ITIS 313
Data and Information Management

Spring 2021

Instructor

- Dr. Mohammed Ali Eltaher

- Ph.D., in Computer Science & Engineering
  - Social Media Data Mining
- M.S., in Intelligent System
- B.S., in Computer Science

- Email: bineltaher@gmail.com

Course objectives

- Understand the role of databases and database management systems in managing organizational data and information.
- Use at least one conceptual data modeling technique (such as entity-relationship modeling) to capture the information requirements for an enterprise domain.
- Understand the purpose and principles of normalizing a relational database structure.
- Use the data definition, data manipulation, and data control language components of SQL in the context of one widely used implementation of the language.

Course objectives

- Understand the concept of database transaction and apply it appropriately to an application context.
- Understand the role of databases and database management systems in the context of enterprise systems.
- Understand the key principles of data security and identify data security risk and violations in data management system design.
- Understand the core concepts of data quality and their application in an organizational context.
- Understand how structured, semi-structured, and unstructured data are all essential elements of enterprise information and knowledge management.

Course Content

Week	*Knowledge Unit*	*Topics Covered*
1	Introduction	Database approach , Types of database management systems, Basic file processing concepts ,Physical data storage concepts , and File organizations techniques issues
2, 3	Data Model	Conceptual data model , Logical data model , and Physical data model
4,5	Database languages and administration	SQL: DDL, DML, and DCL Data and database administration
6,7	DBMS	Using a DBMS from an application development environment Use of DBMS in an enterprise system context
8, 9	Data security management	Basic data security principles Data security implementation
10	Data quality management	Data quality principles , Data quality audits, and Data quality improvement
11,12	Business intelligence	On-line analytic processing , Data warehousing, and Data mining

Work and Grading

- Homework assignments
- Midterms, Final
- Labs
- Participation can help on margins
- Academic honesty policy

11e

Database Systems
Design, Implementation, and Management

Coronel | Morris

Data vs. Information

Data

- Raw facts
  - Raw data - Not yet been processed to reveal the meaning
- Building blocks of information
- **Data management**
  - Generation, storage, and retrieval of data

Information

- Produced by processing data
- Reveals the meaning of data
- Enables **knowledge** creation
- Should be accurate, relevant, and timely to enable good decision making

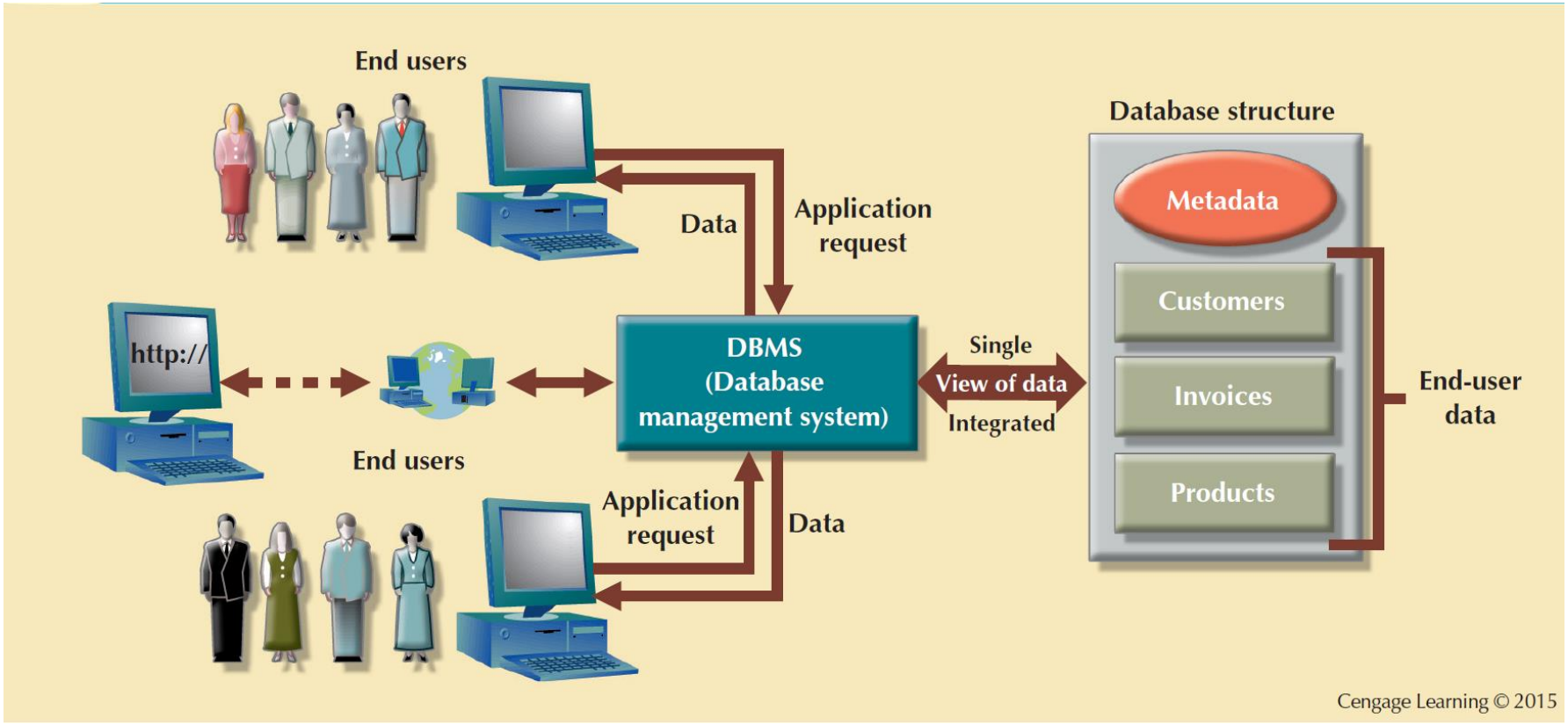
Database

- Shared, integrated computer structure that stores a collection of:
  - End-user data - Raw facts of interest to end user
  - **Metadata**: Data about data, which the end-user data are integrated and managed
    - Describe data characteristics and relationships
- **Database management system (DBMS)**
  - Collection of programs
  - Manages the database structure
  - Controls access to data stored in the database

Role of the DBMS

- Intermediary between the user and the database
- Enables data to be shared
- Presents the end user with an integrated view of the data
- Receives and translates application requests into operations required to fulfill the requests
- Hides database's internal complexity from the application programs and users

Figure 1.2 - The DBMS Manages the Interaction between the End User and the Database



Advantages of the DBMS

- Better data integration and less data inconsistency
  - **Data inconsistency**: Different versions of the same data appear in different places
- Increased end-user productivity
- Improved:
  - Data sharing
  - Data security
  - Data access
  - Decision making
    - **Data quality**: Promoting accuracy, validity, and timeliness of data

Types of Databases

- **Single-user database**: Supports one user at a time
  - **Desktop database**: Runs on PC
- **Multiuser database**: Supports multiple users at the same time
  - **Workgroup databases**: Supports a small number of users or a specific department
  - **Enterprise database**: Supports many users across many departments

Types of Databases

- **Centralized database**: Data is located at a single site
- **Distributed database**: Data is distributed across different sites
- **Cloud database**: Created and maintained using cloud data services that provide defined performance measures for the database

Types of Databases

- **General-purpose databases**: Contains a wide variety of data used in multiple disciplines
- **Discipline-specific databases**: Contains data focused on specific subject areas

Types of Databases

- **Operational database**: Designed to support a company's day-to-day operations
- **Analytical database**: Stores historical data and business metrics used exclusively for tactical or strategic decision making
  - **Data warehouse**: Stores data in a format optimized for decision support

Types of Databases

- **Online analytical processing (OLAP)**
  - Enable retrieving, processing, and modeling data from the data warehouse
- **Business intelligence**: Captures and processes business data to generate information that support decision making

Types of Databases

- **Unstructured data**: It exists in their original state
- **Structured data**: It results from formatting
  - Structure is applied based on type of processing to be performed
- **Semistructured data**: Processed to some extent
- **Extensible Markup Language (XML)**
  - Represents data elements in textual format

Database Design

- Focuses on the design of the database structure that will be used to store and manage end-user data
- Well-designed database
  - Facilitates data management
  - Generates accurate and valuable information
- Poorly designed database causes difficult-to-trace errors

Evolution of File System Data Processing

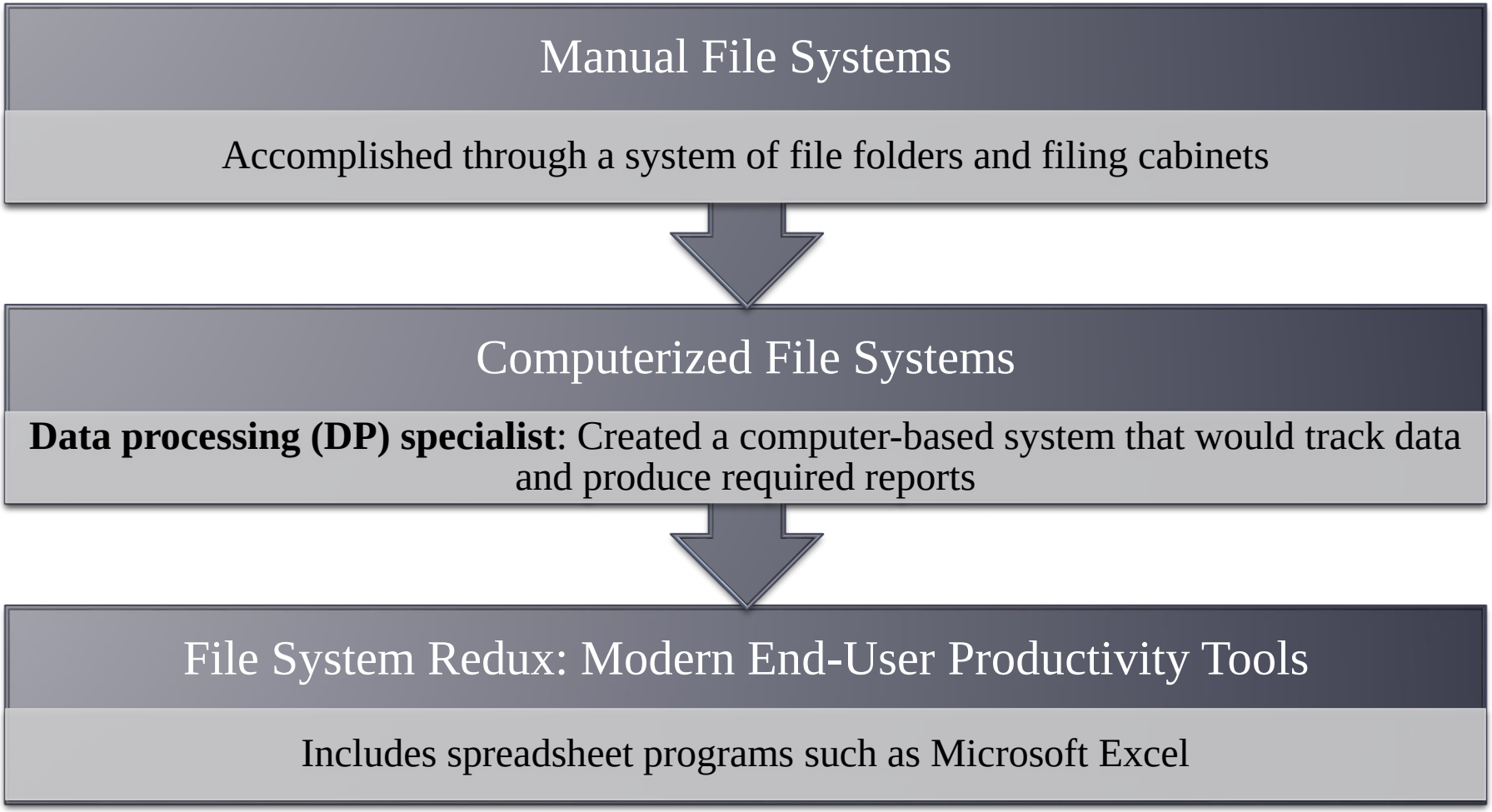
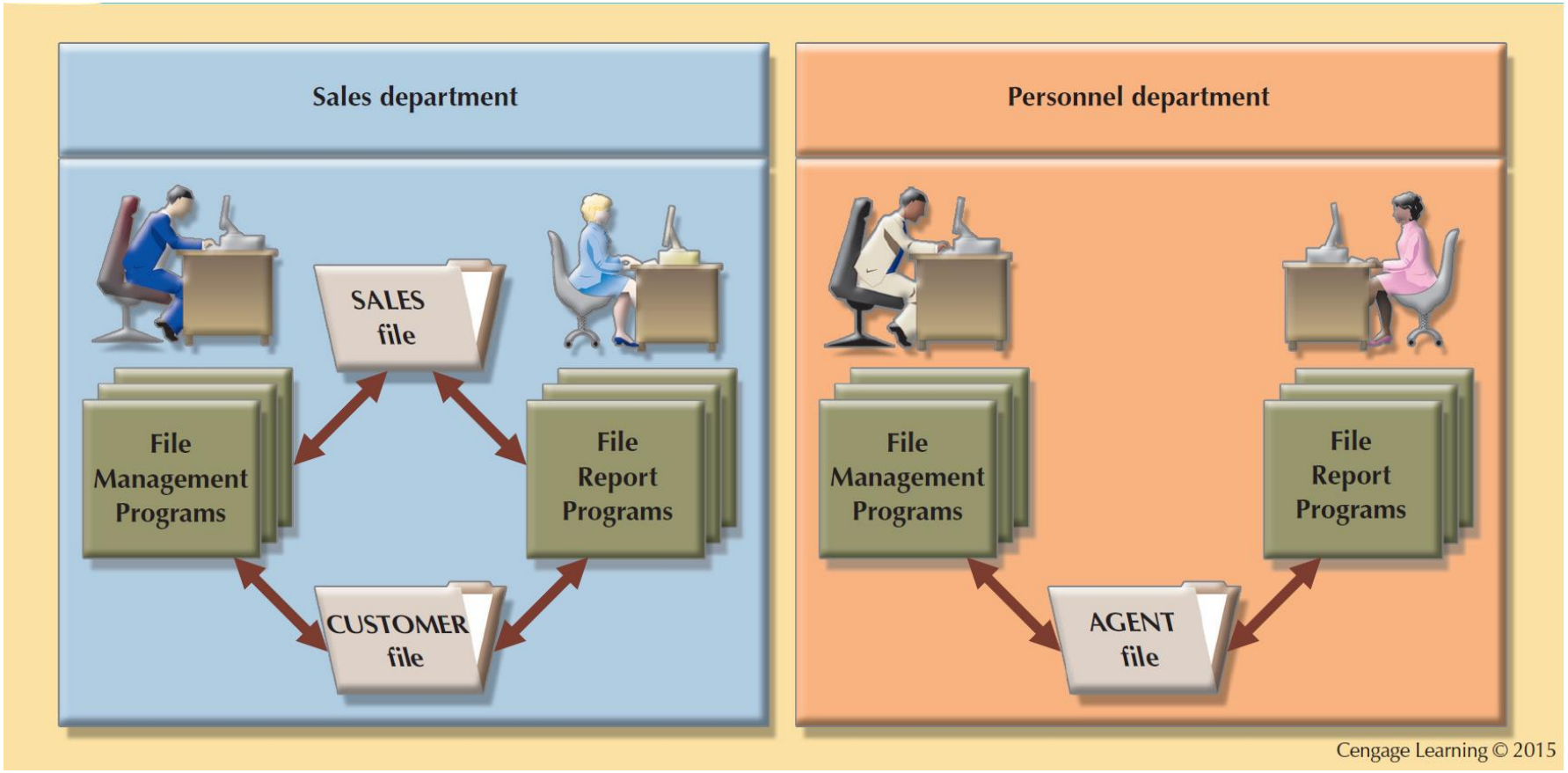


Table 1.2 - Basic File Terminology

TERM	DEFINITION
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data have little meaning unless they have been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Cengage Learning © 2015

Figure1.6 - A Simple File System



Problems with File System Data Processing

- Lengthy development times
- Difficulty of getting quick answers
- Complex system administration
- Lack of security and limited data sharing
- Extensive programming

Structural and Data Dependence

- **Structural dependence**: Access to a file is dependent on its own structure
  - All file system programs are modified to conform to a new file structure
- **Structural independence**: File structure is changed without affecting the application's ability to access the data

Structural and Data Dependence

- Data dependence
  - Data access changes when data storage characteristics change
- Data independence
  - Data storage characteristics is changed without affecting the program's ability to access the data
- Practical significance of data dependence is difference between logical and physical format

Data Redundancy

- Unnecessarily storing same data at different places
- **Islands of information**: Scattered data locations
  - Increases the probability of having different versions of the same data

Data Redundancy Implications

- Poor data security
- Data inconsistency
- Increased likelihood of data-entry errors when complex entries are made in different files
- **Data anomaly**: Develops when not all of the required changes in the redundant data are made successfully

Types of Data Anomaly

- Update Anomalies
- Insertion Anomalies
- Deletion Anomalies

Lack of Design and Data-Modeling Skills

- Evident despite the availability of multiple personal productivity tools being available
- Data-modeling skills is vital in the data design process
- Good data modeling facilitates communication between the designer, user, and the developer

Database Systems

- Logically related data stored in a single logical data repository
  - Physically distributed among multiple storage facilities
- DBMS eliminates most of file system's problems
- Current generation DBMS software:
  - Stores data structures, relationships between structures, and access paths
  - Defines, stores, and manages all access paths and components

Figure 1.8 - Contrasting Database and File Systems

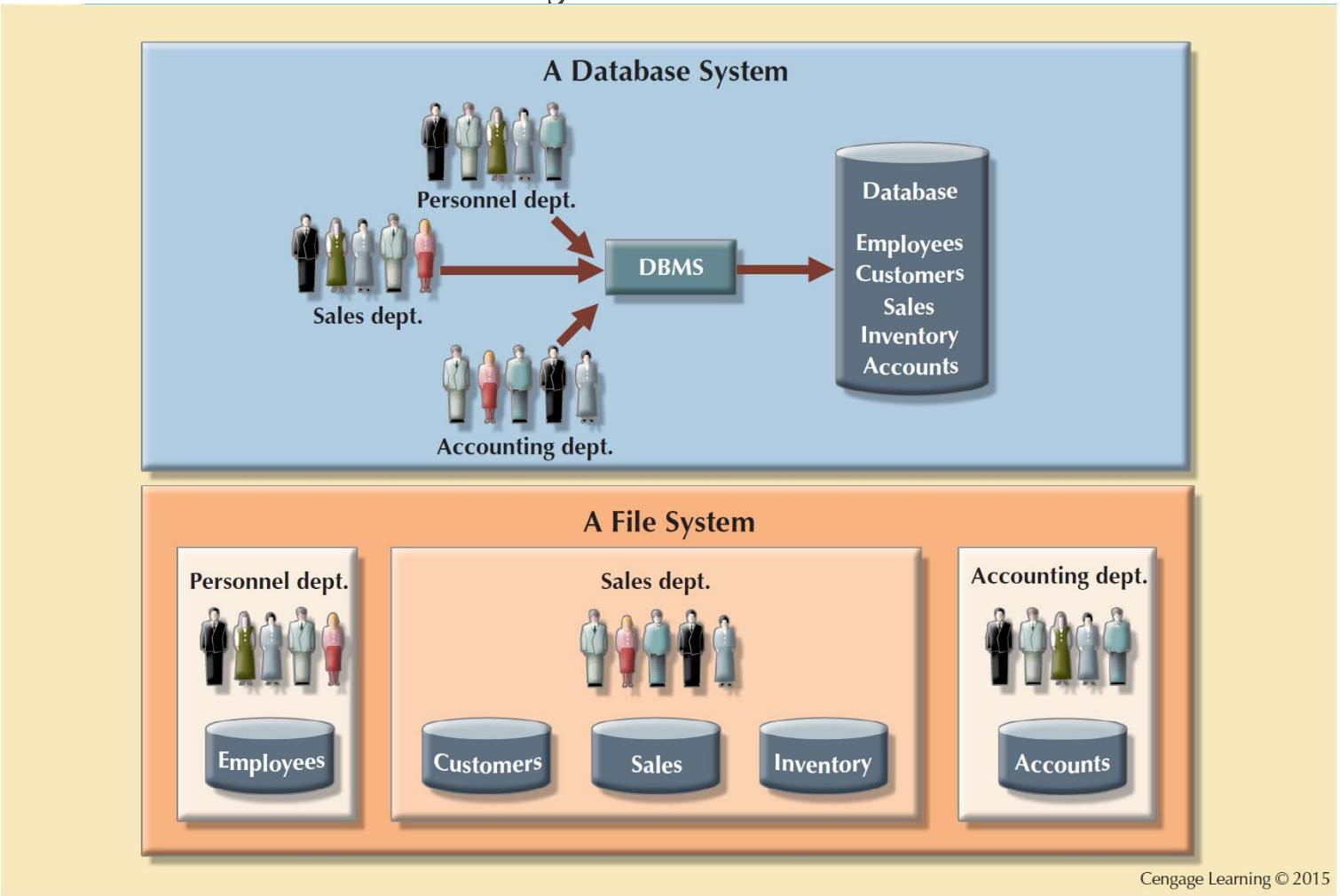
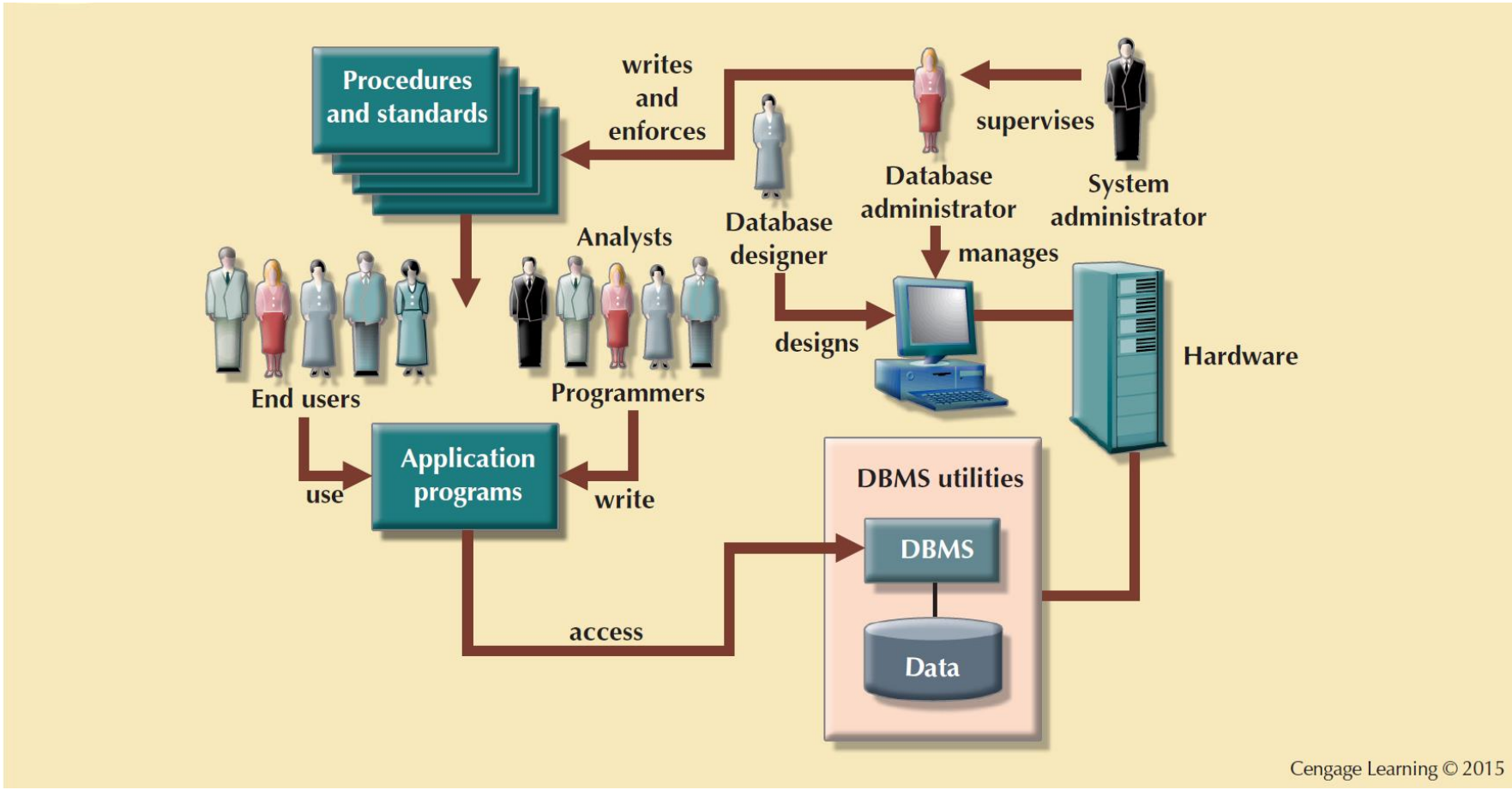


Figure 1.9 - The Database System Environment



DBMS Functions

Data dictionary management

- **Data dictionary**: Stores definitions of the data elements and their relationships

Data storage management

- **Performance tuning**: Ensures efficient performance of the database in terms of storage and access speed

Data transformation and presentation

- Transforms entered data to conform to required data structures

Security management

- Enforces user security and data privacy

DBMS Functions

Multiuser access control

- Sophisticated algorithms ensure that multiple users can access the database concurrently without compromising its integrity

Backup and recovery management

- Enables recovery of the database after a failure

Data integrity management

- Minimizes redundancy and maximizes consistency

DBMS Functions

Database access languages and application programming interfaces

- **Query language**: Lets the user specify what must be done without having to specify how
- **Structured Query Language (SQL)**: De facto query language and data access standard supported by the majority of DBMS vendors

Database communication interfaces

- Accept end-user requests via multiple, different network environments

Disadvantages of Database Systems

- Increased costs
- Management complexity
- Maintaining currency
- Vendor dependence
- Frequent upgrade/replacement cycles