INFORMATION SYSTEM DEPARTMENT
INFORMATION TECHNOLOGY FACULTY – HCM
UNIVERSITY OF SCIENCE

ADVANCED DATABASE

Chapter 01

# OVERVIEW OF DATABASE DESIGN



KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

fit@hcmus

cdio 4.0

fit@hcmus

# Content

- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities

# Content

- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities

# Why databases?

FIGURE 1.1 THE PERVASIVE NATURE OF DATABASES

## A Day In Susan's Life

See how many databases she interacts with each day

| *Before leaving for work, Susan checks her Facebook and Twitter accounts* | *On her lunch break, she picks up her prescription at the pharmacy* | *After work, Susan goes to the grocery store* | *At night, she plans for a trip and buys airline tickets and hotel reservations online* | *Then she makes a few online purchases* |
|---|---|---|---|---|
|  |  |  |  |   |
| Where is the data about the friends and groups stored?<br>Where are the "likes" stored and what would they be used for? | Where is the pharmacy inventory data stored?<br>What data about each product will be in the inventory data?<br>What data is kept about each customer and where is it stored? | Where is the product data stored?<br>Is the product quantity in stock updated at checkout?<br>Does she pay with a credit card? | Where does the online travel website get the airline and hotel data from?<br>What customer data would be kept by the website?<br>Where would the customer data be stored? | Where are the product and stock data stored?<br>Where does the system get the data to generate product "recommendations" to the customer?<br>Where would credit card information be stored? |
|   |   |   |   |   |

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Why databases?

- **Data**
  - Generated through our daily life
  - Essential for organizations to survive and prosper
- **Characteristics**
  - Ubiquitous
  - Pervasive
  - Enormous

➔ How to process, manipulate, and manage this large amount of data efficiently to support organizations?

➔ **Database Systems**



(Coronel and Morris, 2015) (Hoffer et al., 2019)

cdio 4.0
fit@hcmus

# Data vs. Information

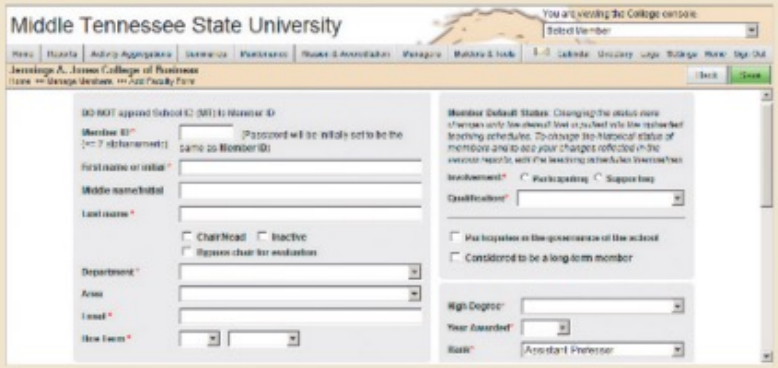
- **Data**
  - Raw facts
  - Not yet processed to reveal meaning to the end user
  - Building blocks of information
- **Information**
  - Results from processing raw data to reveal meaning
  - Requires context
  - Bedrock of knowledge
  - Should be accurate, relevant, and timely

(Coronel and Morris, 2015) (Hoffer et al., 2019)

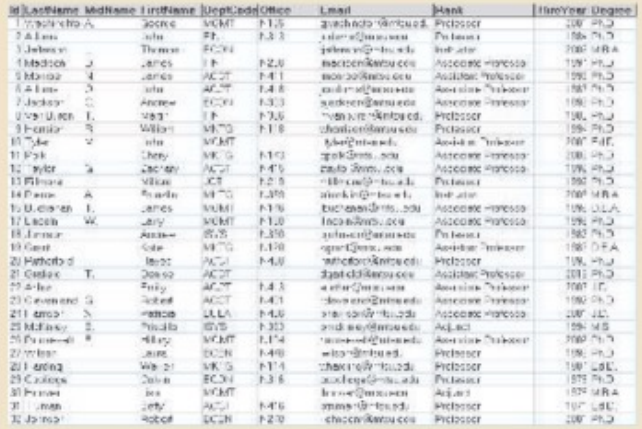
# Data vs. Information

## FIGURE 1.2 TRANSFORMING RAW DATA INTO INFORMATION

a) Data entry screen



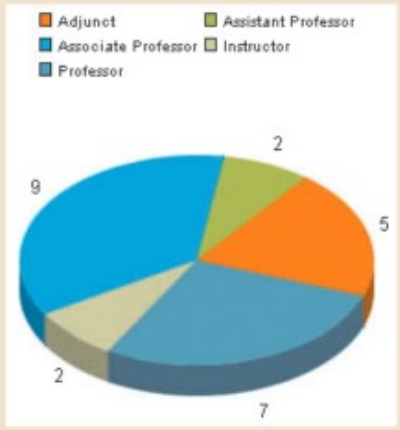

b) Raw data



c) Information in summary format

| Rank | COUNT | %/INFS | TOT/COL | %/COL. TOT. | %/COL. FAC. |
|---|---|---|---|---|---|
| Adjunct | 5 | 20.00% | 23 | 21.74% | 3.27% |
| Assistant Professor | 2 | 8.00% | 28 | 7.14% | 1.31% |
| Associate Professor | 9 | 36.00% | 37 | 24.32% | 5.88% |
| Instructor | 2 | 8.00% | 18 | 11.11% | 1.31% |
| Professor | 7 | 28.00% | 47 | 14.89% | 4.58% |

d) Information in graphical format




**Context helps users understand data (Meta-data)**



(Coronel and Morris, 2015) (Hoffer et al., 2019)

cdio 4.0

fit@hcmus

# Data vs. Information

**Context helps users understand data (Meta-data)**

**TABLE 1-1 Example Metadata for Class Roster**

| Data Item | | | Metadata | | | |
|---|---|---|---|---|---|---|
| Name | Type | Length | Min | Max | Description | Source |
| Course | Alphanumeric | 30 | | | Course ID and name | Academic Unit |
| Section | Integer | 1 | 1 | 9 | Section number | Registrar |
| Semester | Alphanumeric | 10 | | | Semester and year | Registrar |
| Name | Alphanumeric | 30 | | | Student name | Student IS |
| ID | Integer | 9 | | | Student ID (SSN) | Student IS |
| Major | Alphanumeric | 4 | | | Student major | Student IS |
| GPA | Decimal | 3 | 0.0 | 4.0 | Student grade point average | Academic Unit |

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Definitions

- **Data**: stored representations of meaningful objects and events.
  - Structured: numbers, text, dates
  - Unstructured: images, video, documents
- **Information**: data processed to increase knowledge in the person using the data.
- **Database**: organized collection of logically related data.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

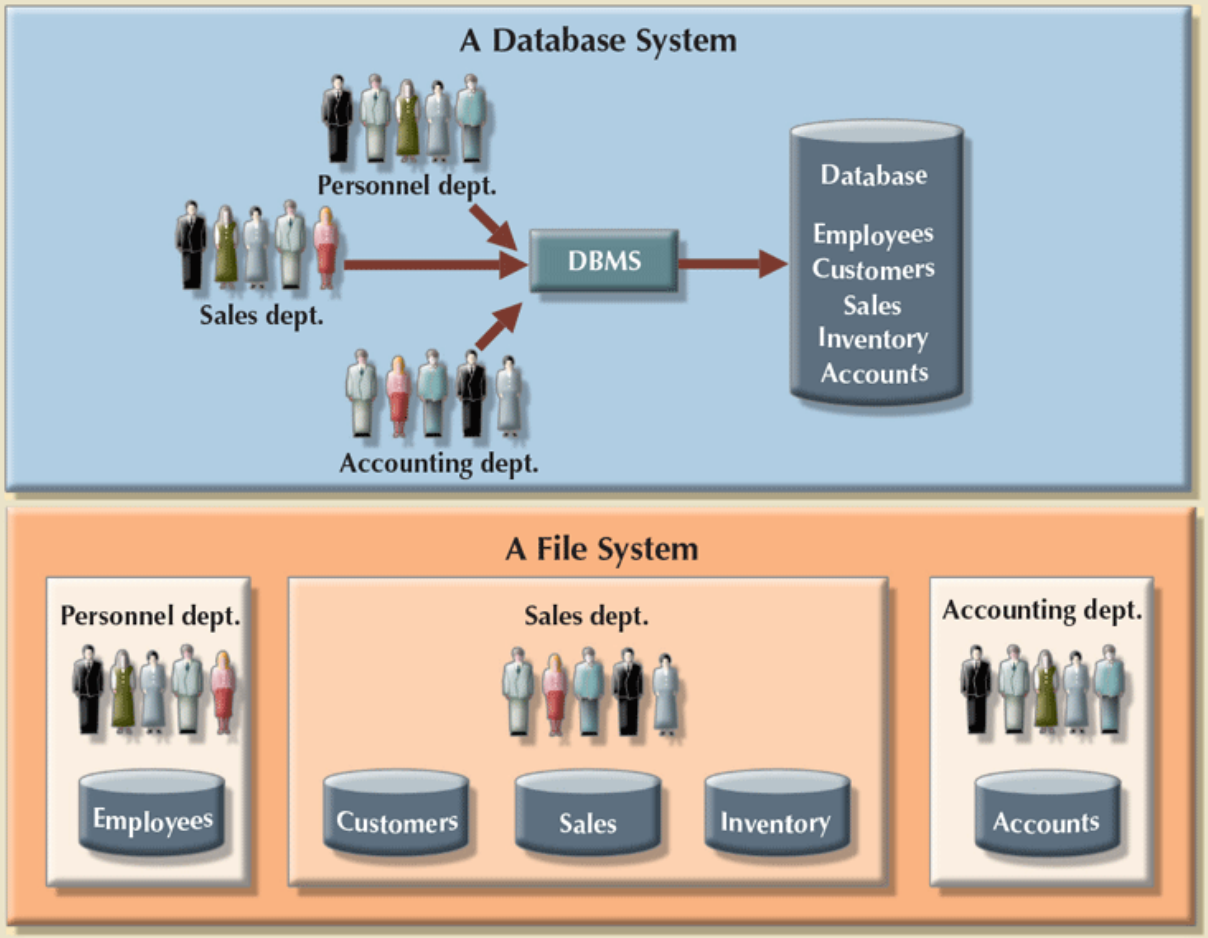
# Content

- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities

(Coronel and Morris, 2015) (Hoffer et al., 2019)

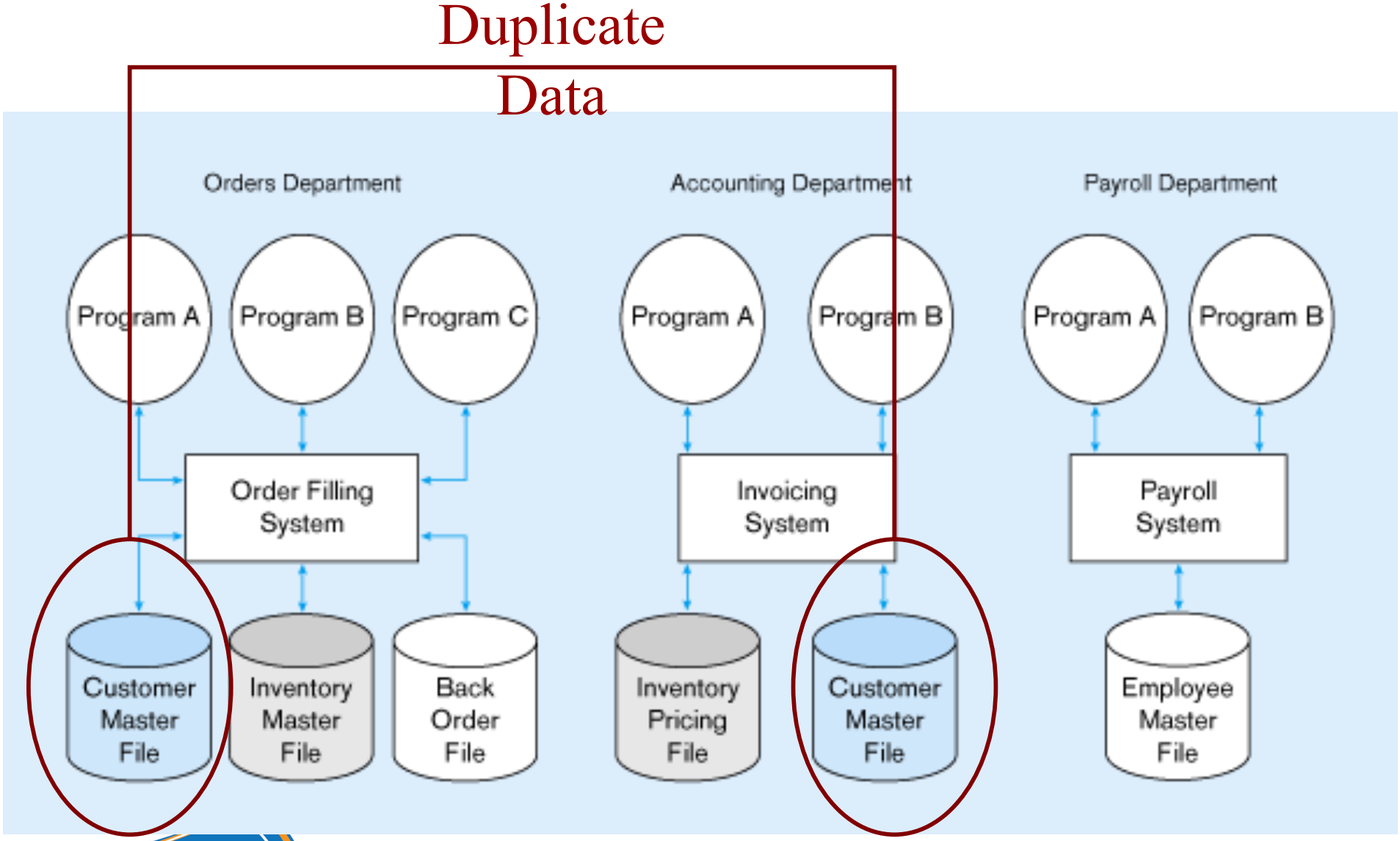
# File System vs. Database System

FIGURE 1.10 CONTRASTING DATABASE AND FILE SYSTEMS




(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Disadvantages of File Processing




(Coronel and Morris, 2015) (Hoffer et al., 2019)

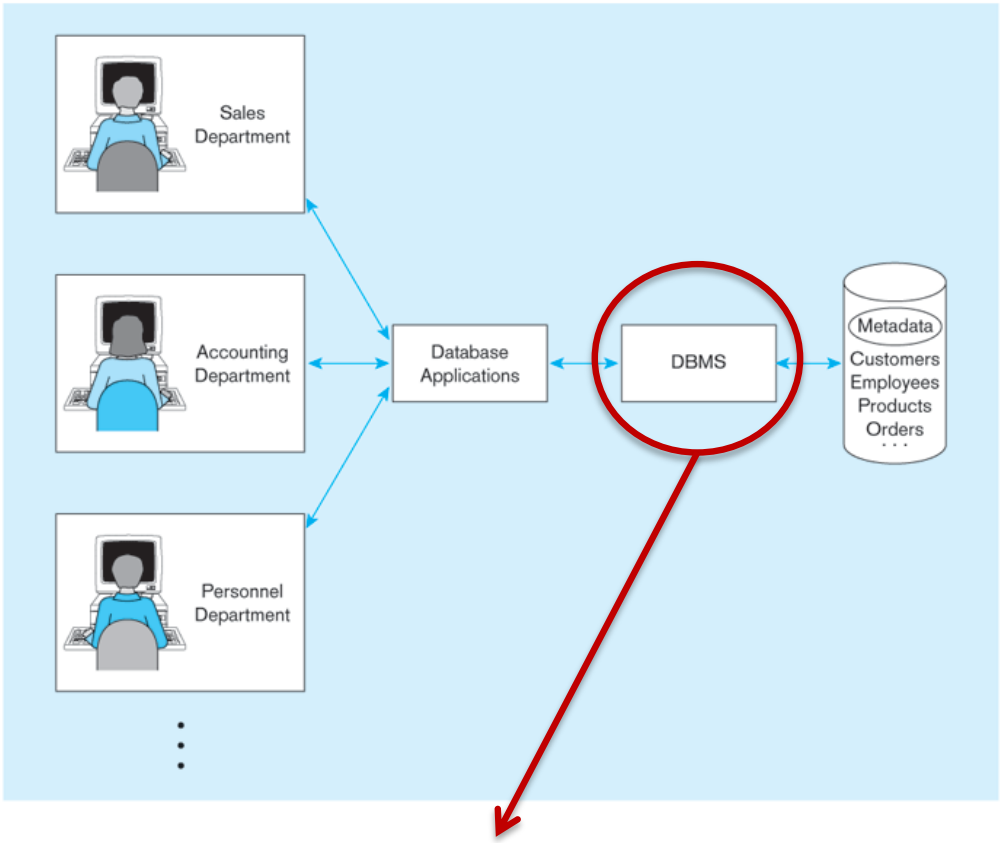
# Disadvantages of File Processing

- Program-Data Dependence
  - All programs maintain metadata for each file they use
- Duplication of Data
  - Different systems/programs have separate copies of the same data
- Limited Data Sharing
  - No centralized control of data
- Lengthy Development Times
  - Programmers must design their own file formats
- Excessive Program Maintenance
  - 80% of of information systems budget

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database System

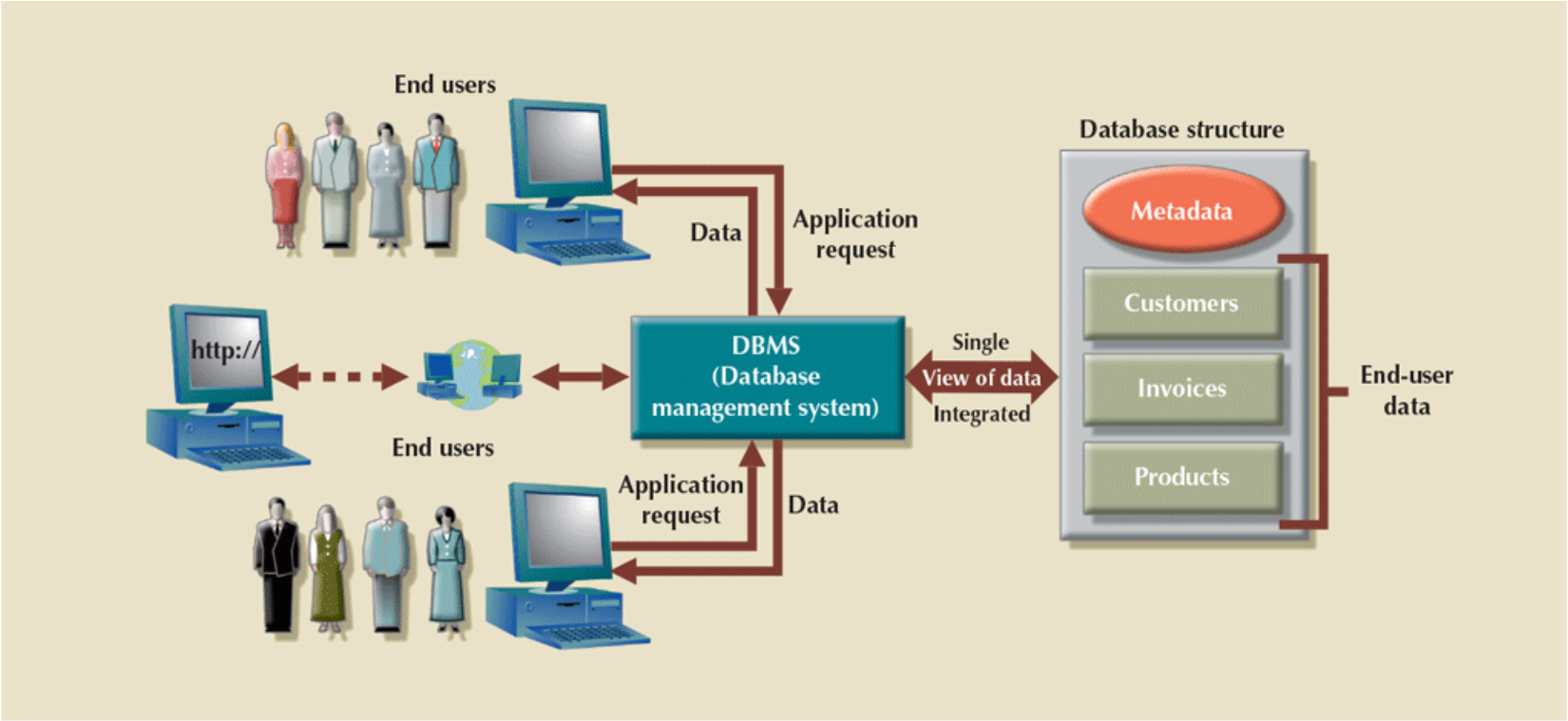
- Central repository of shared data
- Data is managed by a controlling agent
- Stored in a standardized, convenient form
- Requires a Database Management System (DBMS)




*Intermediary between the user and the database*

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database Management System




**The DBMS manages the interaction between the user and the database**

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Role of DBMS

- Manages interactions between users and databases
- Enables data to be shared
- Presents the end user with an integrated view of data
- Provides efficient and effective data management
- Improves sharing, security, integration, access, decision-making, productivity, etc.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

cdio 4.0

fit@hcmus

# DBMS Functions

- Data dictionary management
  - Data dictionary stores definitions of 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
  - Data is formatted to conform to required data structures
- Security management
  - Enforces user security and data privacy



(Coronel and Morris, 2015) (Hoffer et al., 2019)

# 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



(Coronel and Morris, 2015) (Hoffer et al., 2019)

# 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)**: a structured 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



(Coronel and Morris, 2015) (Hoffer et al., 2019)

# DBMS Disadvantages

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

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Content

- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities

(Coronel and Morris, 2015) (Hoffer et al., 2019)

fit@hcmus

# Categories of Databases

*Classification based on numbers of users supported*

- **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.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Categories of Databases

*Classification based on location*

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

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Categories of Databases

*Classification based on the type of stored information*

- **General-purpose** database contains a wide variety of data used in multiple disciplines.
- **Discipline-specific** database contains data focused on specific subject areas.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Categories of Databases

*Classification based on the purpose of databases*

- **Operational database** is 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 that can facilitate decision supports.
  - **Online analytical processing** (OLAP) are tools for working with data warehouses.
- **Business intelligence** captures and processes business data to generate information that support decision making.



(Coronel and Morris, 2015) (Hoffer et al., 2019)

fit@hcmus

# Categories of Databases

*Classification based on the structure degree of data*

- **Unstructured data** exists in its original (raw) state.
- **Structured data** results from formatting.
  - **Structure is applied based on type of processing to be performed.**
- **Semi-structured data** is processed to some extent.
  - **Extensible Markup Language (XML)** represents data elements in textual format.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Content

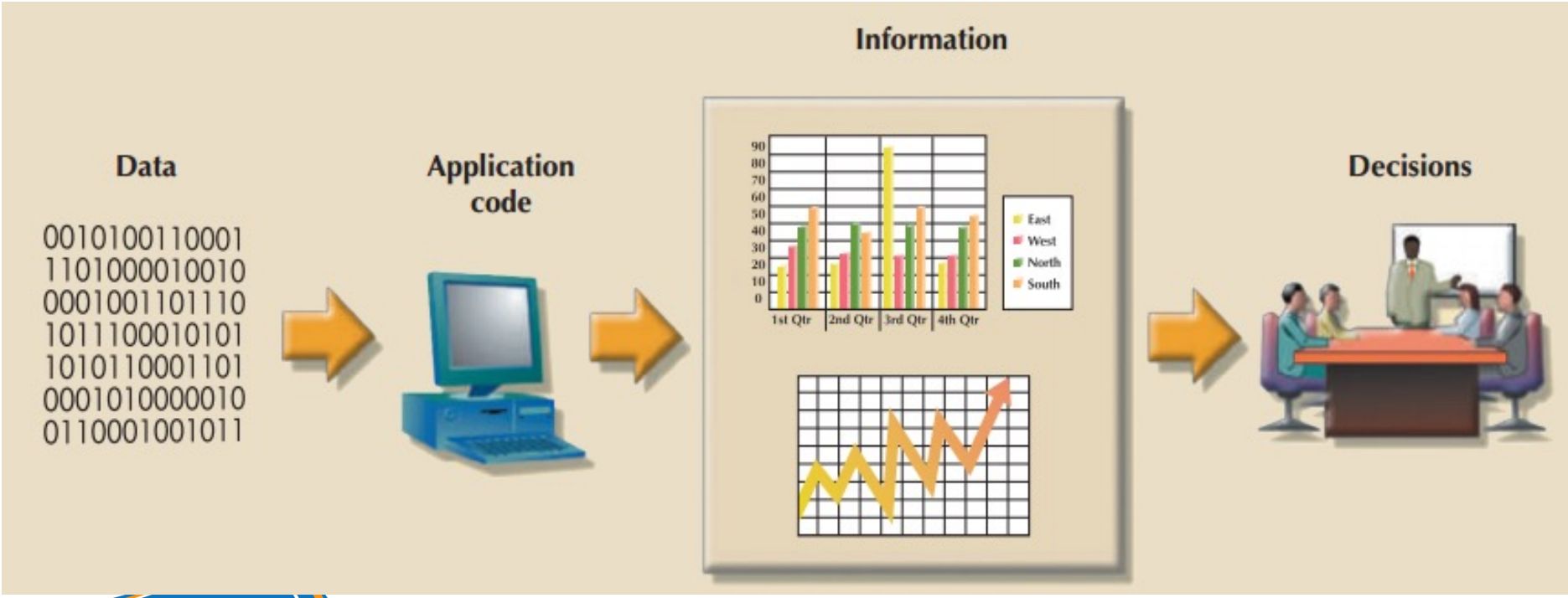
- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities



(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Information System

- Provides for data collection, storage, and retrieval.
- Composed of:
  - People, hardware, software (DBMS and others)
  - Database(s), application programs, and procedures




(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Information System

- **Systems analysis** is the process that establishes the need for an information system and its extent.
- **Systems development** is the process of creating an information system.
- **Database development** describes the process of **database design** and **implementation**.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Why Database Design Is Important?

- Focuses on design of 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 that may lead to poor decision making.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

fit@hcmus

# Why Database Design Is Important?

**SinhVien (mãsv, họtên, mãlớp, tênlớp, môn học, điểm)**

| Mãsv | Hotên | Mãlớp | Tênlớp | Mônhọc | Điểm |
|---|---|---|---|---|---|
| S1 | Tiến | L1 | MT01 | M1 | 9 |
| S1 | Tiến | L1 | MT01 | M2 | 7 |
| S1 | Tiến | L1 | MT01 | M3 | 8 |
| S2 | Trúc | L1 | MT01 | M1 | 9 |
| S2 | Trúc | L1 | MT01 | M2 | 8 |
| S3 | Hiền | L2 | MT02 | M1 | 5 |

- **repetition anomaly**
- **update anomaly**
- **insertion anomaly**
- **deletion anomaly**

- Update **S1**?
- Add a new student **S4**?
- Delete **S3**?

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Type of Database Design

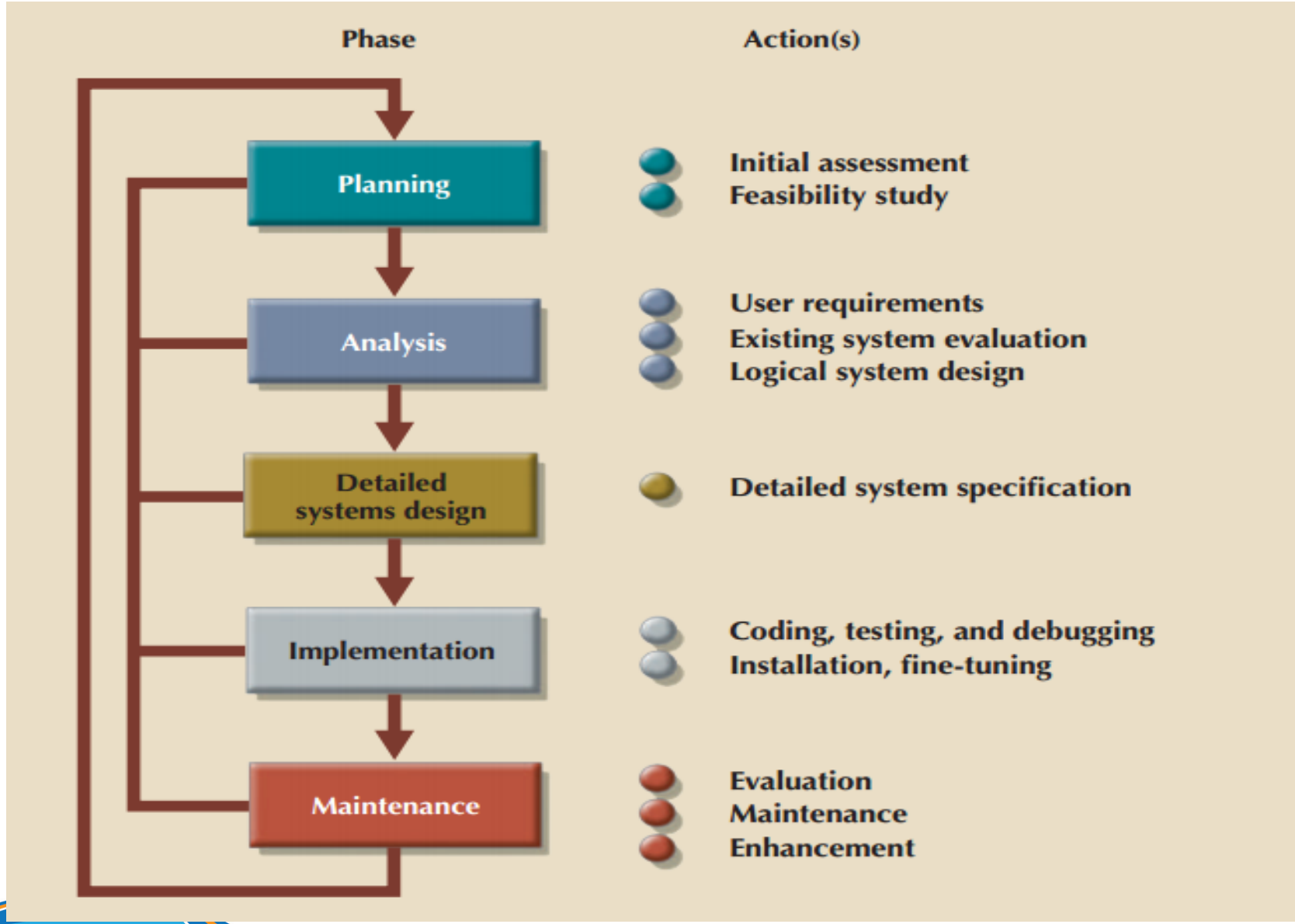
- **Existing Database**
  - Old data: worksheet, files.
  - Extracting data from another databases
  - Design using Normal Form
- **Design a new database**
  - Database analysis based on user requirements.
  - Database design for data extracting.
- **Re-design Database**
  - Transform existing database to a new one.
  - Integrated two or more databased together.
  - Giving techniques & design methods for databases.
  - Applying normal form principle and mapping rules between models





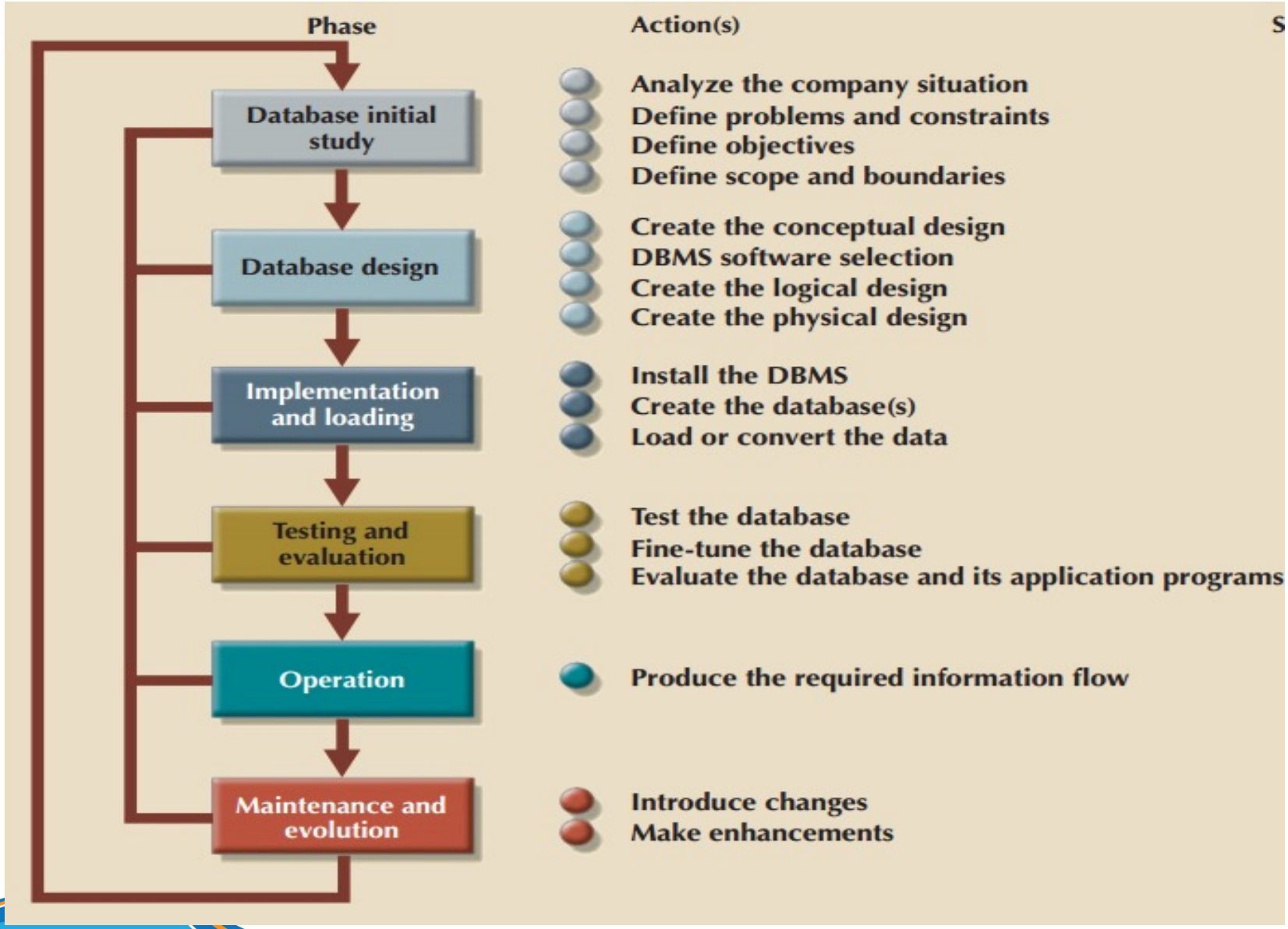
(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Systems Development Life Cycle (SDLC)




(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database Life Cycle (DBLC)






(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database Initial Study (Phase 1)

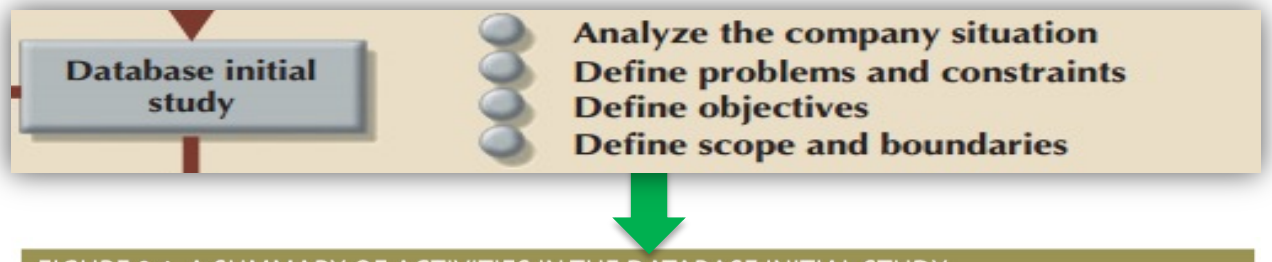
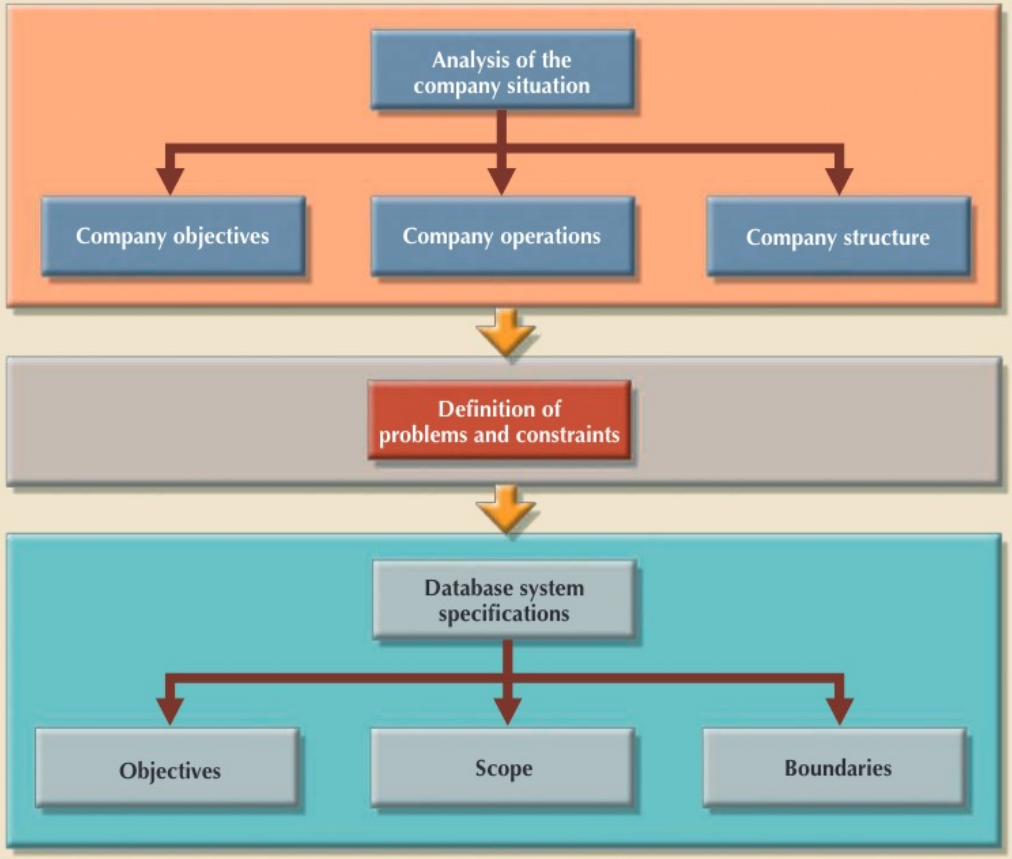



FIGURE 9.4 A SUMMARY OF ACTIVITIES IN THE DATABASE INITIAL STUDY




(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database Design (Phase 2)

- Support company operations and objectives.
- Most critical DBLC phase:
  - Ensures final product meets user and system requirements.
- Points for examining:
  - Must concentrate on the data characteristics required to build the database model.
  - Two views of the data within the system: the business view of data as a source of information and the designer's view of the data structure, its access, and the activities.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Database Design (Phase 2)


FIGURE 9.6 DATABASE DESIGN PROCESS

| Section | Stage | Steps | Activities |
|---|---|---|---|
| 9-4 | Conceptual Design | • Data analysis and requirements | • Determine end-user views, outputs, and transaction requirements |
| | | • Entity Relationship modeling and normalization | • Define entities, attributes, domains, and relationships<br>• Draw ER diagrams; normalize entity attributes |
| | | • Data model verification | • Identify ER modules and validate insert, update, and delete rules<br>• Validate reports, queries, views, integrity, access, and security |
| | | • Distributed database design* | • Define the fragmentation and allocation strategy |
| | | | DBMS and Hardware Independent |
| 9-5 | DBMS Selection | Select the DBMS | • Determine DBMS and data model to use |
| | | | DBMS Dependent |
| 9-6 | Logical Design | • Map conceptual model to logical model components | • Define tables, columns, relationships, and constraints |
| | | • Validate logical model using normalization | • Normalized set of tables |
| | | • Validate logical model integrity constraints | • Ensure entity and referential integrity; define column constraints |
| | | • Validate logical model against user requirements | • Ensure the model supports user requirements |
| | | | Hardware Dependent |
| 9-7 | Physical Design | • Define data storage organization | • Define tables, indexes, and views' physical organization |
| | | • Define integrity and security measures | • Define users, security groups, roles, and access controls |
| | | • Determine performance measures* | • Define database and query execution parameters |

* See Chapter 12, Distributed Database Management Systems
* See Chapter 11, Database Performance Tuning and Query Optimization

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Implementation and Loading (Phase 3)

- Install the DBMS
  - **Virtualization** is a technique that creates logical representations of computing resources that are independent of the underlying physical computing resources.
- Create databases
  - The creation of special **storage-related constructs** to house the end user tables may by required.
- Load or convert the data
  - The **aggregation of data** from multiple sources may be required.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Testing and Evaluation (Phase 4)

- Test the database
  - Physical security, password security, access rights, audit trails, data encryption, diskless workstations.
- Fine-tune the database
  - Database performance during high-volumes of operations and data evolution.
- Evaluate the database and its application programs
  - Integration issues and deployment plan will be refined.
  - Backup and recovery plans are tested.
  - User training is conducted.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Operation (Phase 5)

- The targeted end users have entered the operations phase.
- Unforeseen problems may occur during this phase.
- The demand for change leads to phase 6, maintenance and evolution.

(Coronel and Morris, 2015) (Hoffer et al., 2019)

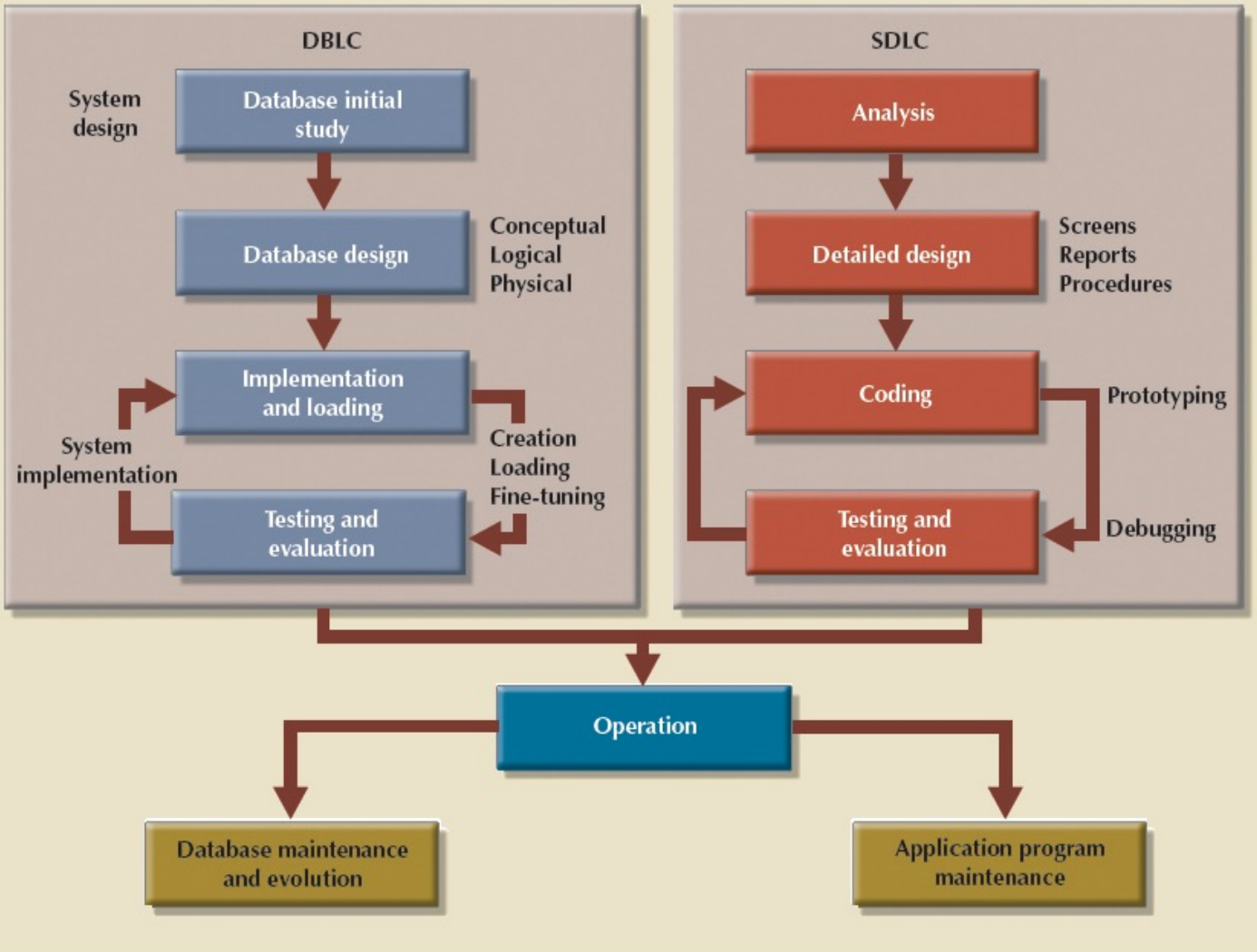
# Maintenance and Evolution (Phase 6)

- Preventive maintenance (backup)
- Corrective maintenance (recovery)
- Adaptive maintenance (enhancing performance)
- Assignment of access permissions and their maintenance for new and old users
- Generation of database access statistics
- Periodic security audits
- Periodic system-usage summaries

(Coronel and Morris, 2015) (Hoffer et al., 2019)

# DBLC vs. SDLC

FIGURE 9.8 PARALLEL ACTIVITIES IN THE DBLC AND THE SDLC




(Coronel and Morris, 2015) (Hoffer et al., 2019)

# Content

- Database
- Database Systems
- Categories of Databases
- System Development Process
- Database Career Opportunities

(Coronel and Morris, 2015) (Hoffer et al., 2019)

cdio 4.0

fit@hcmus

# Database career oppotunities

| DATABASE CAREER OPPORTUNITIES | | |
|---|---|---|
| **JOB TITLE** | **DESCRIPTION** | **SAMPLE SKILLS REQUIRED** |
| Database Developer | Create and maintain database-based applications | Programming, database fundamentals, SQL |
| Database Designer | Design and maintain databases | Systems design, database design, SQL |
| Database Administrator | Manage and maintain DBMS and databases | Database fundamentals, SQL, vendor courses |
| Database Analyst | Develop databases for decision support reporting | SQL, query optimization, data warehouses |
| Database Architect | Design and implementation of database environments (conceptual, logical, and physical) | DBMS fundamentals, data modeling, SQL, hardware knowledge, etc. |
| Database Consultant | Help companies leverage database technologies to improve business processes and achieve specific goals | Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc. |
| Database Security Officer | Implement security policies for data administration | DBMS fundamentals, database administration, SQL, data security technologies, etc. |
| Cloud Computing Data Architect | Design and implement the infrastructure for next-generation cloud database systems | Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc. |
| Data Scientist | Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors | Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization |



(Coronel and Morris, 2015) (Hoffer et al., 2019)

# References

(Coronel and Morris, 2015, **chapter 1, 9**)

**Database System: Design, Implementation, and Management, 12th Edition, Carlos Coronel & Stevem, Morris, 2015.**

(Hoffer et al., 2019, **chapter 1**)

**Modern Database Management, 13th Edition, Jeffrey A. Hoffer, V. Ramesh, Heikki Topi, 2019.**

# THE END