Database Systems Overview

Truong Tuan Anh
CSE-HCMUT

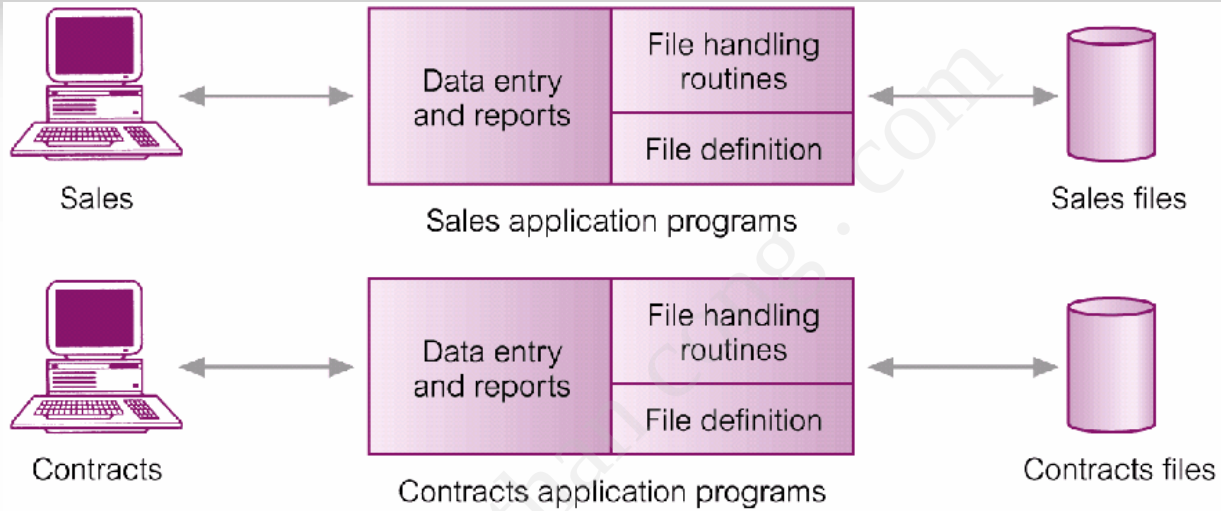
Outline

- File-based Approach and Database Approach
- Three-Schema Architecture and Data Independence
- Database Languages
- Data Models, Database Schema, Database State

File-based Approach

- Data is stored in one or more separate computer files
- Data is then processed by computer programs - applications

File-based Approach



Sales Files

PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

PrivateOwner (ownerNo, fName, lName, address, telNo)

Client (clientNo, fName, lName, address, telNo, prefType, maxRent)

Contracts Files

Lease (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

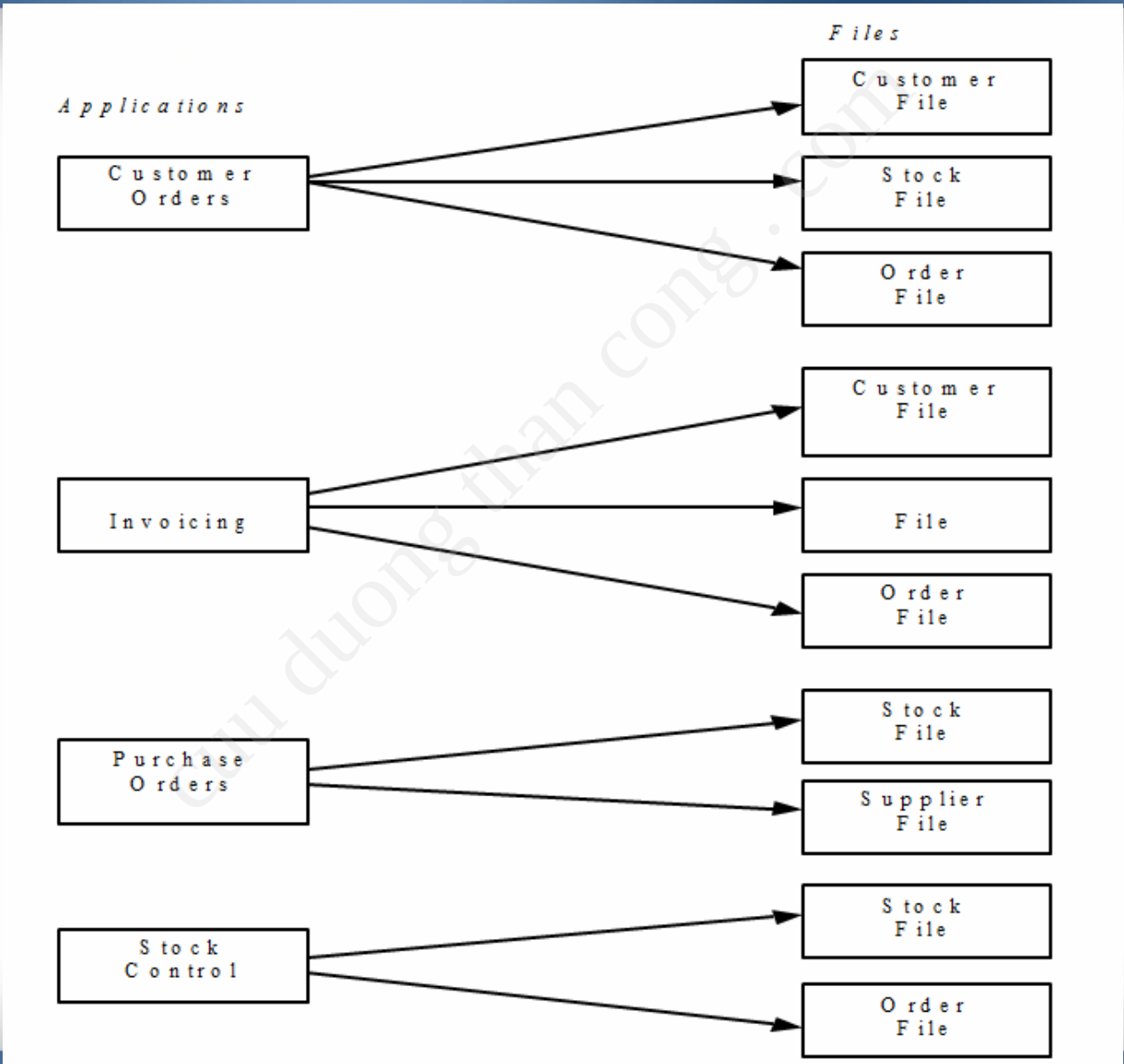
PropertyForRent (propertyNo, street, city, postcode, rent)

Client (clientNo, fName, lName, address, telNo)

File-based Approach

- Problems/Limitations
  - Data Redundancy
  - Data Inconsistency
  - ...

File-based Approach

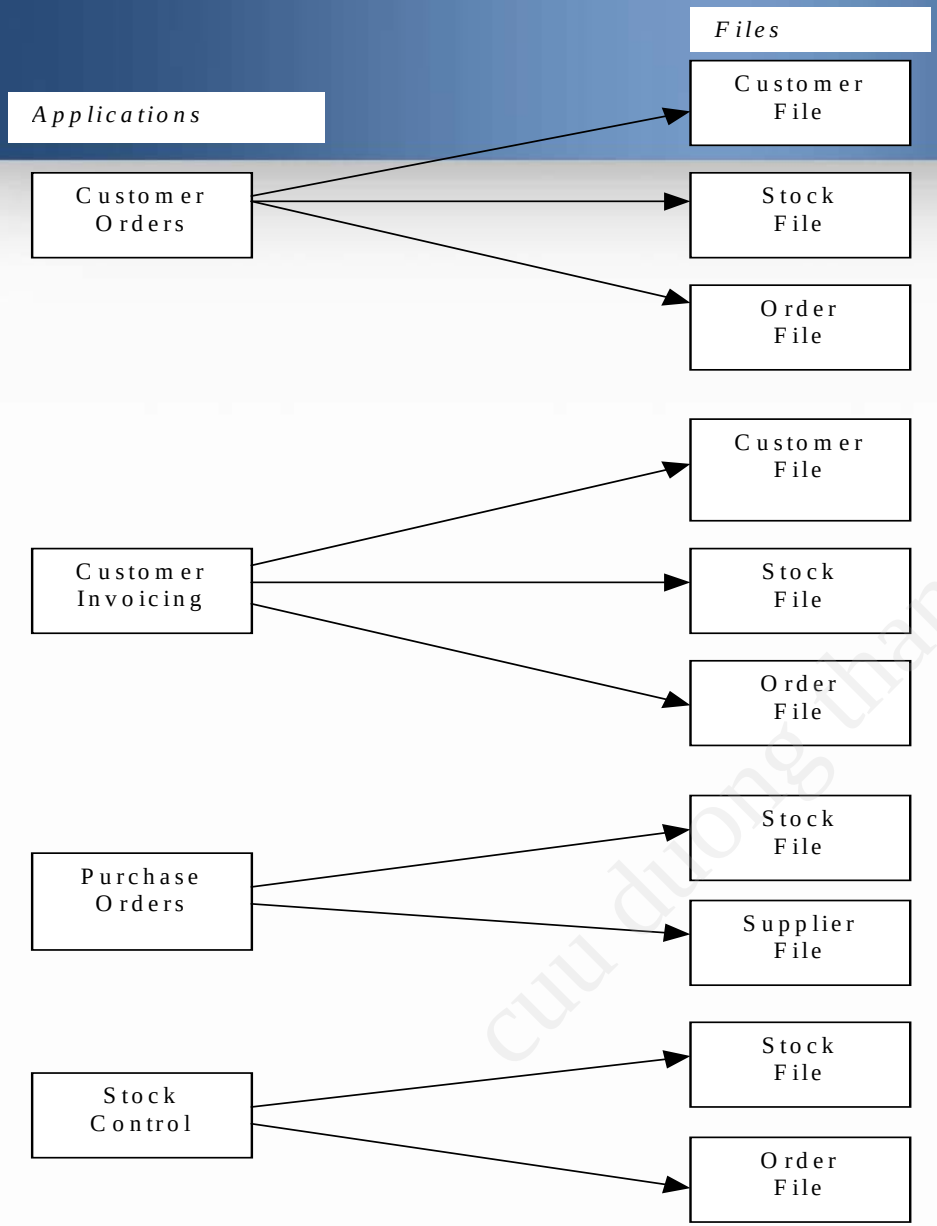


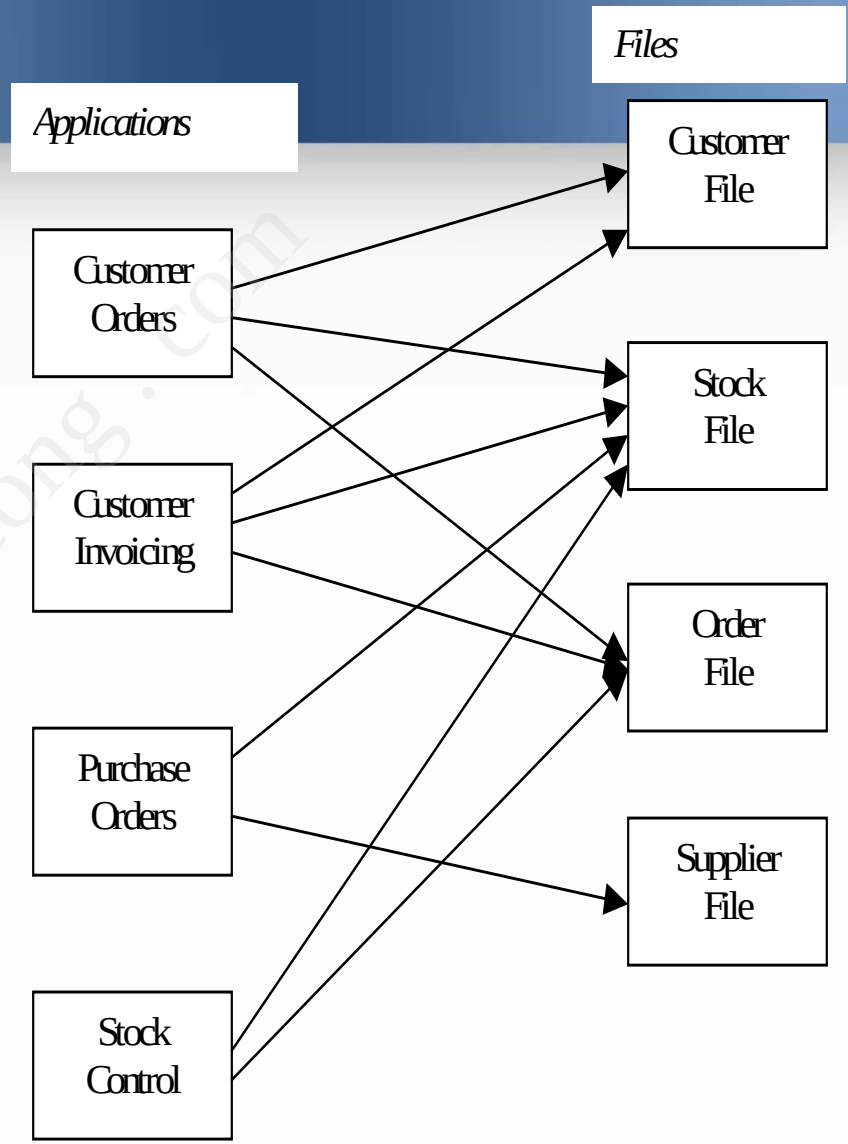
File-based Approach

- Shared File Approach
  - Data (files) is ***shared*** between different applications
  - **Data redundancy** problem is alleviated.
  - **Data inconsistency** problem across different versions of the same file is solved.

File-based Approach

- Shared File Approach
  - Other problems:
    - ***Rigid data structure:*** If applications have to share files, the file structure that suits one application might not suit another.
    - ***Physical data dependency:*** If the structure of the data file needs to be changed in some way, this alteration will need to be reflected in all application programs that use that data file.
    - ***No support of concurrency control:*** While a data file is being processed by one application, the file will not be available for other applications or for ad-hoc queries.



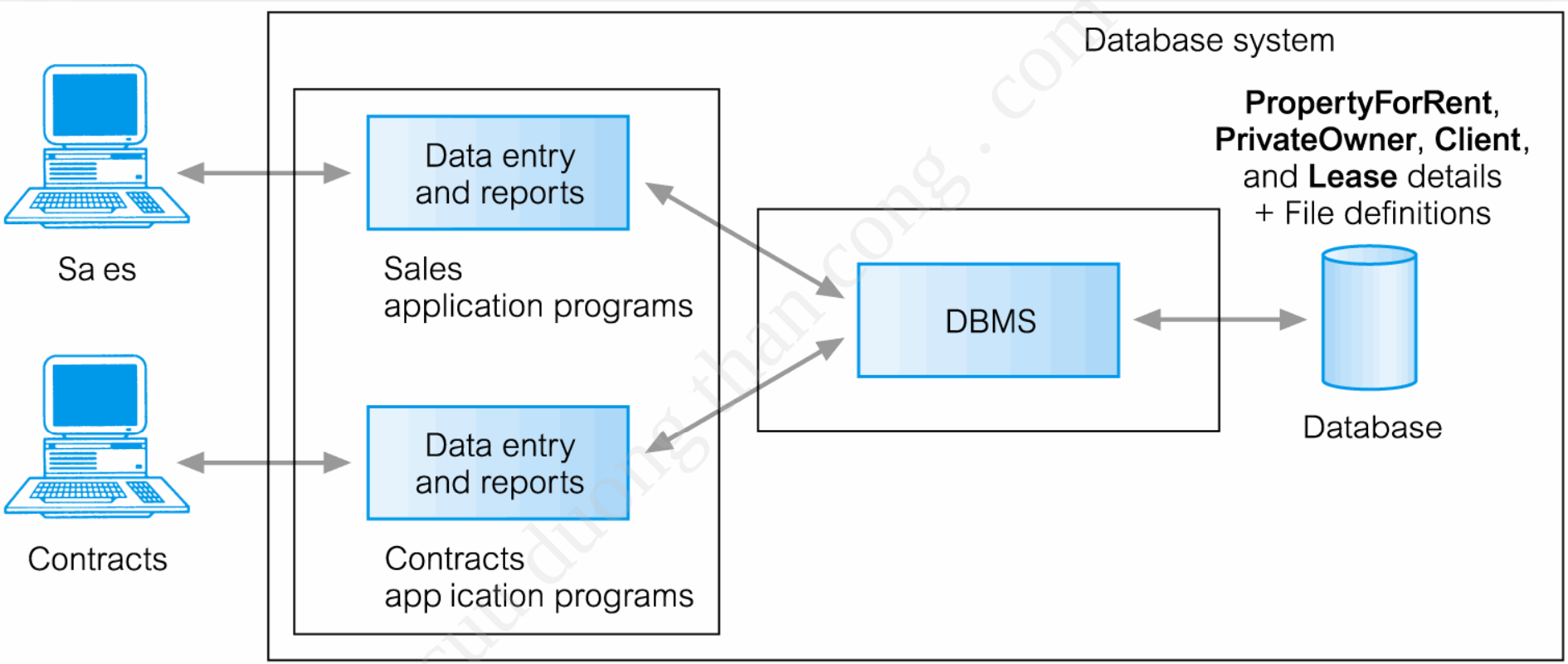


Shared file approach

Database Approach

- Why?
  - Definition of data was embedded in application programs, rather than being stored separately and independently
  - No control over access and manipulation of data beyond that imposed by application programs
- New approach:
  - **The Database and Database Management System (DBMS).**

Database Approach



PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)
PrivateOwner (ownerNo, fName, lName, address, telNo)
Client (clientNo, fName, lName, address, telNo, prefType, maxRent)
Lease (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentF nish)

Database

- Data
  - Known facts that can be recorded and that have implicit meaning
  - Information? Knowledge?
  - More: www.whatis.com
- Database: Shared collection of logically related data and a description of this data, designed to meet the information needs of an organization

Database

- **System catalog (metadata)** provides description of data to enable program–data independence.
- Logically related data comprises entities, attributes, and relationships of an organization's information.
- **D**ata**B**ase **M**anagement **S**ystem (**DBMS**): a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications (*or a software system that enables users to define, create, maintain, and control access to the database*)

DBMS

- Data Definition Language (DDL)
  - Permits specification of data types, structures and any data constraints to be stored in the database
  - All specifications are stored in the database
- Data manipulation language (DML).
  - Query language: retrieve (query), update (insert, delete, modify)
- Controlled access to database may include:
  - a security system
  - an integrity system
  - a concurrency control system
  - a recovery control system
  - a user-accessible catalog

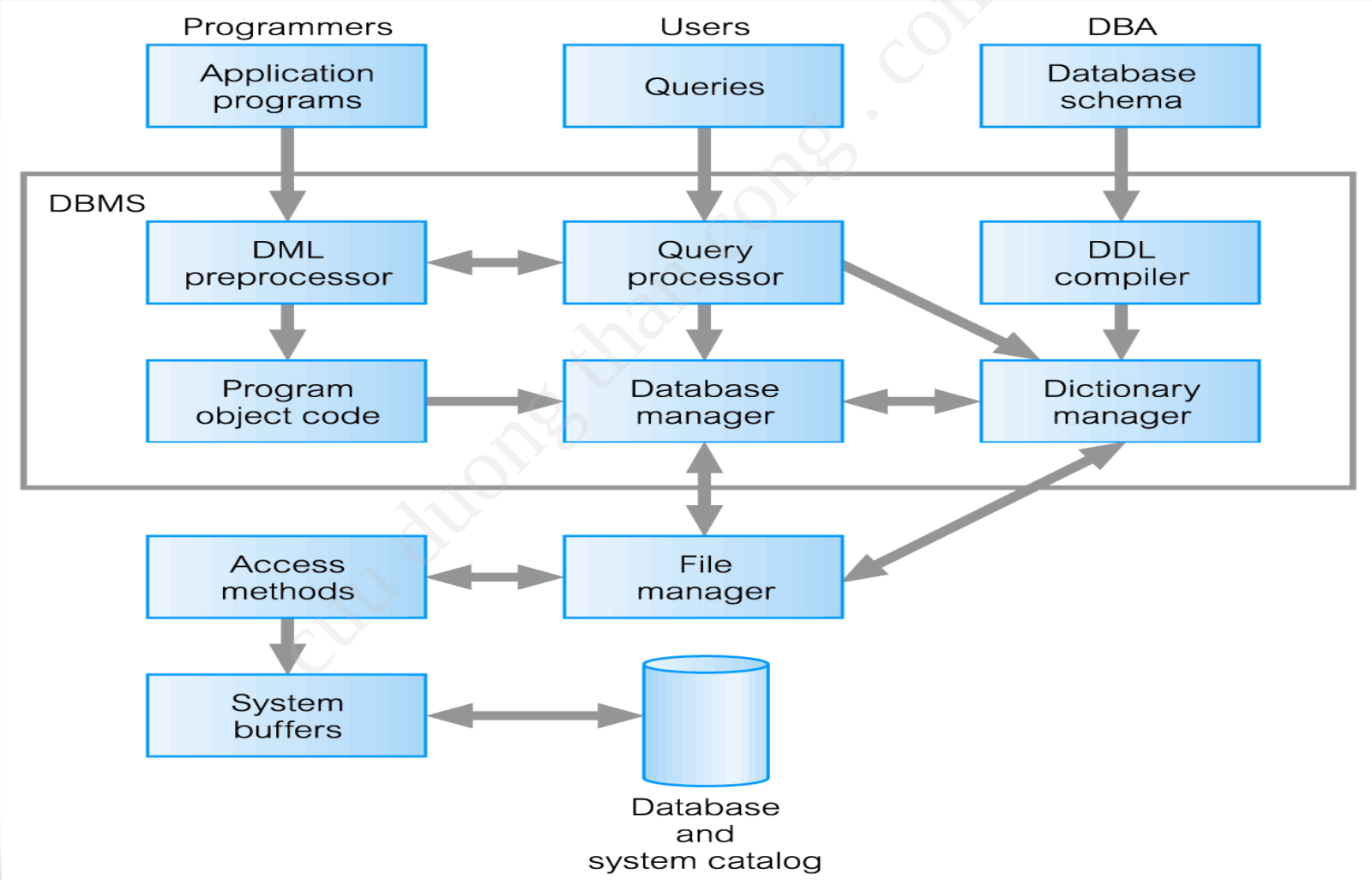
Database System

Database System = the Database + DBMS software

Database System

- Roles in the Database Environment
  - Database Administrator (DBA): responsible for
    - authorizing access to DB
    - coordinating & monitoring its use
    - acquiring software and hardware resources
    - security breach, poor response time
  - Database Designers: responsible for:
    - identifying the data to be stored in DB
    - choosing appropriate structures to represent and store this data
  - Application Programmers
  - End Users
  - …

Database System



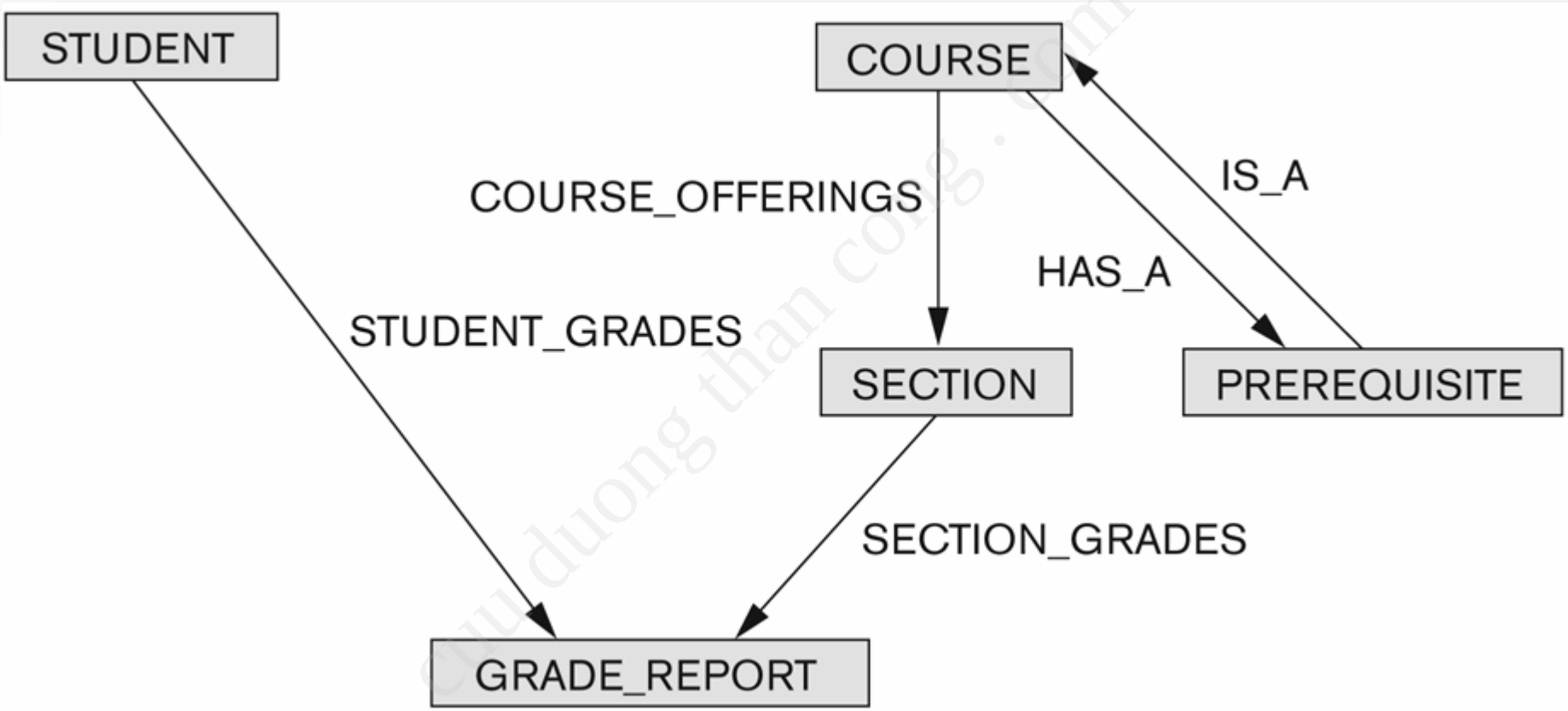
Database System

- Characteristics of the Database Approach:
  - Self-describing nature of a database system
  - Insulation between programs and data, and data abstraction
    - Program-data independence + Program-operation independence = Data abstraction
    - A data model is a type of data abstraction
  - Support of multiple views of the data
  - Sharing of data and multi-user transaction processing
  - …

Database Approach

- History of database systems
  - First generation: Hierarchical and Network
  - Second generation: Relational
  - Third generation: Object-Relational, Object-Oriented
- Brief history of database applications
  - ...

Network Model Schema: Example



Relational Model Schema: Example

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	04	King
92	CS1310	Fall	04	Anderson
102	CS3320	Spring	05	Knuth
112	MATH2410	Fall	05	Chang
119	CS1310	Fall	05	Anderson
135	CS3380	Fall	05	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

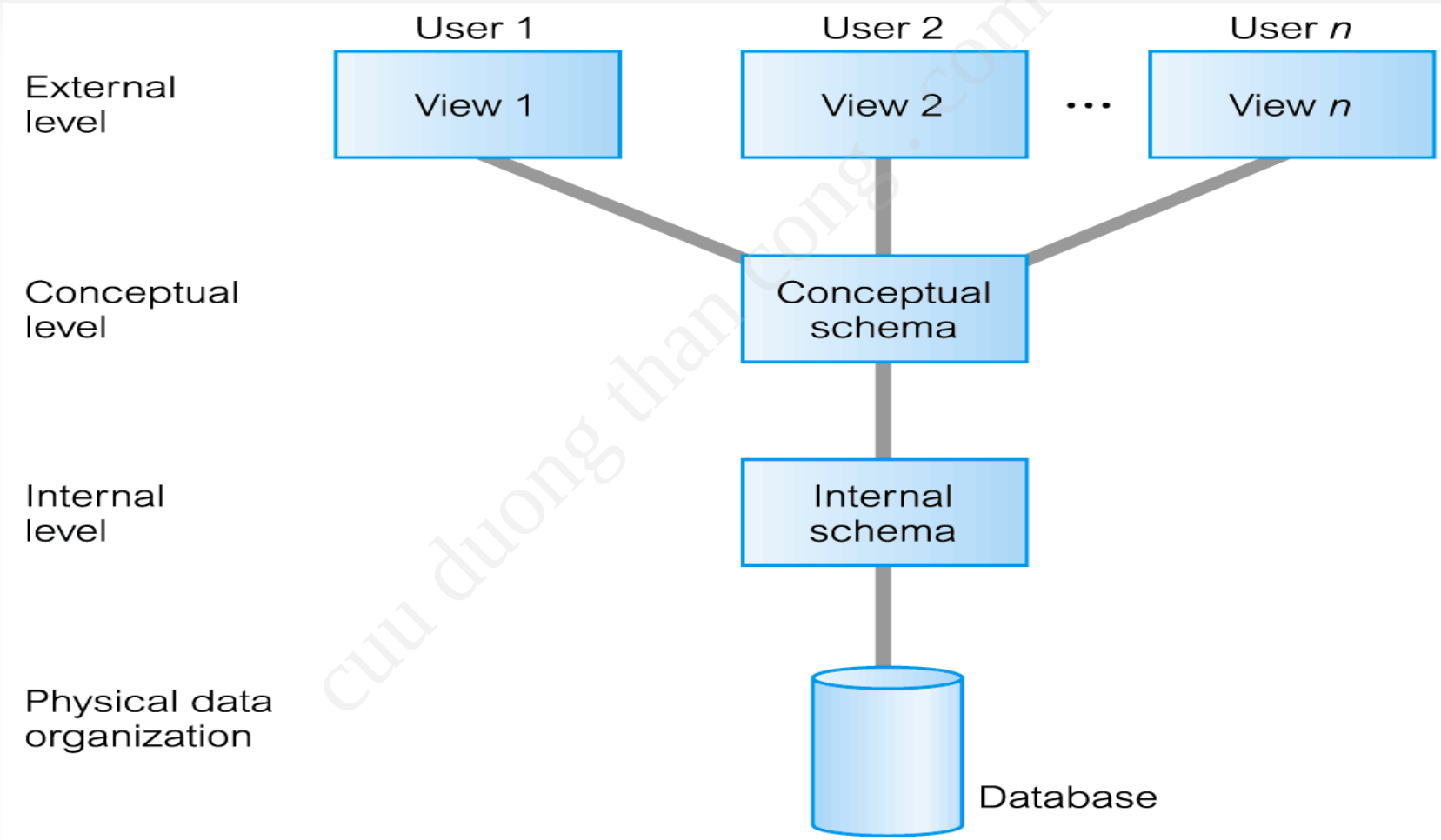
Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Three-Schema Architecture and Data Independence

Three-Schema Architecture

- Objectives of Three-Schema Architecture:
  - All users should be able to access same data
  - Users should not need to know physical database storage details
  - DBA should be able to change database storage structures without affecting the users' views
  - Internal structure of database should be unaffected by changes to physical aspects of storage
  - DBA should be able to change conceptual structure of database without affecting all users

Three-Schema Architecture



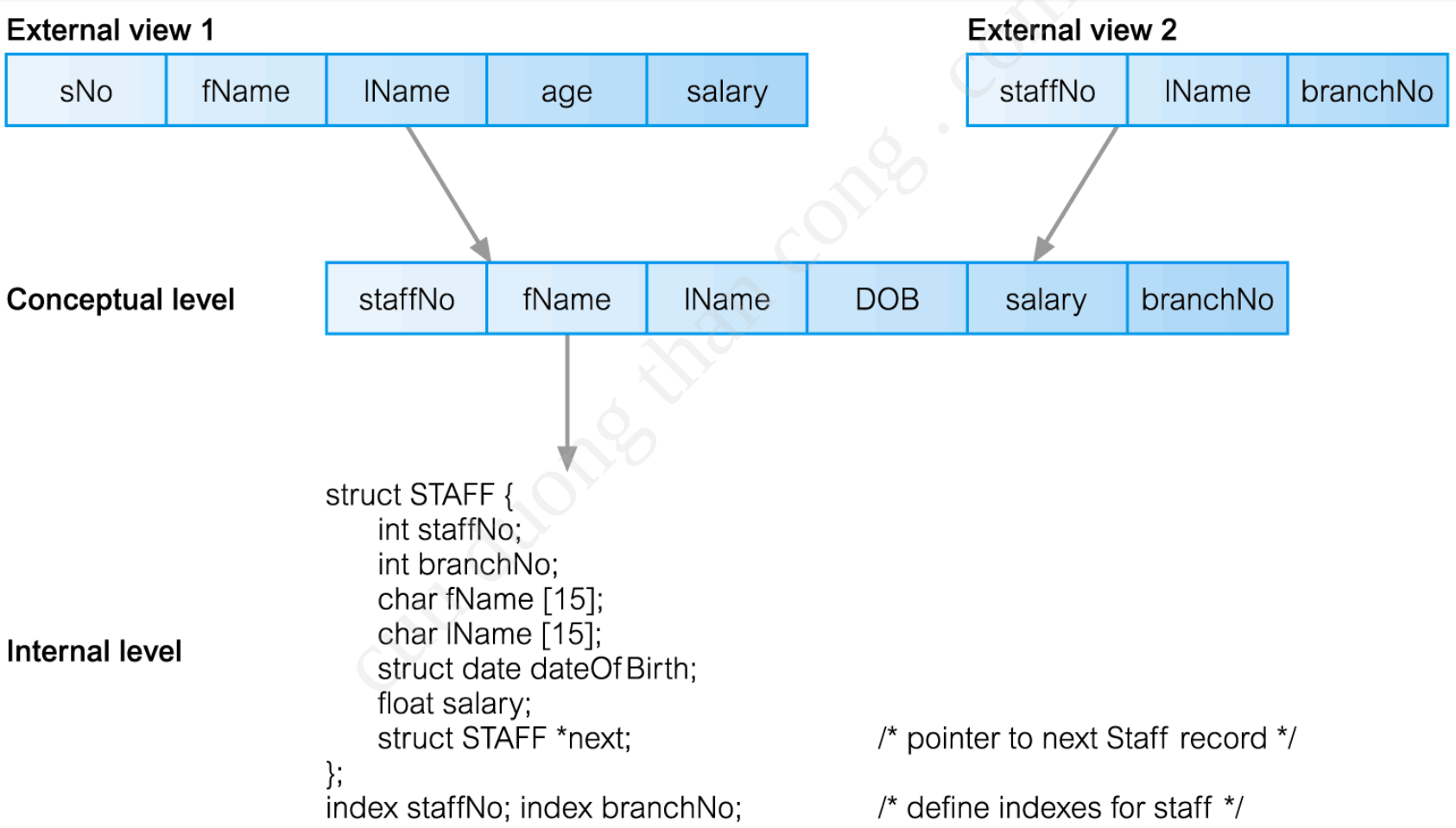
Three-Schema Architecture

- External Level
  - Users' view of the database
  - Describes that part of database that is relevant to a particular user
- Conceptual Level
  - Community view of the database
  - Describes what data is stored in database and relationships among the data

Three-Schema Architecture

- Internal Level
  - Physical representation of the database on the computer.
  - Describes how the data is stored in the database

Three-Schema Architecture



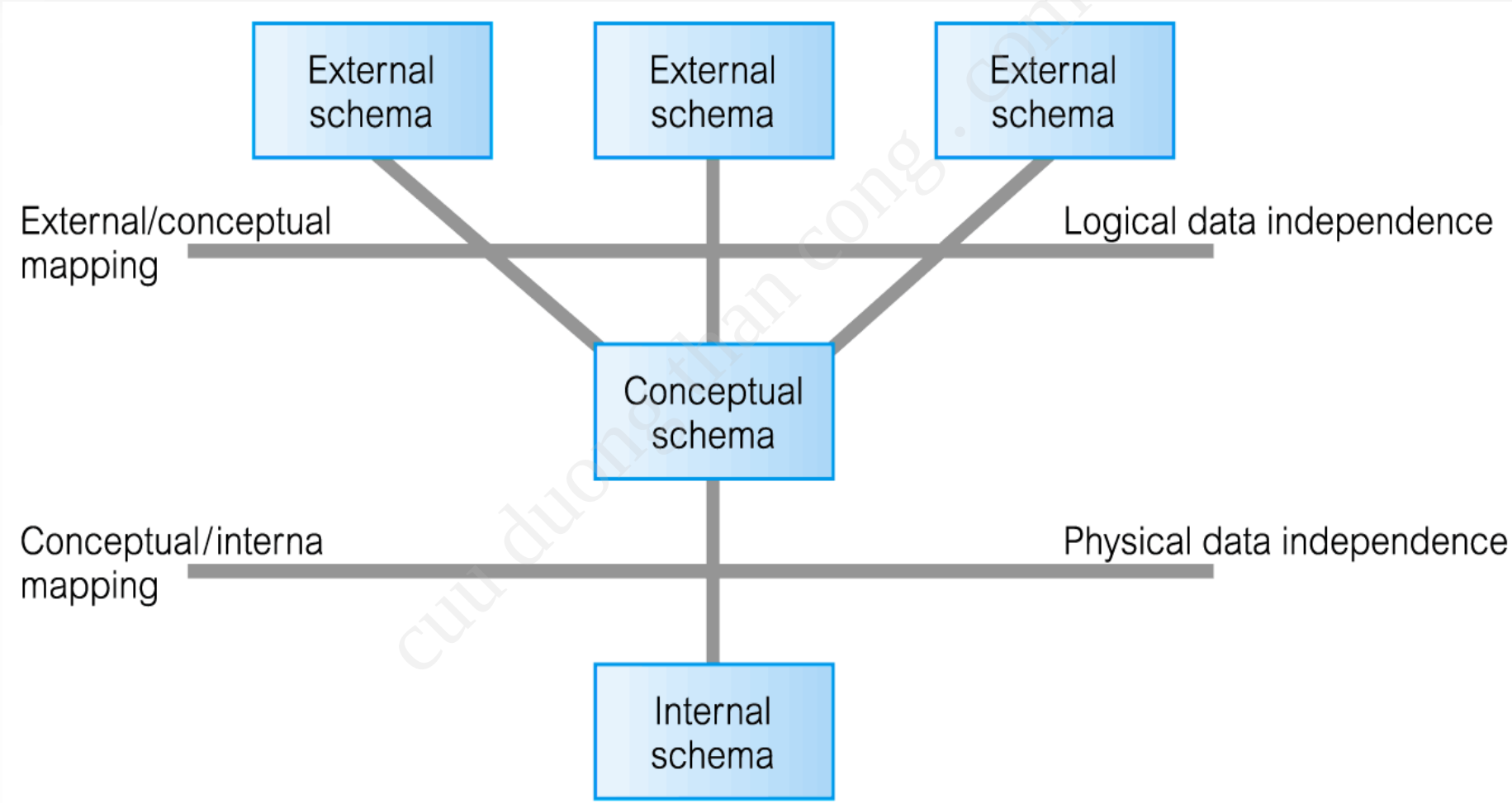
Data Independence

- **Data Independence** is the capacity to change the schema at one level of a database system without having to change the schema at the next higher levels
- **Logical Data Independence**
  - Refers to immunity of external schemas to changes in conceptual schema
  - Conceptual schema changes (e.g. addition/removal of entities) should not require changes to external schema or rewrites of application programs

Data Independence

- **Physical Data Independence**
  - Refers to immunity of conceptual schema to changes in the internal schema
  - Internal schema changes (e.g. using different file organizations, storage structures/devices) should not require changes to conceptual or external schemas

Three-Schema Architecture and Data Independence



Database Languages

Database Languages

- **Data Definition Language (DDL)** allows the DBA or user to describe and name entities, attributes, and relationships required for the application plus any associated integrity and security constraints
- **Data Manipulation Language (DML)** provides basic data manipulation operations on data held in the database
- **Data Control Language (DCL)** defines *activities that are not in the categories of those for the DDL and DML*, such as *granting privileges to users*, and defining when proposed changes to a database should be irrevocably made

Database Languages

- Procedural DML allows user to tell system exactly how to manipulate data (e.g., Network and hierarchical DMLs)
- Non-Procedural DML (declarative language) allows user to state what data is needed rather than how it is to be retrieved (e.g., SQL, QBE)
- Fourth Generation Languages (4GLs)
  - Non-procedural languages: SQL, QBE, etc.
  - Application generators, report generators, …

Data Models, Database Schema, Database State

Data Models

- Data Model: An integrated collection of concepts for describing data, relationships between data, and constraints on the data in an organization
- Categories of data models include:
  - Object-based (Conceptual)
    - ERD, Object-Oriented, …
  - Record-based (Representational)
    - Relational, Network, Hierarchical

Describe data at the conceptual & external levels

  - Physical: used to describe data at the internal level

Database Schema and Database State

- **Database Schema:** the description of a database, which is specified during database design and is not expected to change frequently
- **Schema Diagram:** a displayed schema
- **Database State (Snapshot):** the data in the database at a particular moment in time

Database Schema and Database State

STUDENT

Name	Student_number	Class	Major

COURSE

Course_name	Course_number	Credit_hours	Department

PREREQUISITE

Course_number	Prerequisite_number

SECTION

Section_identifier	Course_number	Semester	Year	Instructor