Relational Database Basics Review

IT 4153 Advanced Database

J.G. Zheng

Spring 2012

Overview

- Database approach
- Database system
- Relational model
- Database development

File Processing Approaches

- Based on file systems
  - Data are record in various types of files organized in folders (directories)
- File types
  - Sequential data files
  - Name-value pair files
  - Spreadsheets or list files
  - XML files

Files Processing Problems

- Loose and weak structure (a general structure may be imposed but not enforced)
  - Difficult to handle complex data
  - Low data quality: redundancy and inconsistency
- No central management
  - Difficult to maintain and share in multi-user environments
  - Limited security
- Not scalable: cannot handle large quantity of data efficiently
- Lack of specialized and standardized data management and processing capabilities

Database Approach

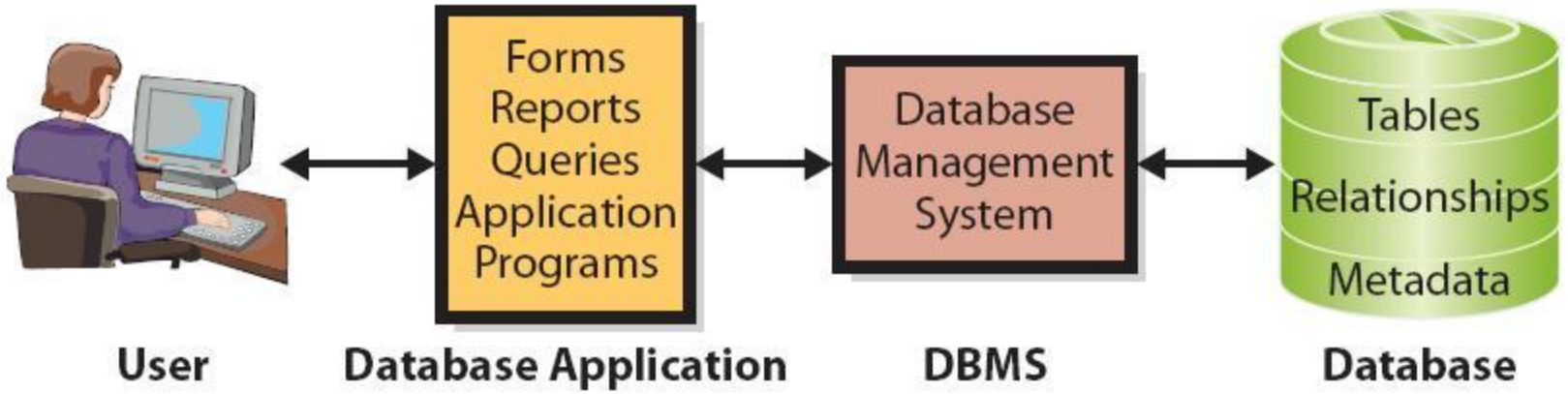
- Database is a structured and self-describing collection of data
  - Structured: structures and rules are consistently and rigorously defined and enforced (integrity)
  - Self-describing: the description of data (data definition, or metadata) is contained within the database
- Centralized management
  - Managed and controlled by specialized programs, called database management systems (DBMS), which provides rich data management functionalities

Advantages and Disadvantages

- Advantages
  - High data quality, integrity, and consistency
  - Reduced data redundancy and application maintenance
  - Easy access and sharing
  - Scalable
  - Improved security
  - Specialized and productive management tool
- Major disadvantages
  - Increased complexity
  - Greater impact of failure

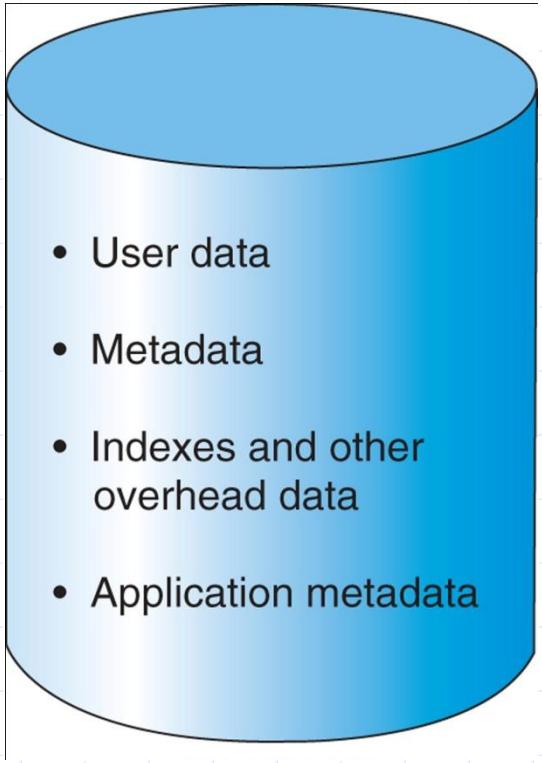
Database System

- A database system is a complete information system
- Basic layers of a database system



Database

- A database is a storage place for data
- What's in the database?
  - Data (tables)
  - Metadata
  - Other data and structures



Metadata

- Metadata are data that describe data (data definitions)

- Metadata is always a part of a database.

Metadata defines tables, columns, data types, keys (relationships), constraints, etc.

USER_TABLES Table

TableName	NumberColumns	PrimaryKey
STUDENT	3	StudentNumber
CLASS	4	ClassNumber
GRADE	3	(StudentNumber, ClassNumber)

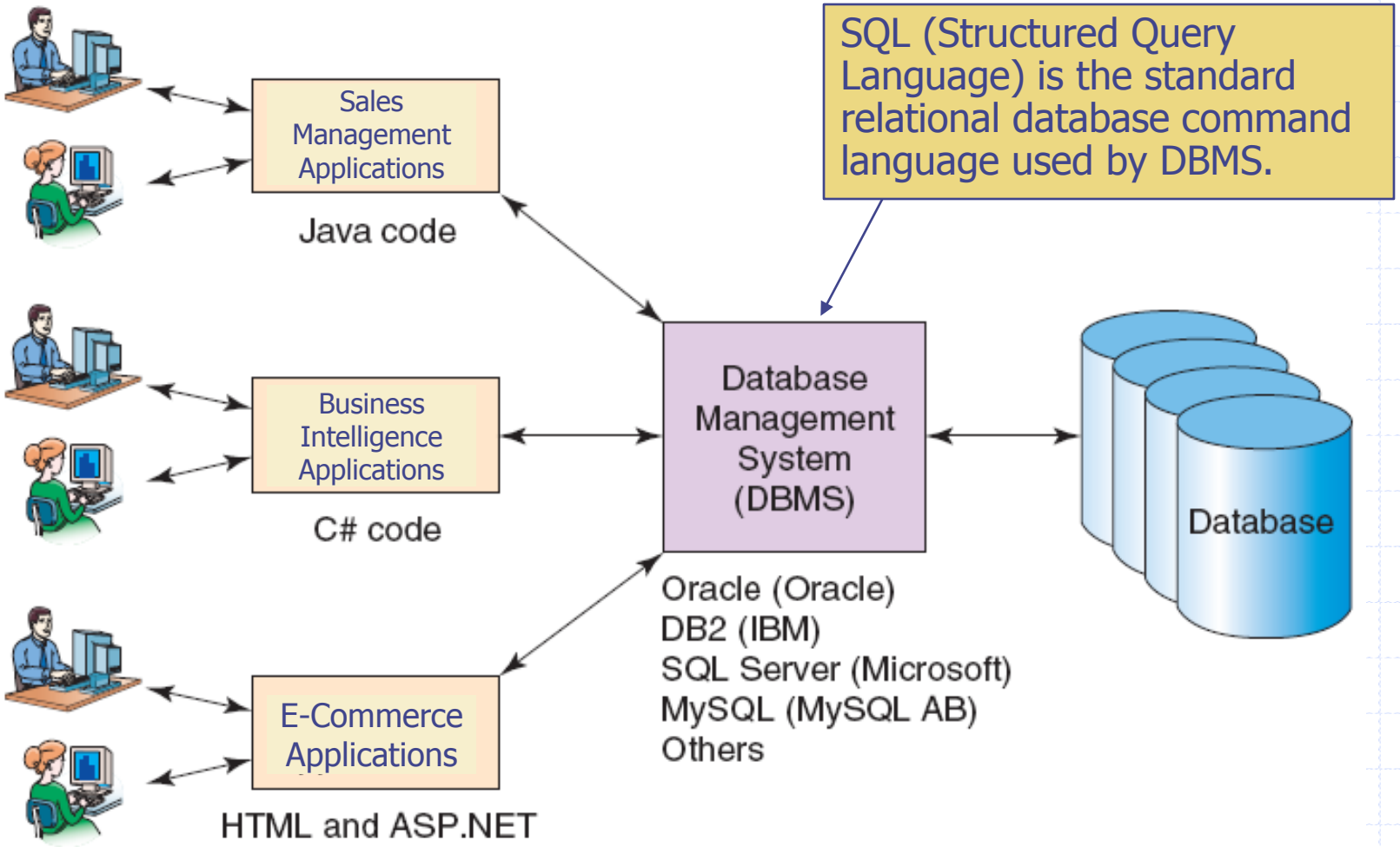
USER_COLUMNS Table

ColumnName	TableName	DataType	Length (bytes)
StudentNumber	STUDENT	Integer	4
LastName	STUDENT	Text	25
FirstName	STUDENT	Text	25
EmailAddress	STUDENT	Text	100
ClassNumber	CLASS	Integer	4
Name	CLASS	Text	25
Term	CLASS	Text	12
Section	CLASS	SmallInteger	2
StudentNumber	GRADE	Integer	4
ClassNumber	GRADE	Integer	4
Grade	GRADE	Decimal	(2, 1)

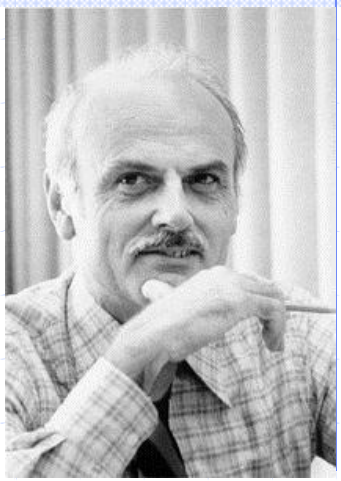
Database Management System

- DBMS serves as a controller (gatekeeper) for databases
- DBMS provides common functionalities and interfaces for managing and controlling database activities, such as
  - creating and maintaining databases and other structures
  - reading, updating and deleting data
  - data backup and recovery
  - controlling concurrency, consistency, and enforcing other rules
  - providing security

Enterprise Database System



Relational Model



- Edgar F. Codd (IBM), 1970
- One sentence to summarize relational database model (extremely brief):

  Data are organized in *relations* (tables), which are linked by *keys* (constraints)

Relation

- A relation is a two-dimensional table that has some specific characteristics:
  1. The table consist of rows and columns
  2. Rows contain data about entity instances
  3. All values in a row describes the same entity instance
  4. Columns contain data about attributes of the entity
  5. All values in a column are of the same kind
  6. Each row is distinct
  7. A cell of the table holds a single value
  8. Each column has a unique name
  9. The order of the rows is unimportant
  10. The order of the columns is unimportant

Terminology Contrast

Database industry	**Table**	**Row**	**Column**
Academic	Relation	Tuple	**Attribute**
File processing	File	**Record**	**Field**

Key

- A key is one or more columns of a relation that is used to uniquely identify a record
  - Primary key
  - Candidate key
  - Alternate key
  - Surrogate key
  - Composite key
  - Foreign key

Candidate Key/Primary Key

- Candidate key
  - The minimum set of column(s) that uniquely identifies a single record (row)
  - Each value in this column is unique in this relation

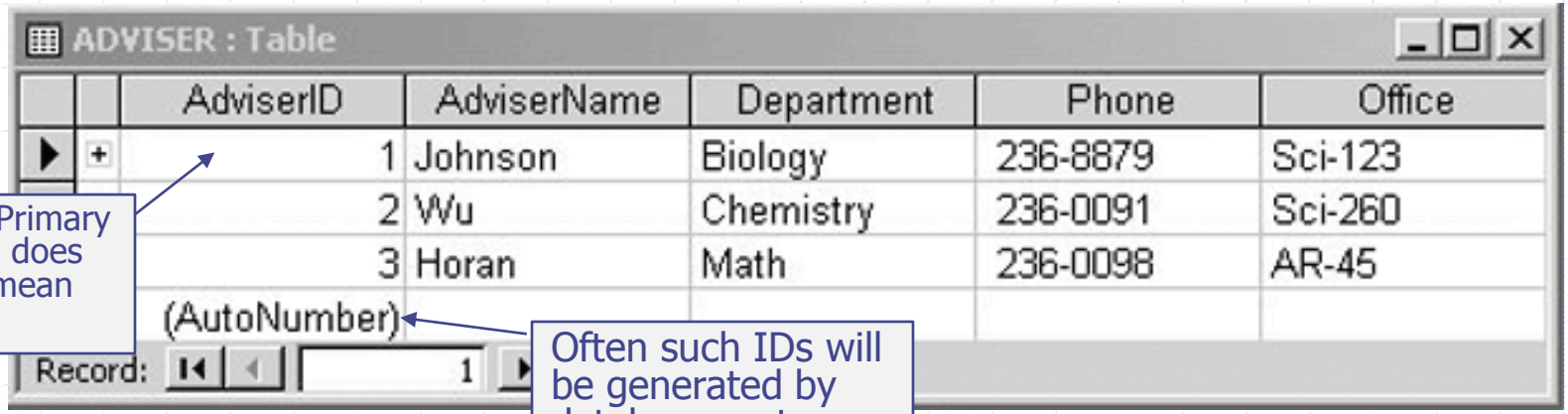
- Primary key
  - Primary key is a column/attribute that is used to uniquely identify a record
  - Is one of the candidate keys chosen to be the identifying key; others become alternate keys
  - Each value of this key column uniquely identifies a single record (row)
  - There is only ONE primary key for a table

Composite Key

- Composite key
  - A composite key contains two or more attributes (columns)
  - All keys can be composite keys
- Example:
  - “FirstName” + “LastName”
  - “FirstName” + “LastName” + “BirthDate”
  - “FirstName” + “LastName” + “BirthDate” + “BirthCity”
  - …

Artificial Primary Key/Surrogate Key

- Sometimes it is difficult to find a natural attribute as a primary key, or it is difficult to use a composite key.
- A column is created arbitrarily and assign each record a unique number/id
  - Product Number, Product Id, Movie Id, Actor Id, etc.



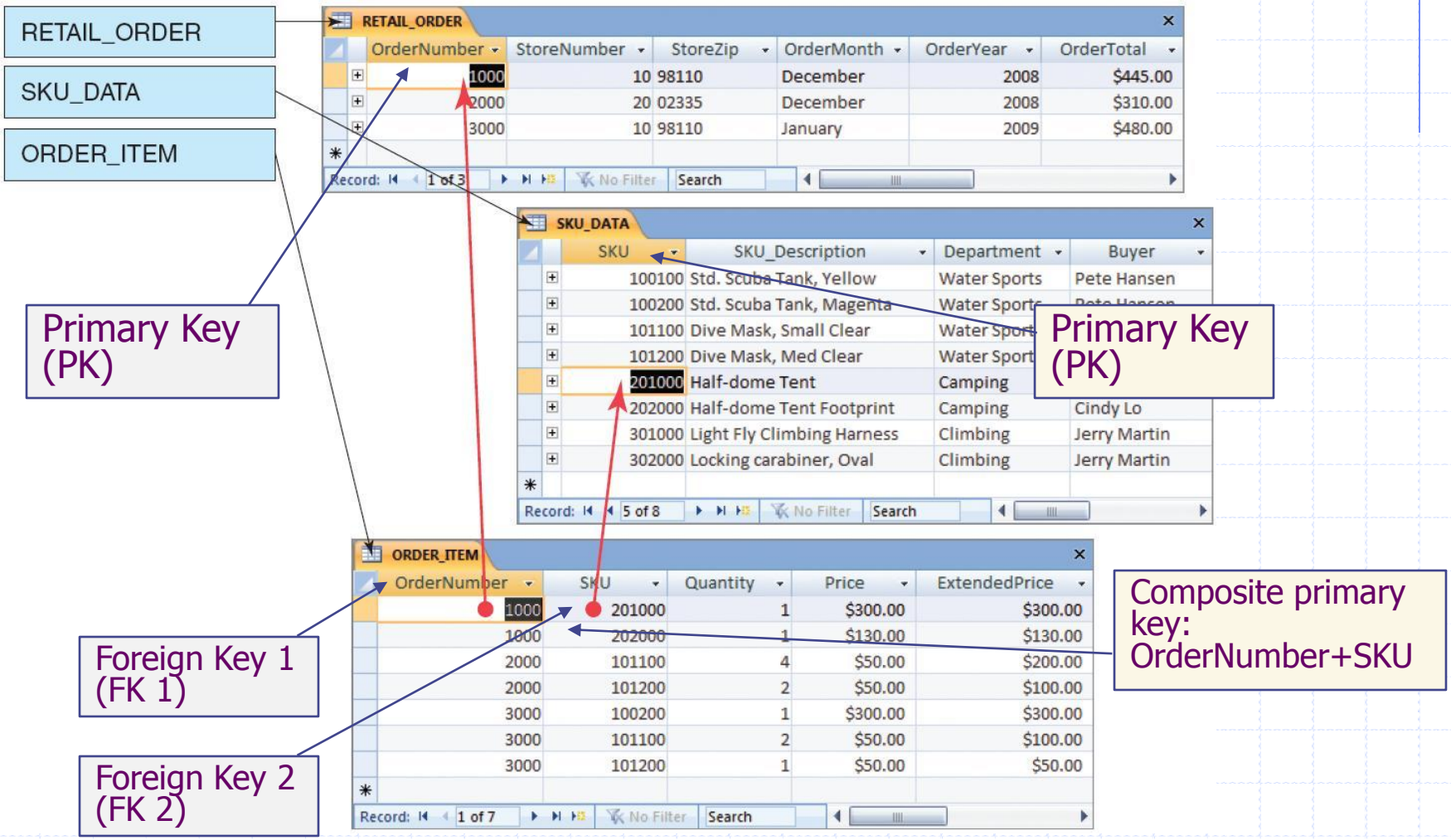
PK Selection Guidelines

- Do not use a field whose value is frequently changed as PK
- Look for single-attribute PK first
- If a PK contains more than 3 columns, consider a surrogate key
- Don’t be limited to sample data; think beyond and consider possible scenarios and requirements

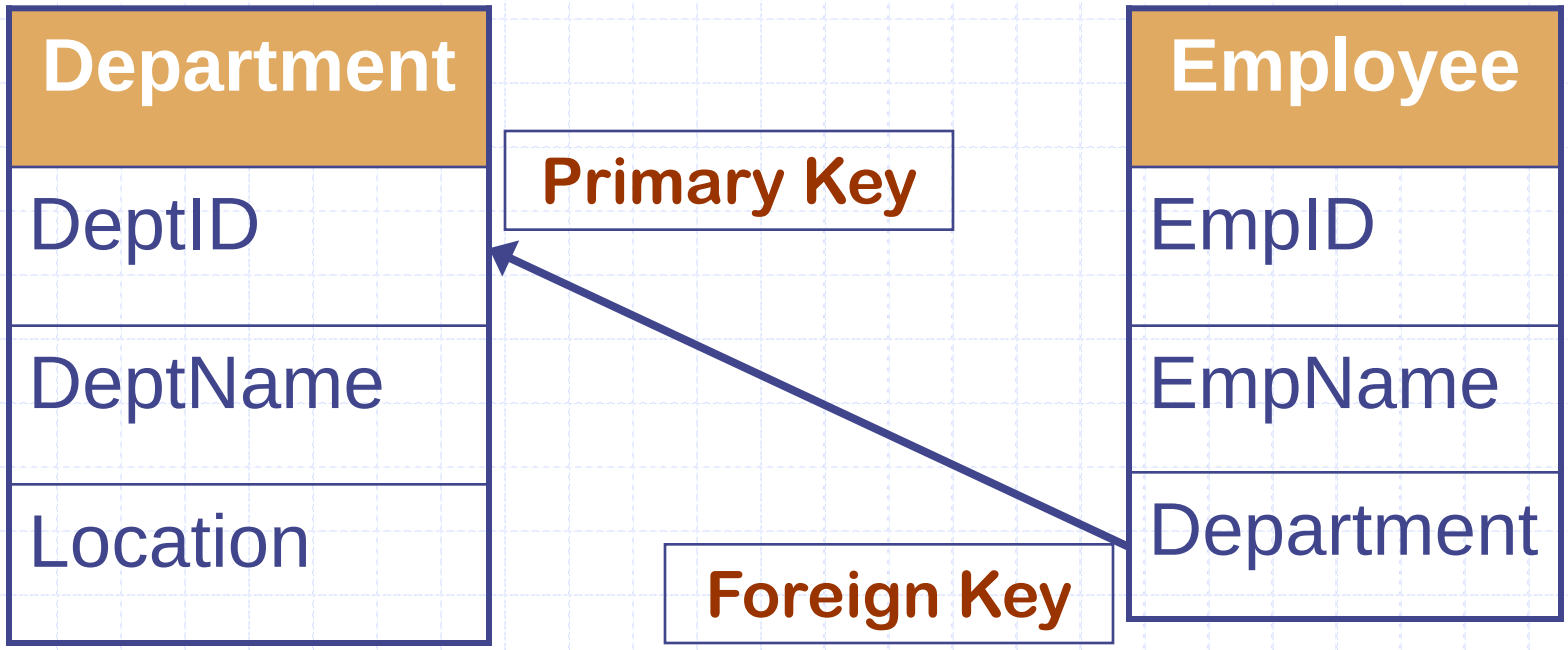
Relationship and Foreign Key (FK)

- Relationship is how tables (relations) are linked
  - It is defined by the *foreign key (FK)* constraint
- A foreign key references a primary key (or any other unique keys) in another table
  - This *pair* of keys are of the *same kind* (may be of different name)

Relationship and FK Example

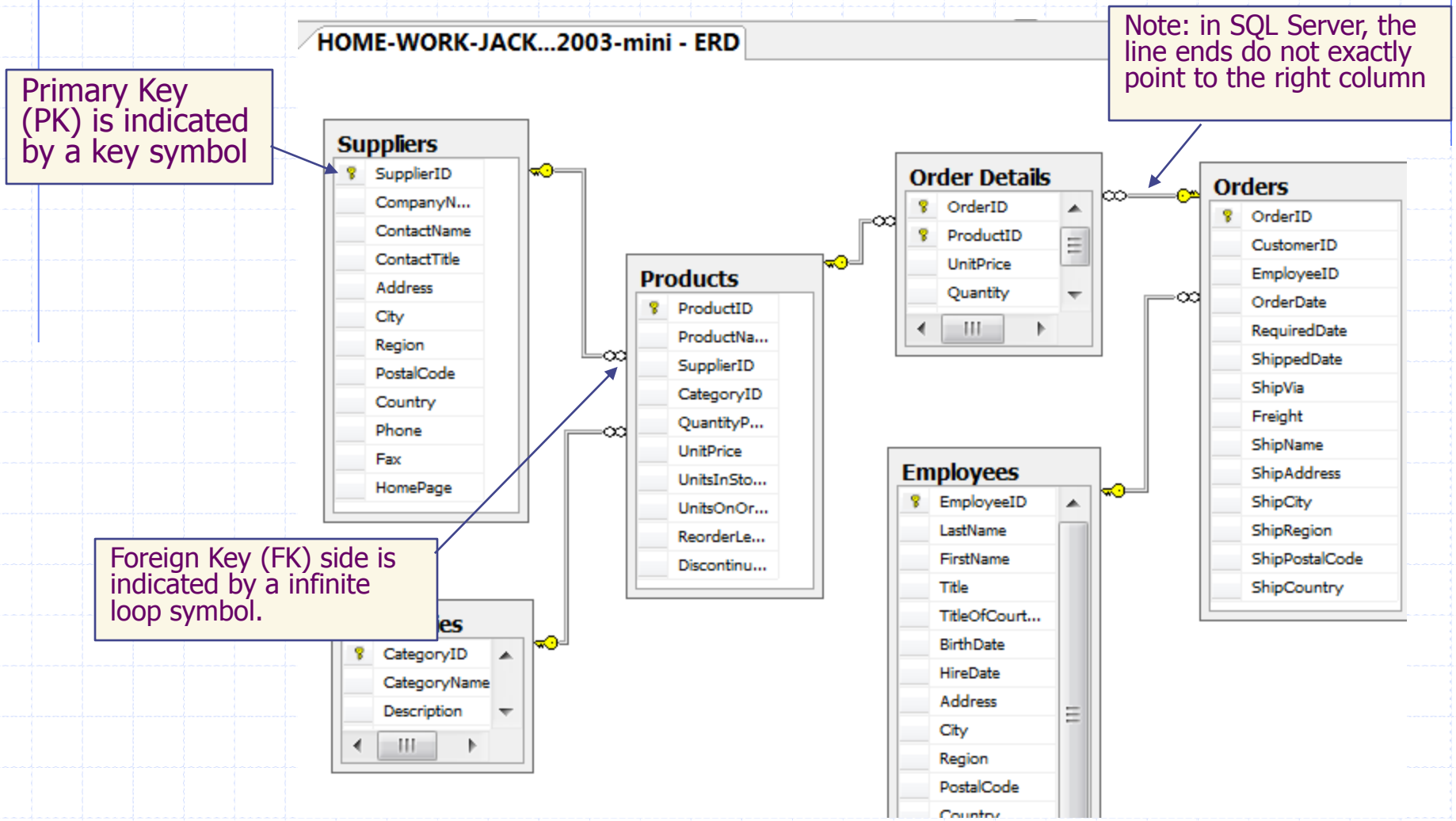


Foreign Key Example



Primary Key and Foreign Key are of the same type (string, number, etc.) and length, but they do not necessarily have the same name.

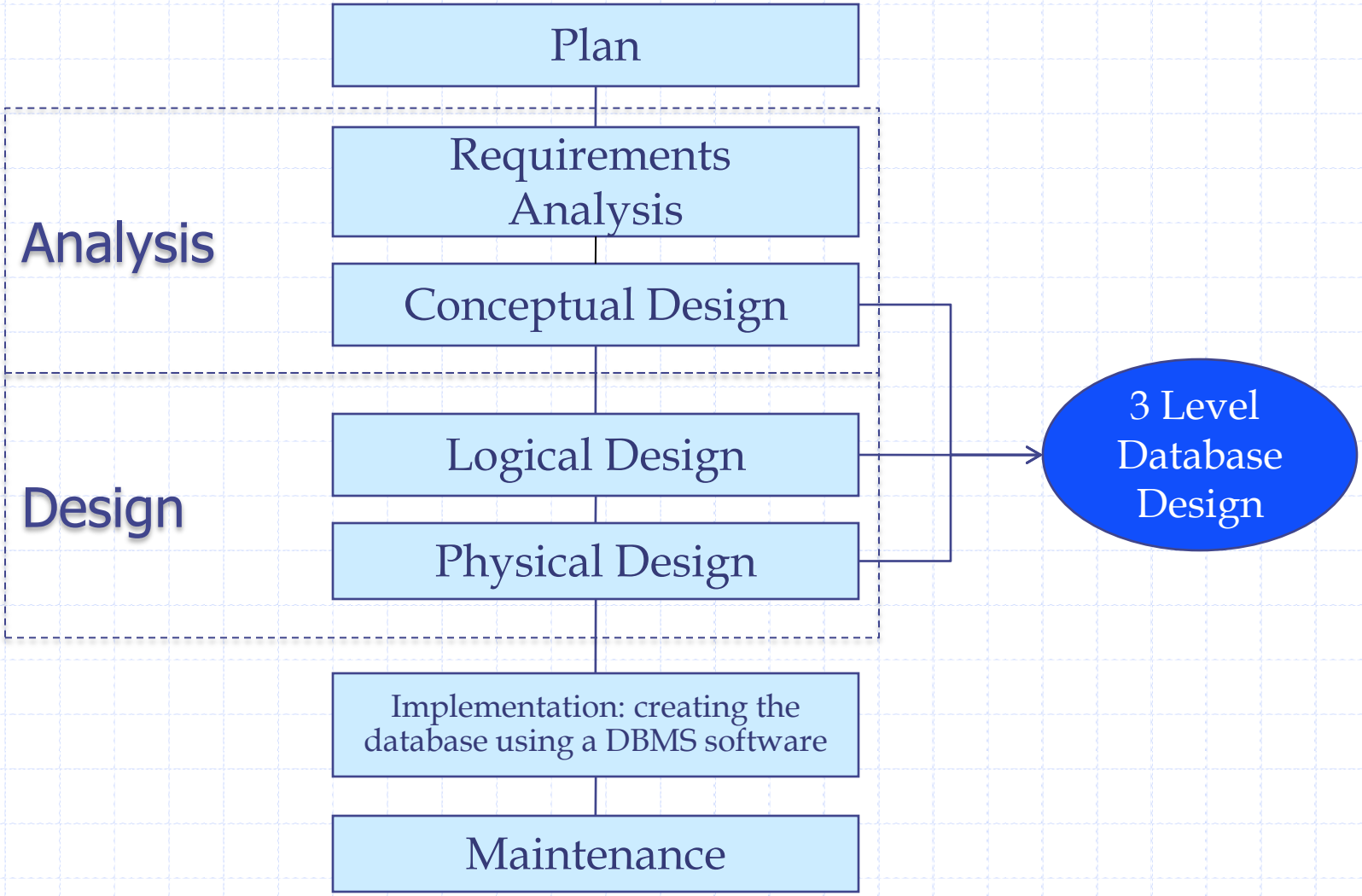
Relationship in SQL Server Database



Schema in Relational Databases

- Schema is the structure described in a formal language supported by the database management system (DBMS)
  - It's a kind of metadata
- Relational database schema commonly defines
  - Tables: name, primary key
  - Columns: name, data type, size, value range, etc.
  - Constraints: all kinds of keys
  - Other structures

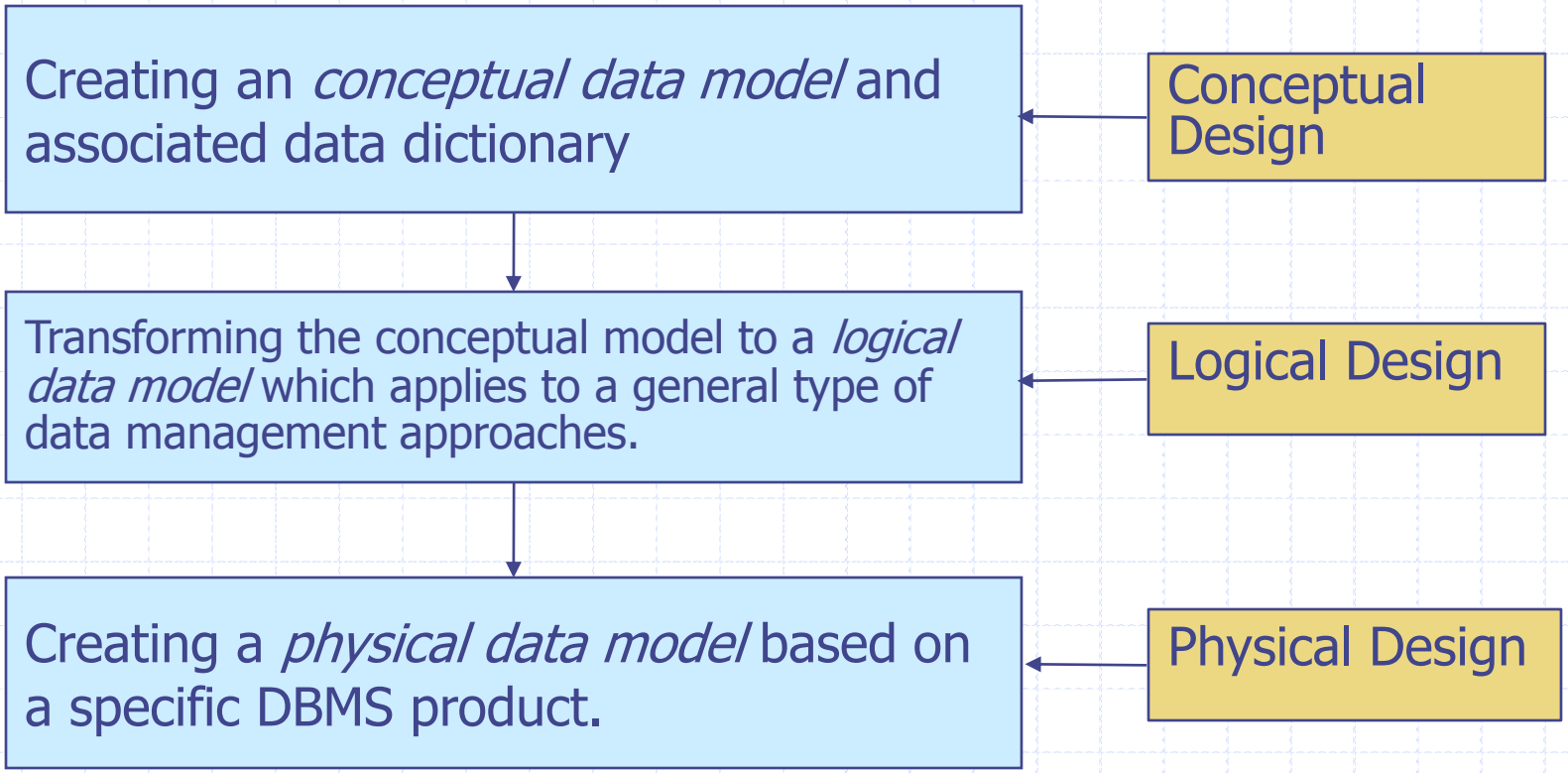
Database Development



Requirement Analysis

- Requirements describe what a system should be or do
  - Functional vs. non-functional requirements
  - Process vs. data requirements
- Issues in requirements engineering
  - How to obtain requirements?
  - How to record/represent requirements?

3 Level Database Design



Data Model

- A model is a general and abstract representation of something more complicated and detailed
  - Process model
  - Data model
  - UML model
- A data model is a general and abstract representation of the structure of data
  - Conceptual
  - Logical
  - Physical

Conceptual Modeling/Design

- Conceptual data model
  - A high level representation of the reality based on human understanding
  - It is abstract, simple, yet meaningful
  - Not tied to any computing technologies
- Examples
  - Entity Relationship Diagram (ERD)
  - Semantic data model
  - Concept diagram
  - Data structure diagram

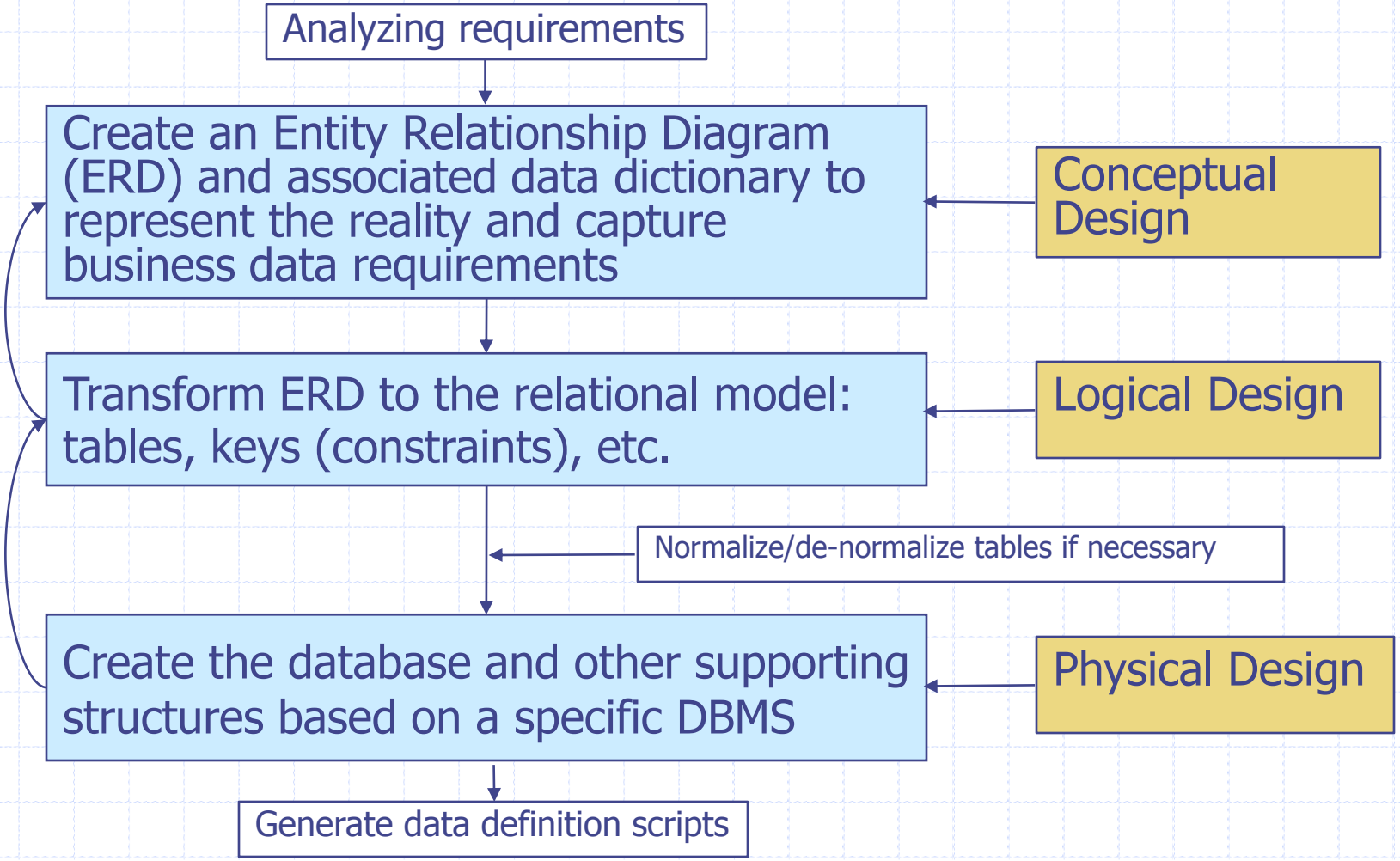
Logical Modeling/Design

- Logical data model
  - A specific data structure that organizes data following specific rules (logics); for example, mathematical or computing rules
  - It is more detailed, structured, actionable and has specific rules
  - Yet it is implementation (product) independent
- Examples
  - Relational data model
  - Object-oriented data model
  - Hierarchical data model (XML)

Physical Modeling/Design

- Physical data model
  - Based on a specific implementation (a software product), which implements the corresponding logical data model and adds more operational details
- In database design
  - More details are added to relational models that directly support the creation of the database in a certain database product
  - DBMS product specific: data types, storage methods, DBMS capability, proprietary functions and rules, etc.
  - Including lower level details and other DBMS structure, such as index, partition, cluster, storage, etc.

A Practical Process for Relational Database Design based on the Three-Data Model Approach



Summary

- Key concepts and terms
  - File processing vs. database processing
  - Database features, advantages, and disadvantages
  - Database, database system, DBMS, database application
  - Data, metadata, database schema
  - Oracle, SQL Server, DB2, MySQL
  - Relation (and its 10 features)
  - Row, column, record, field, attribute
  - Key, primary key, candidate key, surrogate key, composite key, foreign key
  - SQL
- Key concepts
  - Three data models and examples
    - Conceptual, logical, physical
    - ERD
    - Relational model
  - Three-level database design method
    - Conceptual design
    - Logical design
    - Physical design
- Key skills
  - Identify/design the primary key, composite primary key, candidate keys, and foreign keys of a given table/relation.

Good Readings and Resources

- Database system
  - http://en.wikipedia.org/wiki/Database
  - http://en.wikipedia.org/wiki/Database_management_system
- Relational model
  - http://en.wikipedia.org/wiki/Relational_model
- The database development life cycle
  - http://openlearn.open.ac.uk/mod/oucontent/view.php?id=399373
- Data modeling 101
  - http://www.agiledata.org/essays/dataModeling101.html
- Data model
  - http://en.wikipedia.org/wiki/Data_model
- The database report: latest database industry news
  - http://www.tdan.com/featured_columns/db_report.php