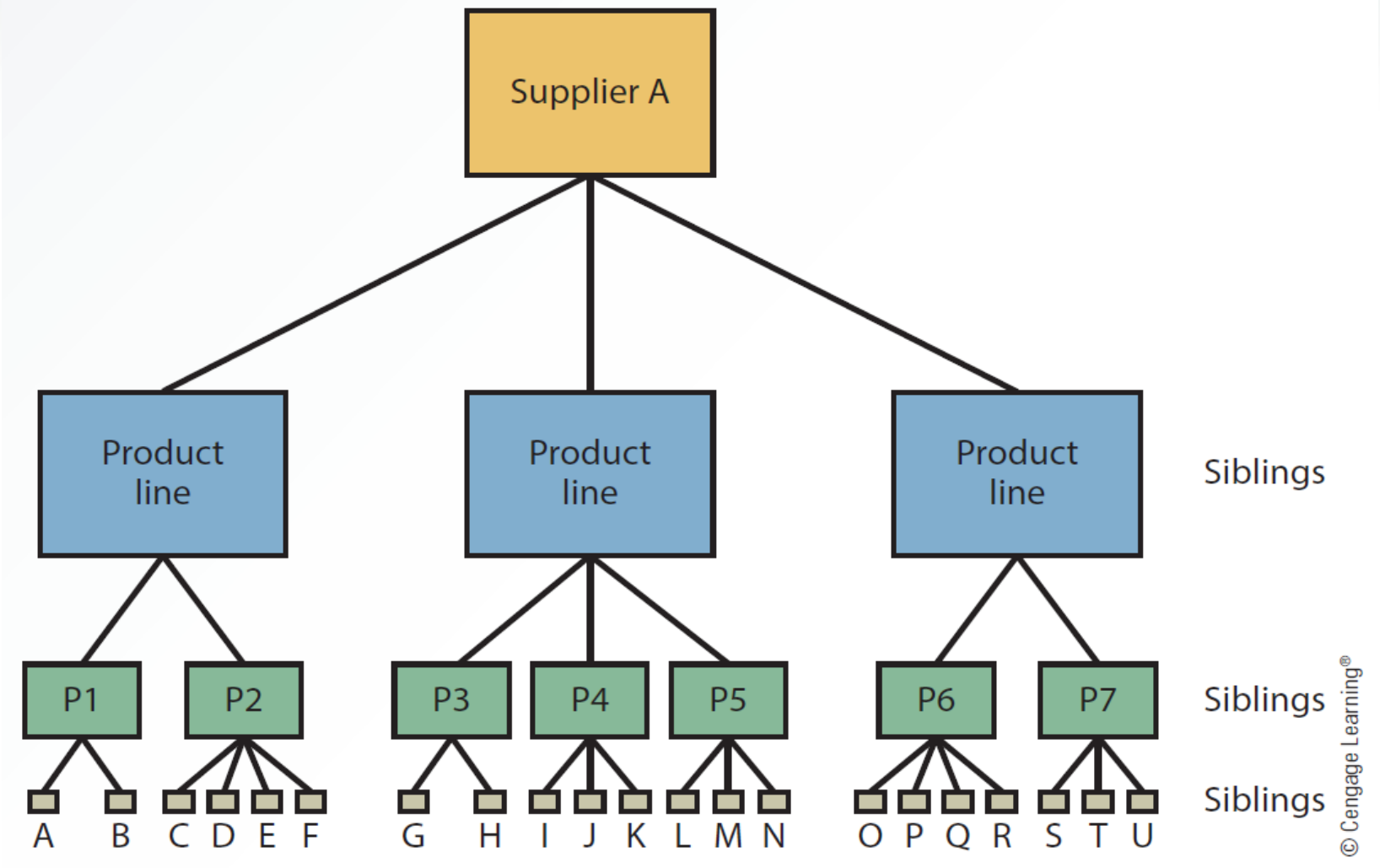
Logical Database Design

- **Physical view**
  - *Involves how data is stored on and retrieved from storage media*
    - Hard disks, magnetic tapes, or CDs
- **Logical view**
  - *Involves how information appears to users and how it can be organized and retrieved*
  - *Includes more than one logical view of data, depending on the user*

Logical Database Design

- **Data model**
  - *Determines how data is created, represented, organized, and maintained*
  - *Contains*
    - Data structure
    - Operations
    - Integrity rules
- **Hierarchical model**
  - *Relationships between records form a treelike structure*

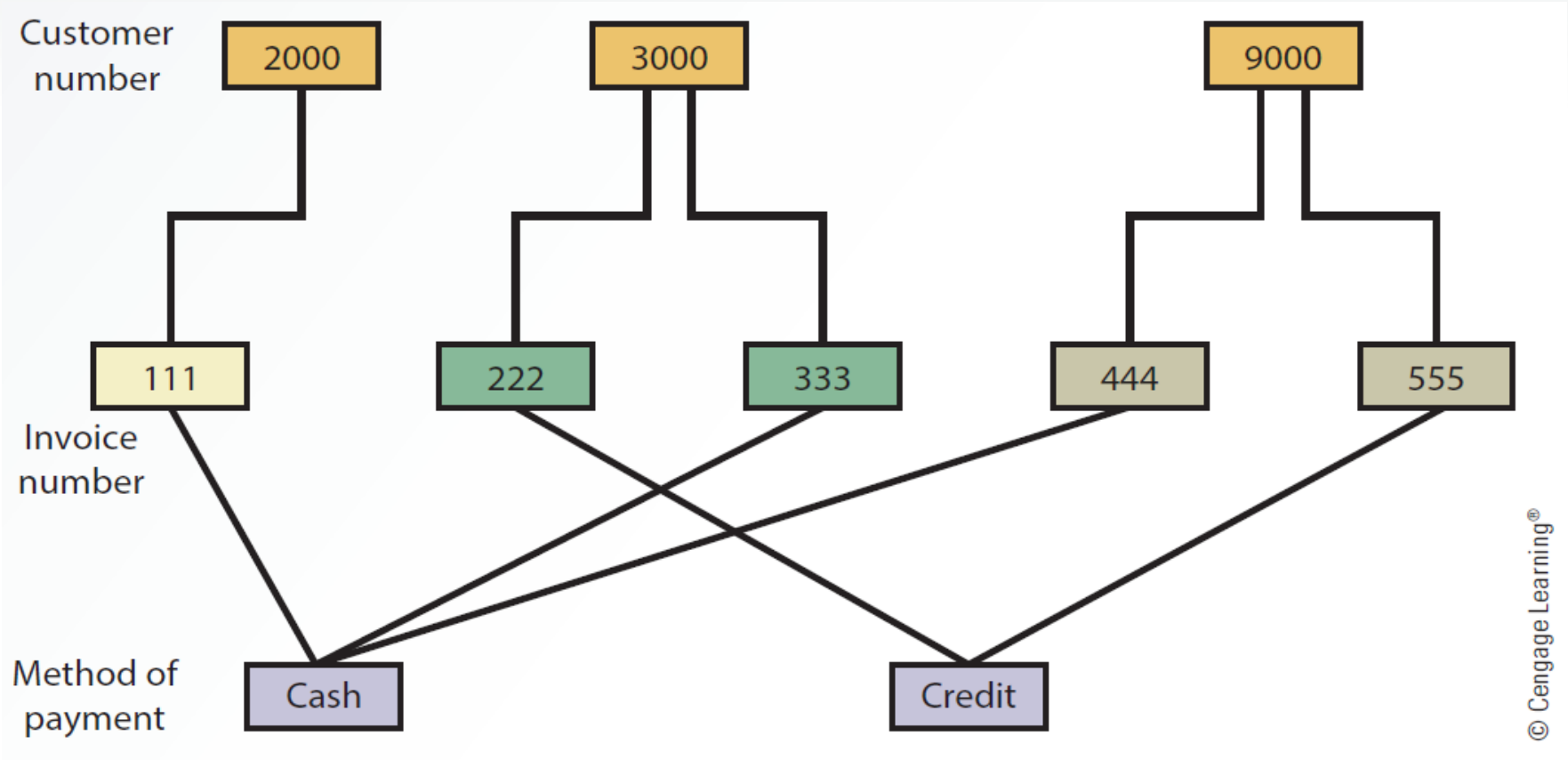
A Hierarchical Model



Logical Database Design

- **Network model**
  - *Similar to the hierarchical model but records are organized differently*
  - *Includes multiple parent and child records*

A Network Model



Relational Model

- Uses a two-dimensional table of rows and columns of data
  - *Rows are records*
  - *Columns are fields*
- **Data dictionary**: Stores definitions
  - *Data types for fields, default values, and validation rules for data in each field*

Relational Model

- **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*

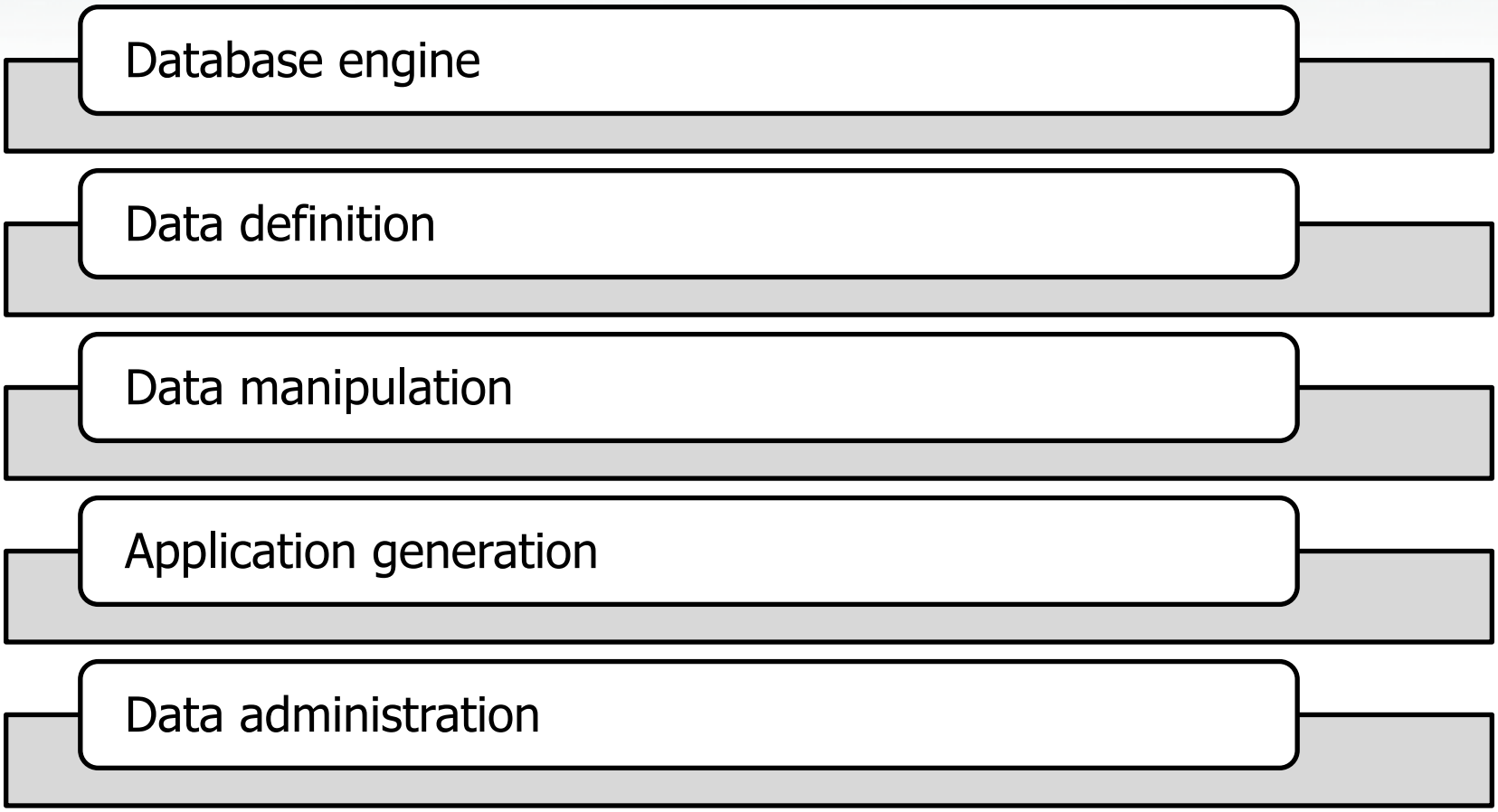
Relational Model

- **Normalization**
  - *Improves database efficiency by eliminating redundant data*
    - Ensures that only related data is stored in a table
  - *Goes through different stages from first normal form (1NF) to fifth normal form (5NF)*

Relational Model

- Retrieves data from tables using operations that pick and combine data from one or more tables
  - *Select*
  - *Project*
  - *Join*
  - *Intersection*
  - *Union*
  - *Difference*

Components of a DBMS



Database Engine

- Heart of DBMS software
- Responsible for data storage, manipulation, and retrieval
- Converts logical requests from users into their physical equivalents
  - *By interacting with other components of the DBMS*

Data Definition

- Creates and maintains the data dictionary
- Defines the structure of files in a database
- Makes changes to a database's structure
  - *Adding and deleting fields*
  - *Changing field size and data type*

Data Manipulation

- Used to add, delete, modify, and retrieve records from a database
- Uses a query language
  - ***Structured Query Language (SQL)***
    - Standard fourth-generation query language that consists of several keywords specifying actions to take
  - ***Query by example (QBE)***
    - Involves requesting data from a database by constructing a statement formed by query forms

Application Generation

- Designs elements of an application using a database
  - *Data entry screens*
  - *Interactive menus*
  - *Interfaces with other programming languages*
- Used by IT professionals and database administrators

Data Administration

- Used for the tasks 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)***

Data Administration

- **Database administrator (DBA)**
  - *Handles database design and management*
    - Setting up database
    - Establishing security measures to determine users' access rights
    - Developing recovery procedures when data is lost or corrupted
    - Evaluating database performance
    - Adding and fine-tuning database functions

Recent Trends in Database Design and Use

- **Data-driven website**
  - *Interface to a database*
  - *Retrieves data and allows users to enter data*
  - *Improves access to information*
  - *Gives users more current information from a variety of data sources*

Recent Trends in Database Design and Use

- **Distributed database**: Stores data on multiple servers throughout an organization
- 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, with each site storing the data used most often*

Recent Trends in Database Design and Use

- **Object-oriented database**: Single object contains data and their relationships
  - *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 single unit*
  - ***Inheritance**: New objects can be created faster and easily by entering new data in attributes*

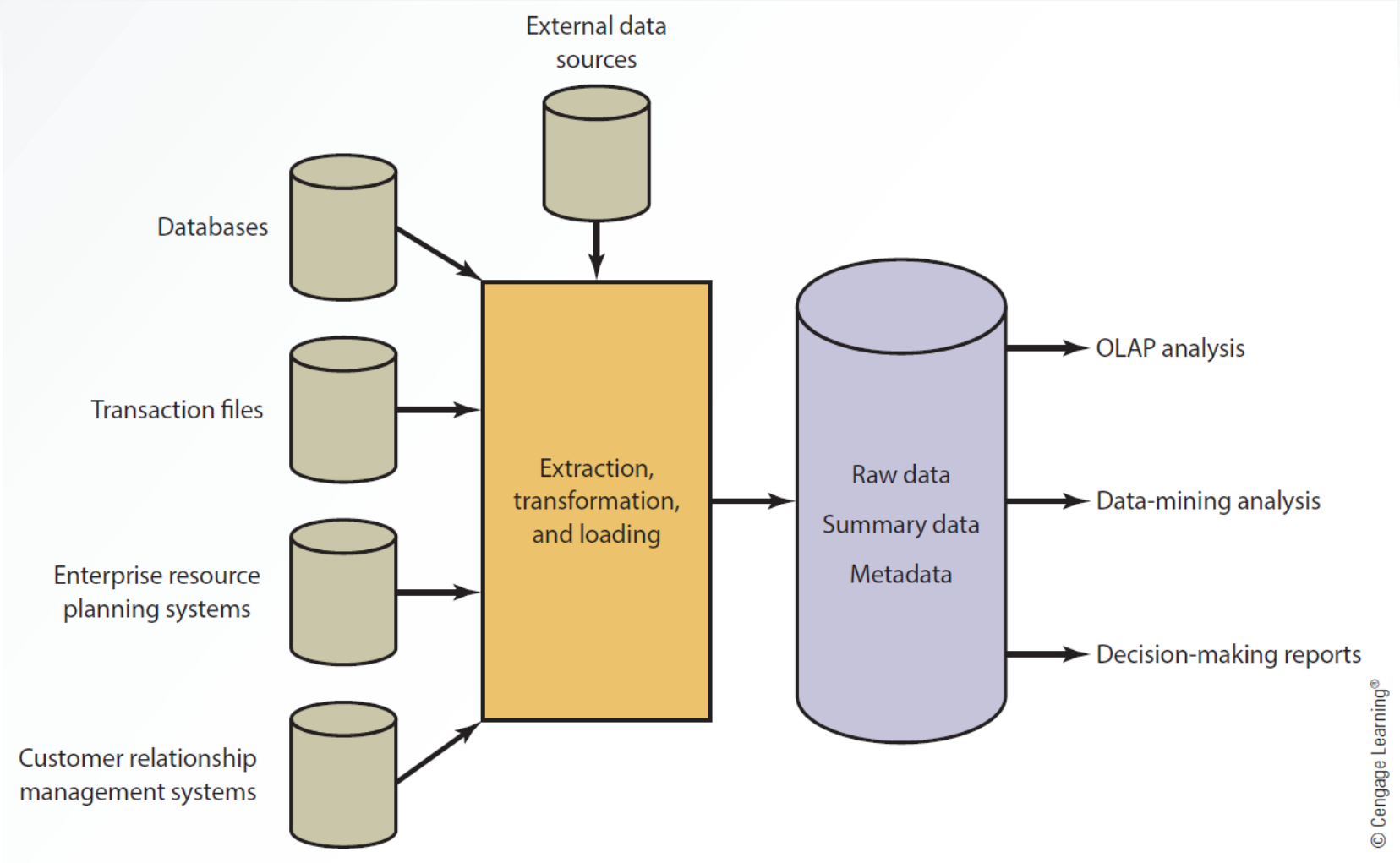
Data Warehouses

- Collection of data from a variety of sources
  - *Used to support decision-making applications and generate business intelligence*
  - *As they store multidimensional data, they are called hypercubes*

Characteristics of Data in a Data Warehouse

- 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 sources of data together provide input for a data warehouse to perform analyses and generate reports
  - *External data sources*
  - *Databases*
  - *Transaction files*
  - *Enterprise resource planning (ERP) systems*
  - *Customer relationship management (CRM) systems*

Extraction, Transformation, and Loading (ETL)

- Processes used in a data warehouse
  - *Extracting data from outside sources*
  - *Transforming data to fit operational needs*
  - *Loading data into the database or data warehouse*

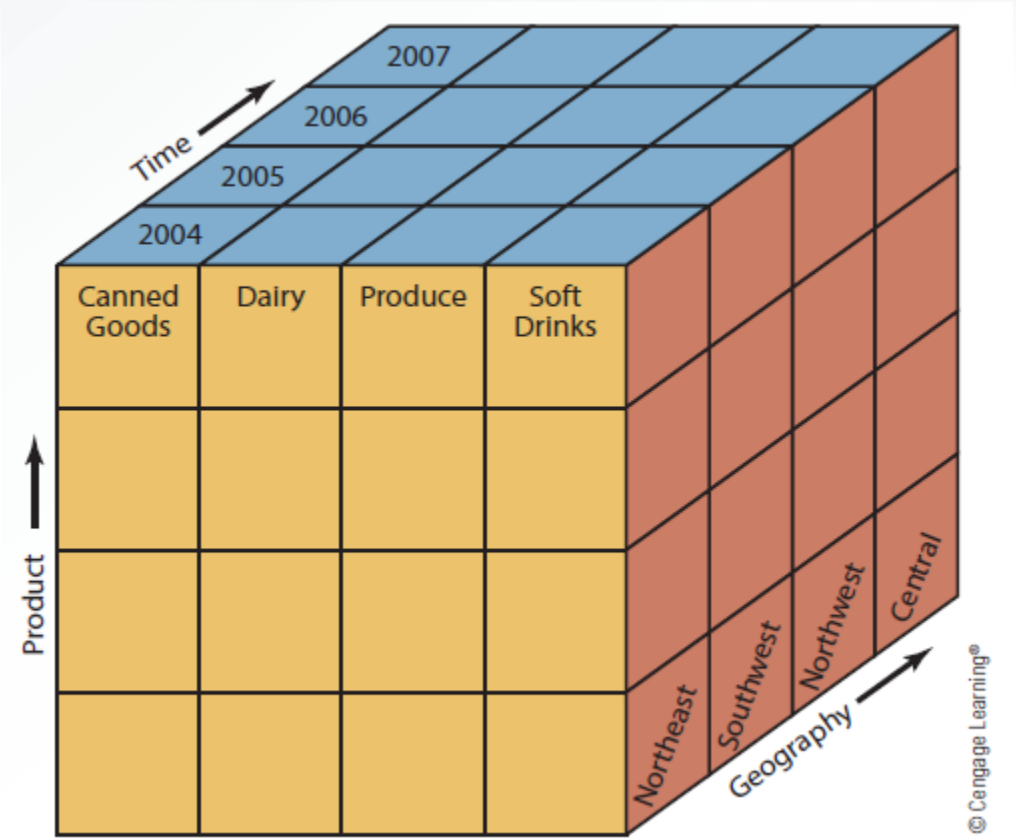
Storage

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

Output

- **Online transaction processing (OLTP)**
  - *Facilitates and manages transaction-oriented applications*
  - *Uses internal data and responds in real time*
- **Online analytical processing (OLAP)**
  - *Generates business intelligence*
  - *Uses multiple sources of information and provides multidimensional analysis*
    - Viewing data based on time, product, and location

Exhibit 3.7 Slicing and Dicing Data



Output

- **Data-mining analysis**: Discovers patterns and relationships
- Data warehouses help generate various types of information and reports for decision making
  - *Cross-reference segments of an organization's operations for comparison purposes*
  - *Generate complex queries and reports faster and easier*
  - *Generate reports efficiently using data from a variety of sources*

Output

- *Find patterns and trends that can't be found with databases*
- *Analyze large amounts of historical data quickly*
- *Assist management in making well-informed business decisions*
- *Manage high demand information from many users with different needs and decision making styles*

Data Mart

- Smaller version of data warehouse, used by single department or function
- Advantages over data warehouses
  - *Access to data is faster due to their smaller size*
  - *Response time for users is improved*
  - *Easy to create because they are smaller and simple*
  - *Less expensive*
  - *Users are targeted better*
- Has limited scope

Business Analytics (BA)

- Uses data and statistical methods to gain insight into the data
- Provides decision makers with information to act on

Types of BA Methods

- Descriptive analytics
  - *Reviews past events*
  - *Analyzes the data*
  - *Provides a report indicating what happened over a given period of time and how to prepare for future*
  - *Reactive strategy*
- Predictive analytics
  - *Prepares decision maker for future events*
  - *Proactive strategy*

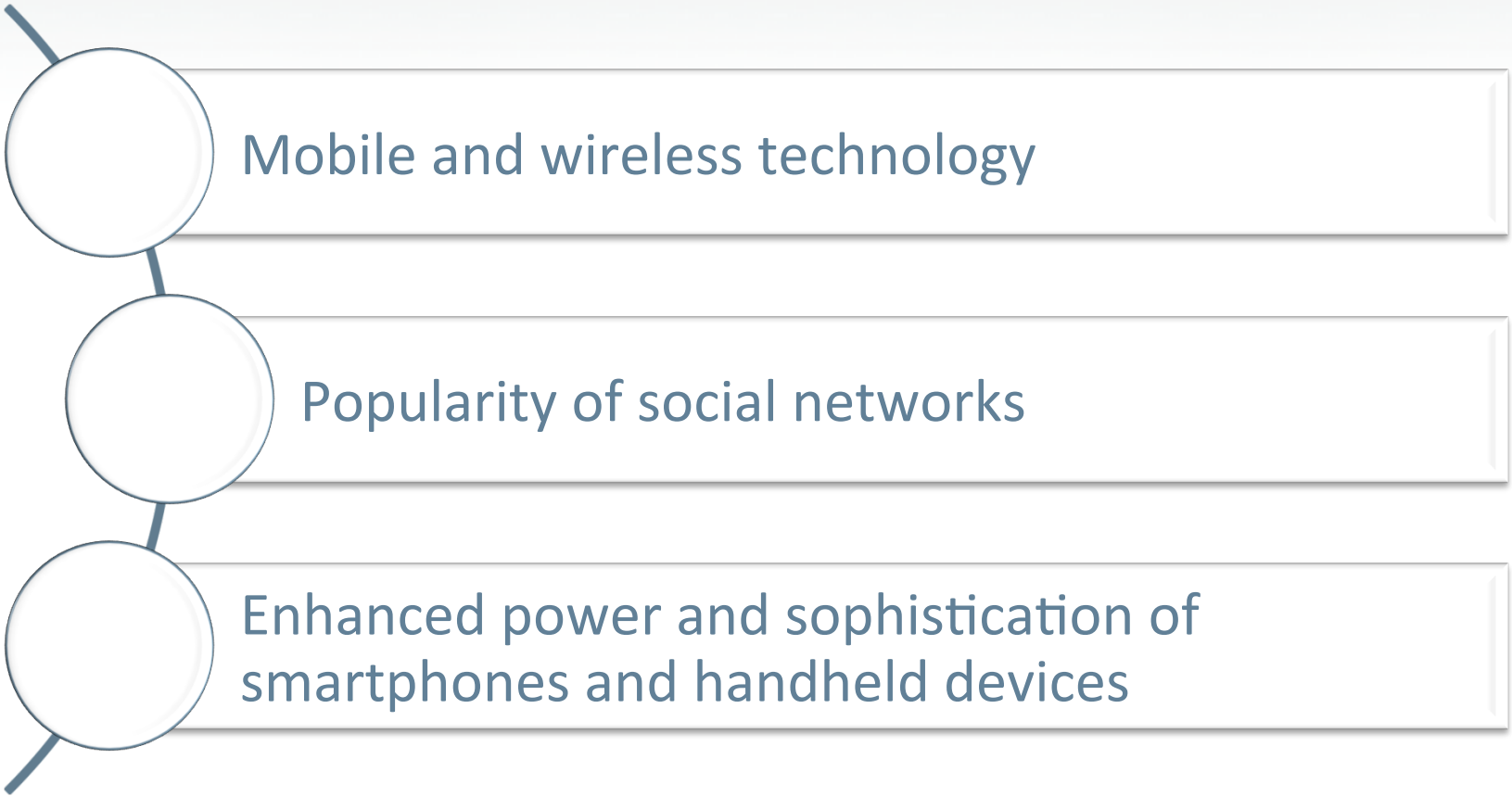
Big Data Era

- **Big data**: Voluminous data which the conventional computing methods are unable to efficiently process and manage
  - *Involves dimensions known as 3 Vs*
    - Volume: Quantity of transactions
    - Variety: Combination of structured and unstructured data
    - Velocity: Speed with which data needs to be gathered and processed

Who Benefits from Big Data?

- Industries benefit and gain a competitive advantage in areas like:
  - *Retail*
  - *Financial services*
  - *Advertising and public relations*
  - *Government*
  - *Manufacturing*
  - *Media and telecommunications*
  - *Energy*
  - *Healthcare*

Factors in the Growth and Popularity of Big Data



KEY TERMS

- Allocation
- Big data
- Business analytics
- Create, read, update, and delete (CRUD)
- Data dictionary
- Data hierarchy
- Data mart
- Data model
- Data warehouse
- Database

KEY TERMS

- Database administrator (DBA)
- Database management system (DBMS)
- Data-driven website
- Data-mining analysis
- Distributed database management system (DDBMS)
- Encapsulation
- Extraction, transformation, and loading (ETL)
- Foreign key
- Fragmentation

KEY TERMS

- Hierarchical model
- Indexed sequential access method (ISAM)
- Inheritance
- Logical view
- Network model
- Normalization
- Object-oriented databases
- Online analytical processing (OLAP)
- Online transaction processing (OLTP)
- Physical view

KEY TERMS

- Primary key
- Query by example (QBE)
- Random access file structure
- Relational model
- Replication
- Sequential access file structure
- Structured Query Language (SQL)