

BỘ MÔN HỆ THỐNG THÔNG TIN
KHOA CÔNG NGHỆ THÔNG TIN – ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP HCM

CƠ SỞ DỮ LIỆU NÂNG CAO

Chương 01: Tổng quan về thiết kế CSDL

Giảng viên: TS. Nguyễn Trần Minh Thư



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



Thông tin giảng viên

- Giảng viên lý thuyết:
 - TS. NGUYỄN TRẦN MINH THU'
 - Email: ntmthu@fit.hcmus.edu.vn
 - Office: l84 – (08)38324467 – 801
 - HTTP: ntmthu.net/csdInc-20th
 - User/pw truy cập: csdInc/2@CsdInc
 - PW mở tài liệu: ntmthu@khtn
- Trợ giảng/Thực hành:
 - 1) ThS. Tiết Gia Hồng
 - Email: tghong@fit.hcmus.edu.vn
 - 2) Ths. Hồ Thị Hoàng Vy
 - Email: hthvy@fit.hcmus.edu.vn
- Giao tiếp/thảo luận: Moodle , Group (Facebook)



4.0

Nội dung

- Tại sao CSDL quan trọng?
- Tại sao thiết kế CSDL quan trọng?
- Có những loại CSDL nào?
- Chu kỳ phát triển hệ thống và phát triển CSDL
- Mục tiêu, nội dung môn học
- Qui định, qui chế
- Hình thức đánh giá

4.0 Tại sao cần CSDL?

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*



Where is the data about the friends and groups stored?
Where are the “likes” stored and what would they be used for?

*On her lunch break,
she picks up her
prescription at the
pharmacy*



Where is the pharmacy inventory data stored?
What data about each product will be in the inventory data?
What data is kept about each customer and where is it stored?

*After work, Susan
goes to the grocery
store*



Where is the product data stored?
Is the product quantity in stock updated at checkout?
Does she pay with a credit card?

*At night, she plans for a trip
and buys airline tickets and
hotel reservations online*



Where does the online travel website get the airline and hotel data from?
What customer data would be kept by the website?
Where would the customer data be stored?

*Then she makes a few
online purchases*



Where are the product and stock data stored?
Where does the system get the data to generate product “recommendations” to the customer?
Where would credit card information be stored?

4.0 Tại sao cần CSDL?

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



Where is the data about the
friends and groups stored?

Where are the "likes" stored
and what would they be
used for?



On her lunch break,
she picks up her
prescription at the
pharmacy



Where is the pharmacy
inventory data stored?

What data about each
product will be in the
inventory data?

What data is kept about
each customer and where
is it stored?



After work, Susan
goes to the grocery
store



Where is the product
data stored?

Is the product quantity in
stock updated at checkout?

Does she pay with a credit
card?



At night, she plans for a trip
and buys airline tickets and
hotel reservations online



Where does the online
travel website get the
airline and hotel data from?

What customer data would
be kept by the website?

Where would the customer
data be stored?



Then she makes a few
online purchases



Where are the product
and stock data stored?

Where does the system get
the data to generate product
"recommendations" to the
customer?

Where would credit card
information be stored?



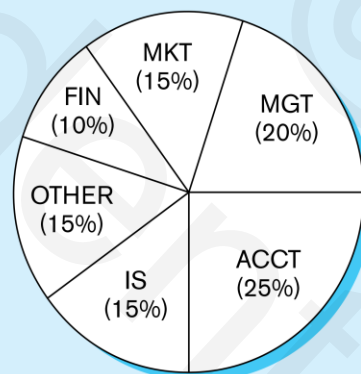
Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145

Class Roster

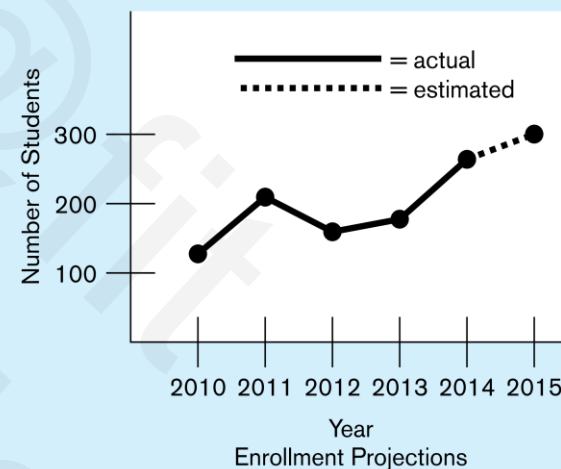
Course: MGT 500 Semester: Spring 2015
Business Policy

Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3



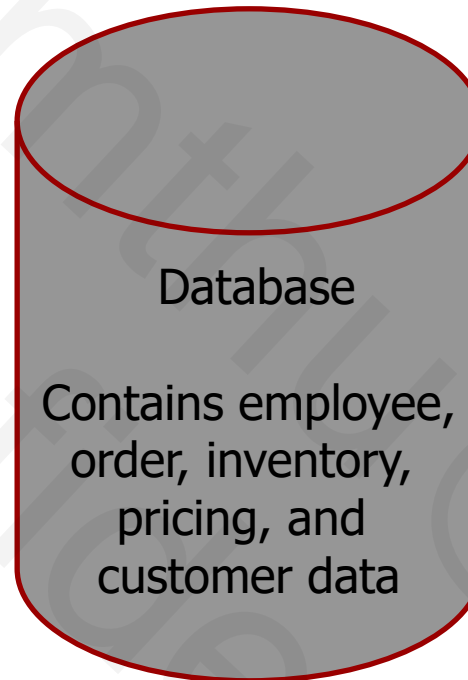
Percent Enrollment by Major (2015)





4.0

Cơ sở dữ liệu (Database)?



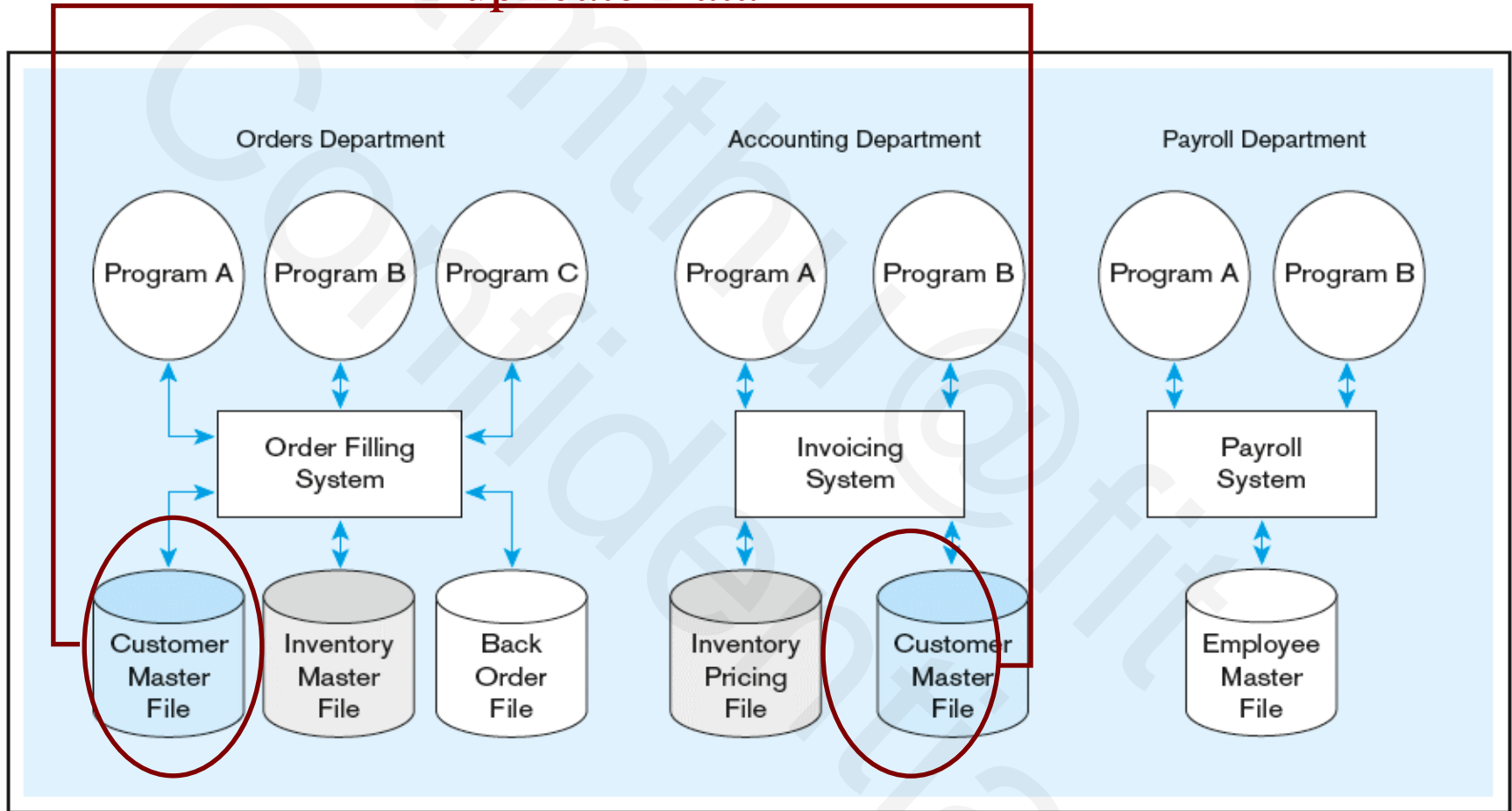
Cơ sở dữ liệu là một tập dữ liệu có cấu trúc và có sự liên quan



4.0

Tại sao cần giải pháp CSDL?

Duplicate Data



Hệ thống xử lý dựa trên file

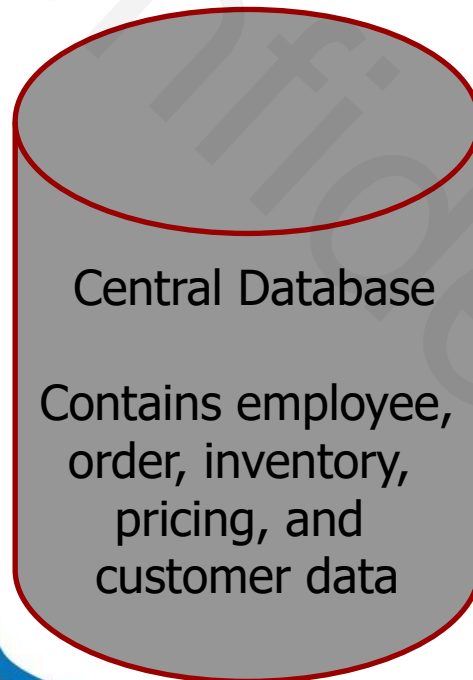
Cơ sở dữ liệu nâng cao



4.0

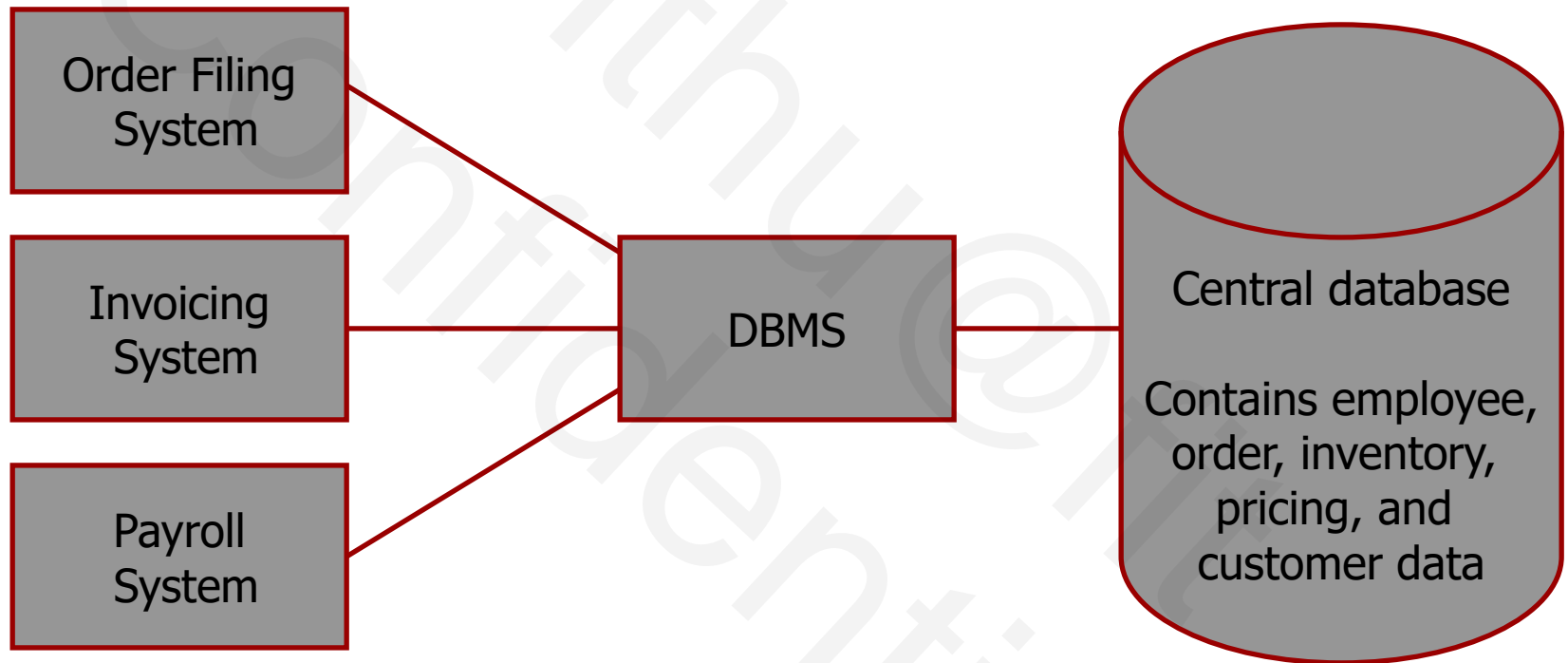
Tại sao cần giải pháp CSDL?

- Kho tập trung dữ liệu trung phục vụ cho sự chia sẻ
- Dữ liệu được quản lý bởi một tác nhân kiểm soát
- Được lưu trữ theo chuẩn, định dạng thuận tiện



Yêu cầu?
Hệ QTCSDL (Database Management System - DBMS)

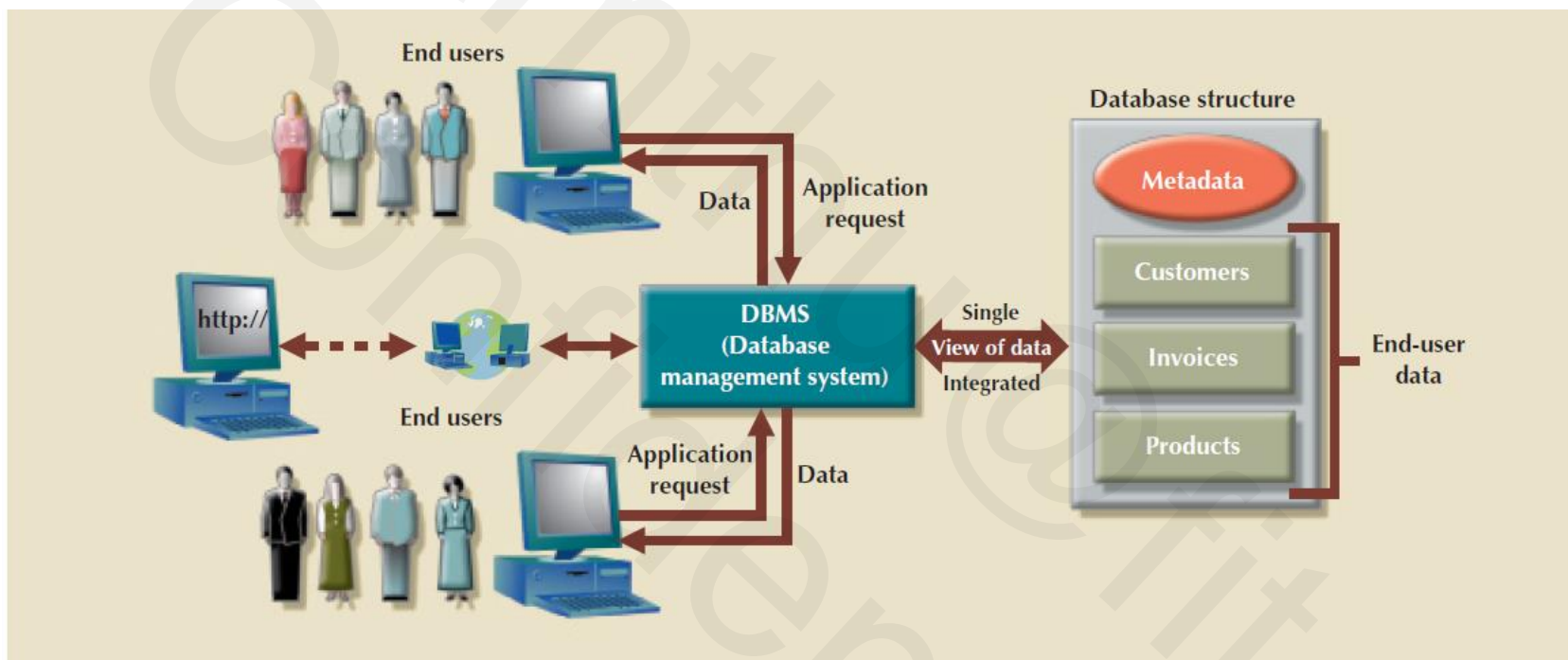
Vai trò của HQT CSDL (DBMS)



4.0 **Vai trò của DBMS**

- Đóng vai trò trung gian giữa người dùng cuối và CSDL
- Cho phép dữ liệu được chia sẻ
- Trình bày cho người dùng cuối chế độ xem dữ liệu tích hợp
- Nhận và chuyển dịch các yêu cầu ứng dụng thành các hoạt động cần thiết để thực hiện.
- Che dấu sự phức tạp bên trong với các chương trình ứng dụng và người dùng

4.0 Vai trò của DBMS



DBMS quản lý sự tương tác giữa người dùng cuối (end user) và CSDL (database)



4.0

Siêu dữ liệu (metadata) là gì?

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

Mô tả đặc tính của dữ liệu bao gồm: kiểu dữ liệu, kích thước, miền giá trị, ngữ cảnh dữ liệu



Siêu dữ liệu (metadata)

The screenshot displays the Microsoft SQL Server Management Studio interface for the 'CISORA1\GRADSQL.CORMOR - dbo.lgcustomer' database. The 'Object Explorer' on the left shows the database structure, including tables like 'dbo.lgcustomer'. The central pane shows the 'Columns' table with columns: cust_code (numeric(38, 0)), cust_name (varchar(20)), cust_iname (varchar(20)), cust_street (varchar(70)), cust_city (varchar(50)), cust_state (char(2)), cust_zip (char(5)), and cust_balance (numeric(8, 2)). The 'Column Properties' pane for 'cust_code' shows details like (Name), Allow Nulls, Data Type, Precision, Scale, and Table Designer settings. The 'Properties' pane on the right shows the table's identity and primary key settings. A blue box labeled 'Metadata' has arrows pointing to these three panes.

Column name	Data type	Allow Nulls
cust_code	numeric(38, 0)	<input type="checkbox"/>
cust_name	varchar(20)	<input type="checkbox"/>
cust_iname	varchar(20)	<input type="checkbox"/>
cust_street	varchar(70)	<input checked="" type="checkbox"/>
cust_city	varchar(50)	<input checked="" type="checkbox"/>
cust_state	char(2)	<input checked="" type="checkbox"/>
cust_zip	char(5)	<input checked="" type="checkbox"/>
cust_balance	numeric(8, 2)	<input checked="" type="checkbox"/>

Property	Value
(Name)	lgcustomer
Database Name	CORMOR
Description	
Schema	dbo
Server Name	cisora1\gradsql
Table Designer	
Identity Column	
Indexable	Yes
Lock Escalation Table	
Regular Data Sp	PRIMARY
Replicated	No
Row GUID Colu	
Text/Image File	PRIMARY

Property	Value
(Name)	cust_code
Allow Nulls	No
Data Type	numeric
Default Value or Binding	
Precision	38
Scale	0
Table Designer	
Collation	<database default>
Computed Column Specification	
Condensed Data Type	numeric(38, 0)
Description	
Deterministic	Yes
DTS-published	No
Full-text Specification	No
Has Non-SQL Server Subscriber	No
Identity Specification	No
Indexable	Yes



4.0

Có các loại CSDL nào?

- **Single-user database:** hỗ trợ 1 người dùng vào 1 thời điểm
 - **Desktop database:** chạy trên PC
- **Multiuser database:** hỗ trợ đa người dùng vào 1 thời điểm
 - **Workgroup databases:** hỗ trợ một nhóm nhỏ người dùng hoặc một phòng ban cụ thể
 - **Enterprise database:** hỗ trợ nhiều người dùng thông qua nhiều phòng ban của một công ty.



Có các loại CSDL nào?

- **Centralized database:** dữ liệu được lưu trữ ở 1 vị trí (site).
- **Distributed database:** dữ liệu được lưu trữ phân tán trên nhiều vị trí
- **Cloud database:** được tạo và duy trì bằng cách sử dụng các dịch vụ dữ liệu đám mây, nhằm cung cấp các phương pháp đo lường hiệu suất được xác định cho cơ sở dữ liệu



4.0

Có những loại CSDL nào?

- **General-purpose databases:** chứa nhiều loại dữ liệu dùng cho nhiều lĩnh vực
- **Discipline-specific databases:** chứa dữ liệu phục vụ cho một lĩnh vực cụ thể
- **Operational database:** thiết kế để hỗ trợ hoạt động hằng ngày của công ty



4.0

Có những loại CSDL nào?

- **Analytical database:** lưu trữ dữ liệu lịch sử và số liệu kinh doanh được sử dụng riêng để ra quyết định chiến lược.
 - **Data warehouse:** lưu dữ liệu ở định dạng tối ưu hóa để hỗ trợ ra quyết định.
 - **Online analytical processing (OLAP)**
 - Công cụ để truy vấn, xử lý, và mô hình hóa dữ liệu từ kho dữ liệu (data warehouse)
- **Business intelligence:** ghi nhận và xử lý dữ liệu kinh doanh để tạo ra thông tin, nhằm hỗ trợ ra quyết định.



4.0

Có những loại CSDL nào?

- **Unstructured data:** dữ liệu ở trạng thái nguyên gốc
- **Structured data:** dữ liệu đã được định dạng
 - Định dạng cấu trúc được áp dụng trên xử lý
- **Semi structured data:** xử lý một số mở rộng
- **Extensible Markup Language (XML)**
 - Thể hiện các thành phần dữ liệu ở định dạng text



Có những loại CSDL nào?

- **Modern Databases:**

- NoSQL (Not only SQL): Key-Value, Document, Graph, Column Family
- Time Series Database
- Realtime Database
- Spatial Database
- Mobile Database



4.0

Hệ thống thông tin (Information System) ?



Hệ thống thông tin: thu thập dữ liệu, lưu trữ, chuyển đổi, truy xuất dữ liệu. Bao gồm:

- Con người, phần cứng, phần mềm
- Cơ sở dữ liệu, chương trình ứng dụng, các thủ tục

Tại sao thiết kế CSDL quan trọng?

Why are there blanks in rows 9 and 10?

How to produce an alphabetical listing of employees?

How to count how many employees are certified in Basic Database Manipulation?

Is Basic Database Manipulation the same as Basic DB Manipulation?

What if an employee acquires a fourth certification?

Do we add another column?

ID	ENum	Name	Title	HireDate	Skill1	Skill1Date	Skill2	Skill2Date	Skill3	Skill3Date
1	02345	Brian Oates	DBA	2/14/1995	Basic Database Management	2/14/2002	Advanced Database Management	2/14/2005	Basic Web Design	8/9/2003
2	08273	Marco Bienz	Analyst	7/28/2006	Basic Web Design	3/8/2009	Advance Process Modeling	8/19/2012		
3	06234	Jasmine Patel	Programmer	8/10/2005	Basic Web Design	8/10/2007	Advanced C# programming	8/10/2007	Basic DB manipulation	1/29/2012
4	03373	Franklin Johnson, Jr.	Purchasing Agent	3/15/2002	Advanced Spreadsheets	6/20/2011				
5	13567	Almond, Robert	Analyst	9/30/2012	Basic Process Modeling	9/30/2014	Basic Database Design	5/23/2015		
6	10282	Richardson, Amanda	Clerk	4/11/2011						
7	09382	Susan Mathis	Database Programmer	8/2/2010	Basic DB Design	8/2/2012	Basic Database Manipulation	8/2/2012	Advanced DB Manipulation	5/1/2013
8	14311	Duong, Lee	Programmer	9/1/2014	Basic Web Design	9/1/2016				
9					Master Database Programming					
10					Basic Spreadsheets					
11	09002	Wade Gaither	Clerk	5/20/2010	Advanced Spreadsheets	5/16/2013	Basic Web Design	5/16/2013		
12	13383	Raymond F. Matthews	Programmer	3/12/2012	Basic C# Programming	3/12/2014				
13	09283	Chavez, Juan	Clerk	7/4/2010						
14	04893	Patricia Richards	DBA	6/11/2004	Advanced Database Management	6/11/2006	Advanced Database Manipulation	9/20/2012		
15	13932	Lee, Megan	Programmer	9/29/2013						

**Thiết kế chứng nhận kỹ năng nhân viên
(thiết kế tồi)**

Tại sao thiết kế CSDL quan trọng?

Table name: EMPLOYEE

Employee_ID	Employee_FName	Employee_LName	Employee_HireDate	Employee_Title
02345	Johnny	Jones	2/14/1995	DBA
03373	Franklin	Johnson	3/15/2002	Purchasing Agent
04893	Patricia	Richards	6/11/2004	DBA
06234	Jasmine	Patel	8/10/2005	Programmer
08273	Marco	Bienz	7/28/2006	Analyst
09002	Ben	Joiner	5/20/2010	Clerk
09283	Juan	Chavez	7/4/2010	Clerk
09382	Jessica	Johnson	8/2/2010	Database Programmer
10282	Amanda	Richardson	4/11/2011	Clerk
13383	Raymond	Matthews	3/12/2012	Programmer
13567	Robert	Almond	9/30/2012	Analyst
13932	Megan	Lee	9/29/2013	Programmer
14311	Lee	Duong	9/1/2014	Programmer

Table name: CERTIFIED

Employee_ID	Skill_ID	Certified_Date
02345	100	2/14/2002
02345	110	8/9/2003
02345	180	2/14/2005
03373	120	6/20/2011
04893	180	6/11/2006
04893	220	9/20/2012
06234	110	8/10/2007
06234	200	8/10/2007
06234	210	1/29/2012
08273	110	3/8/2009
08273	190	8/19/2012
09002	110	5/16/2013
09002	120	5/16/2013
09382	140	8/2/2012
09382	210	8/2/2012
09382	220	5/1/2013
13383	170	3/12/2014
13567	130	9/30/2014
13567	140	5/23/2015
14311	110	9/1/2016

Table name: SKILL

Skill_ID	Skill_Name	Skill_Description
100	Basic Database Management	Create and manage database user accounts.
110	Basic Web Design	Create and maintain HTML and CSS documents.
120	Advanced Spreadsheets	Use of advanced functions, user-defined functions, and macroing.
130	Basic Process Modeling	Create core business process models using standard libraries.
140	Basic Database Design	Create simple data models.
150	Master Database Programming	Create integrated trigger and procedure packages for a distributed environment.
160	Basic Spreadsheets	Create single tab worksheets with basic formulas
170	Basic C# Programming	Create single-tier data aware modules.
180	Advanced Database Management	Manage Database Server Clusters.
190	Advance Process Modeling	Evaluate and Redesign cross-functional internal and external business processes.
200	Advanced C# Programming	Create multi-tier applications using multi-threading
210	Basic Database Manipulation	Create simple data retrieval and manipulation statements in SQL.
220	Advanced Database Manipulation	Use of advanced data manipulation methods for multi-table inserts, set operations, and correlated subqueries.

**Thiết kế chứng nhận kỹ năng nhân viên
(thiết kế tốt)**



4.0

Phân loại thiết kế CSDL

- **Dữ liệu đang tồn tại**

- Dữ liệu từ bảng tính hoặc tập tin
- Rút trích dữ liệu từ cơ sở dữ liệu khác
- Thiết kế sử dụng nguyên lý dạng chuẩn

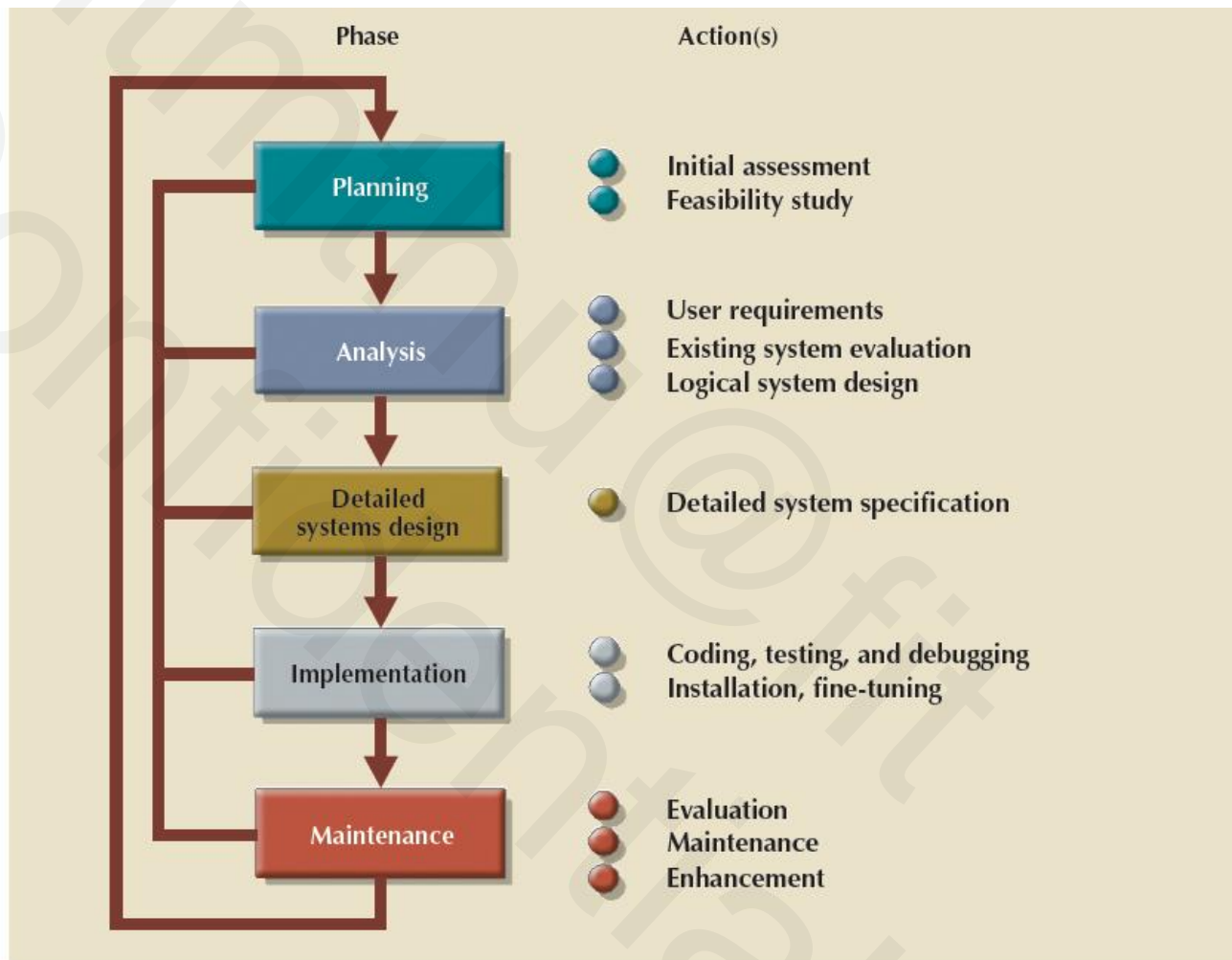
- **Phát triển các hệ thống mới**

- Phân tích mô hình dữ liệu từ yêu cầu của ứng dụng
- Thiết kế dữ liệu phục vụ khai thác.

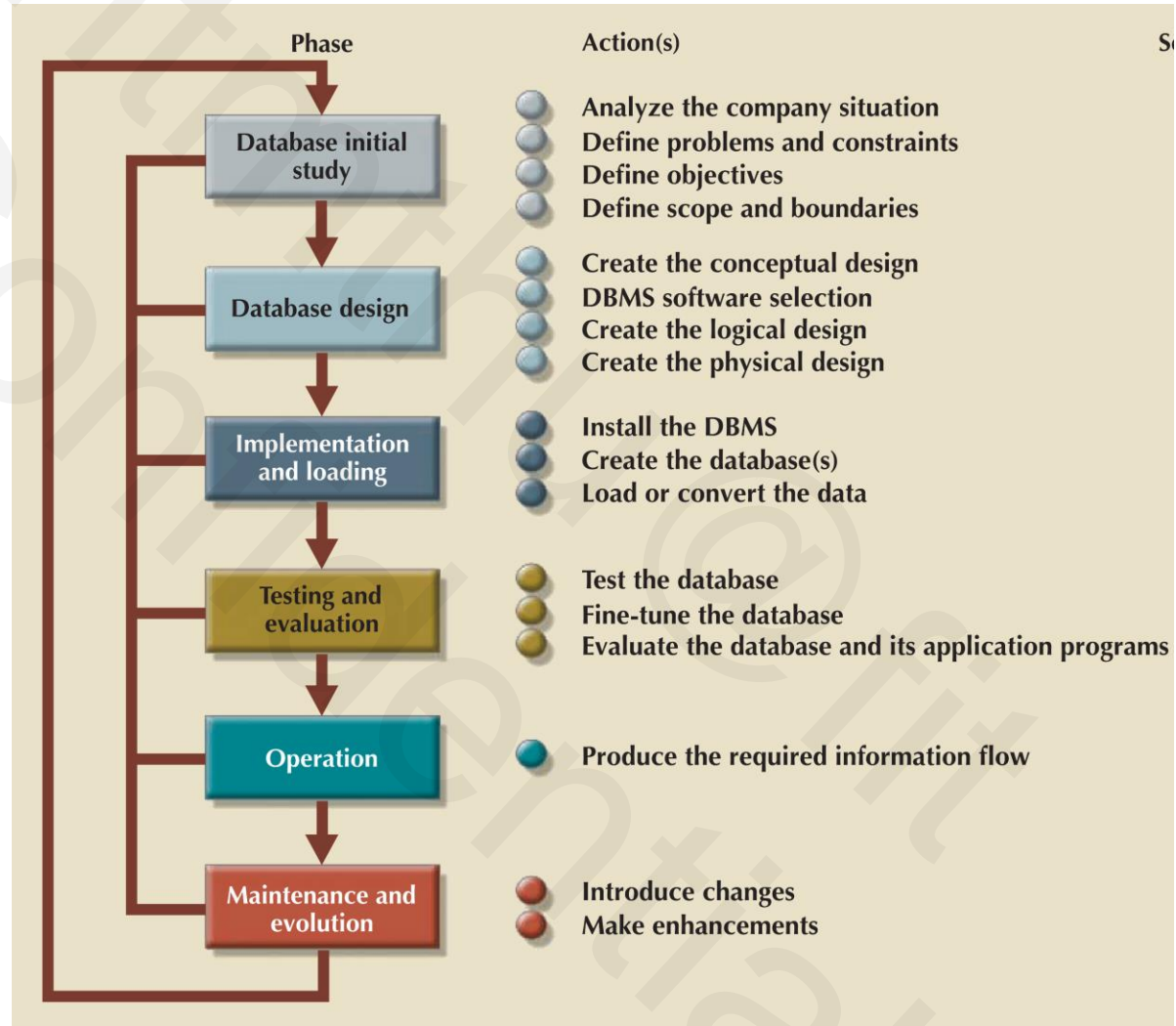
- **Tái thiết kế cơ sở dữ liệu**

- Chuyển cơ sở dữ liệu cũ sang cơ sở dữ liệu mới hơn.
- Tích hợp hai hoặc nhiều CSDL lại.
- Đưa ra các kỹ thuật và thiết kế sử dụng cho CSDL mới.
- Áp dụng nguyên lý dạng chuẩn và qui tắc chuyển đổi mô hình.

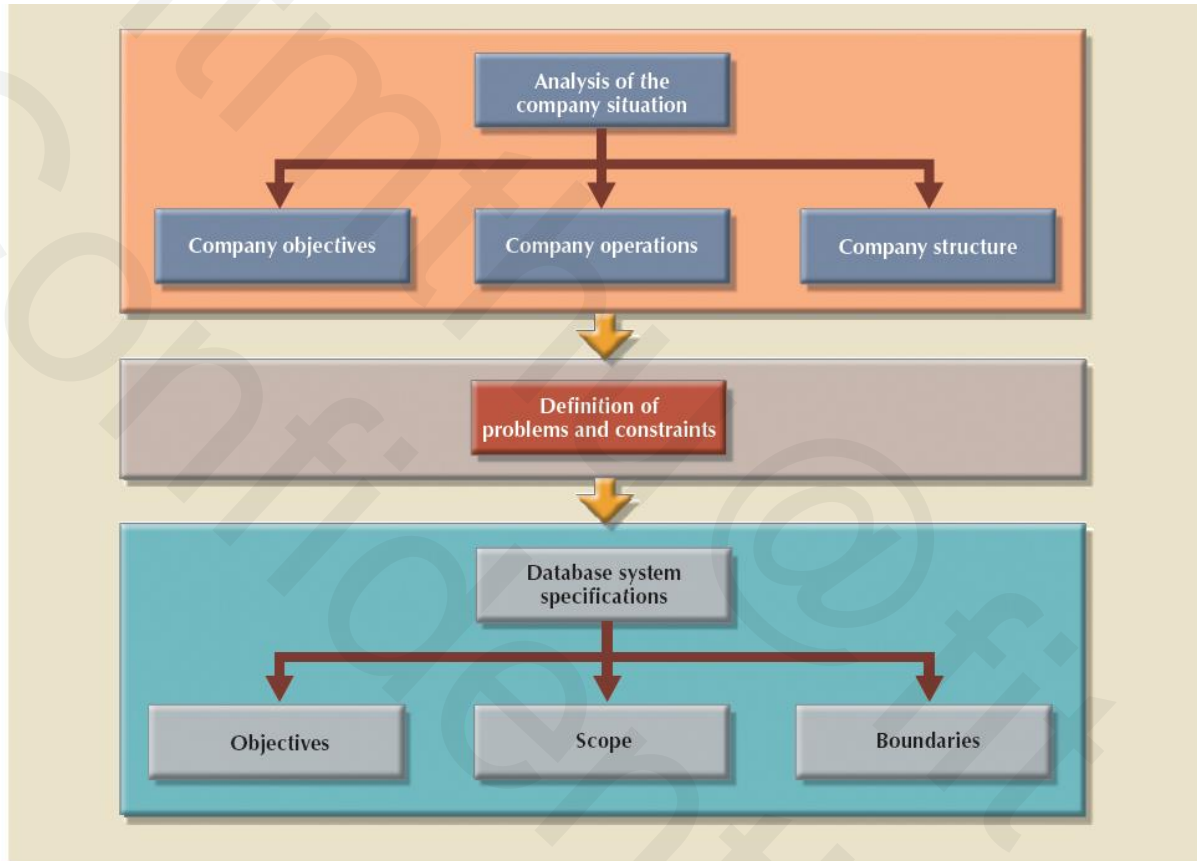
Systems Development Life Cycle (SDLC)



The Database Life Cycle (DBLC)

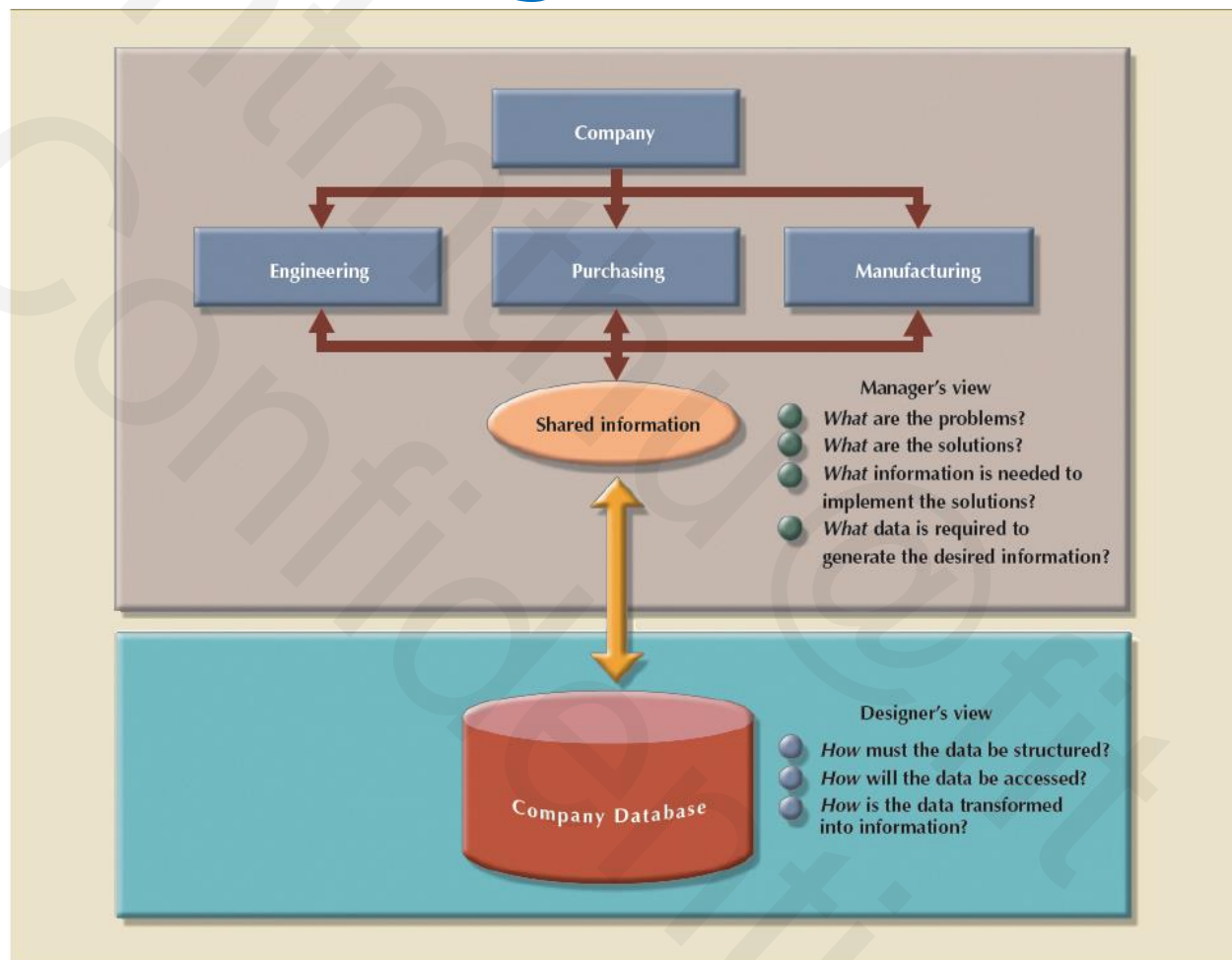


Purpose of Database Initial Study



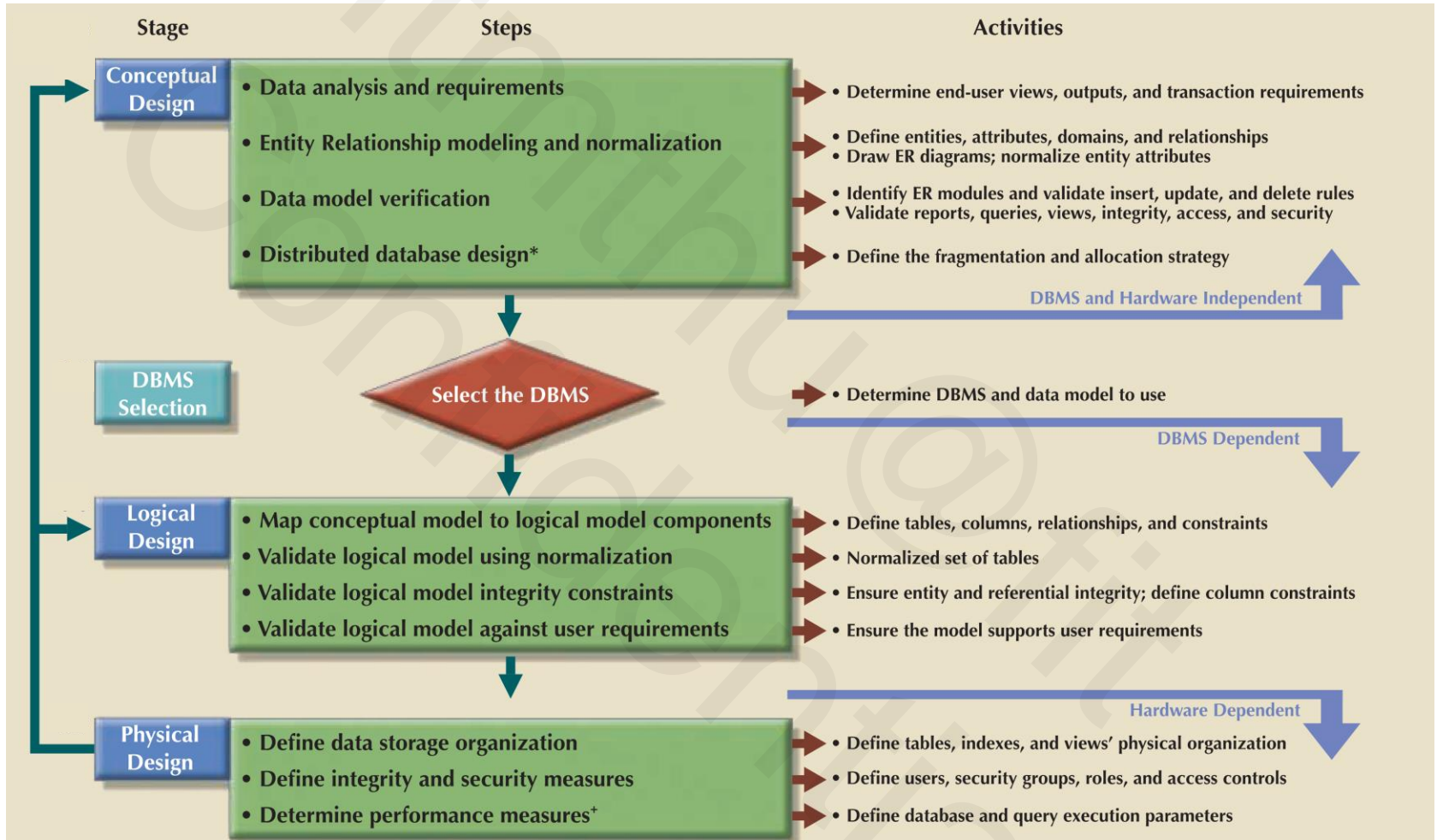
Sơ đồ các hoạt động trong điều tra khởi tạo CSDL

- Hỗ trợ các xử lý và mục tiêu của công ty
- Là phase quan trọng nhất
 - Đảm bảo sản phẩm cuối cùng phải đáp ứng được yêu cầu của người dùng và các yêu cầu hệ thống.
- Điểm quan trọng để kiểm tra sự hoàn thiện của các thủ tục
 - Thành phần dữ liệu là một thành phần quan trọng của toàn hệ thống
 - Các nhà phân tích hệ thống/lập trình thiết kế các thủ tục để chuyển đổi dữ liệu thành thông tin.
 - Thiết kế CSDL là một quy trình lặp lại.

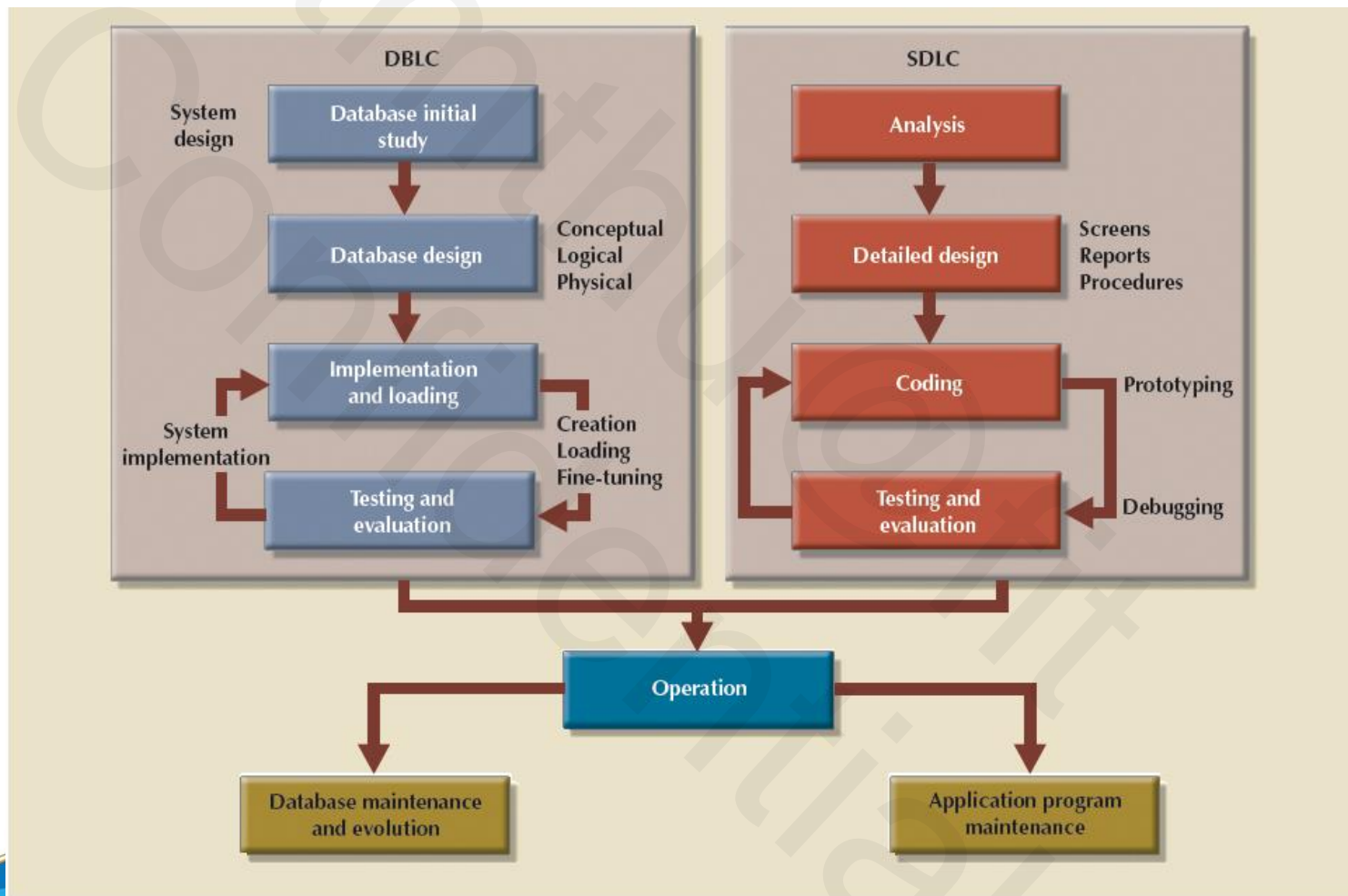


Hai cách nhìn về dữ liệu: người quản lý & người thiết kế

Database Design Process



Parallel Activities in the DBLC and the SDLC





4.0 Mục tiêu môn học

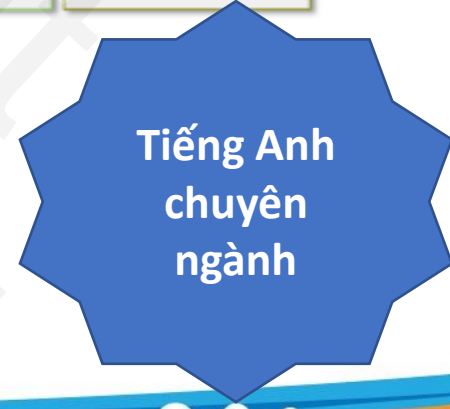
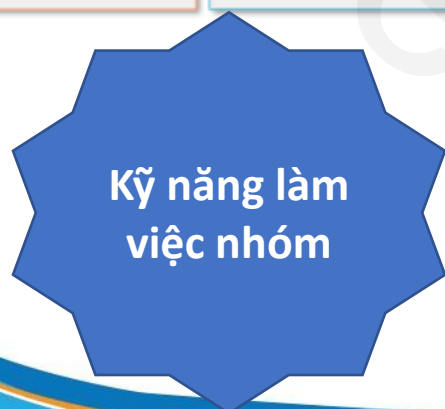
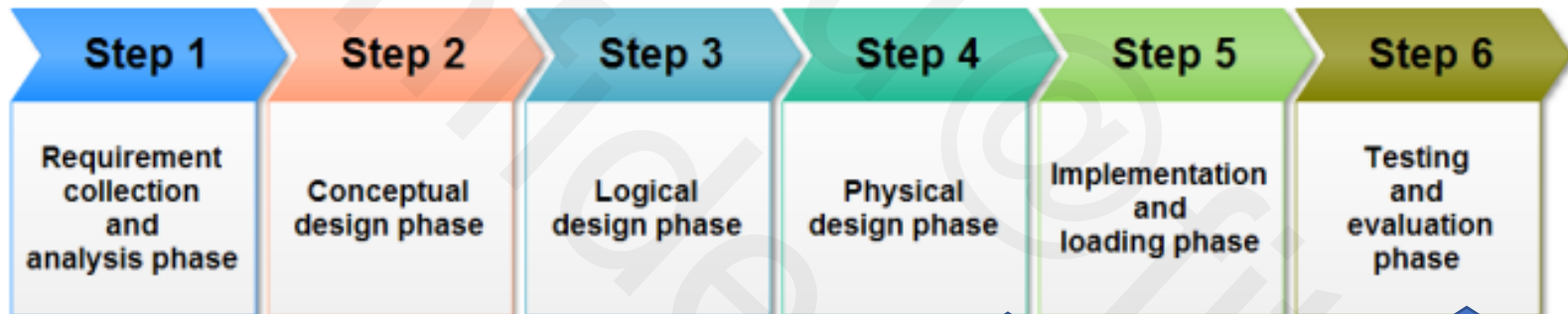
- Biết các vấn đề đặt ra cho chuyên viên thiết kế CSDL quan hệ
- Vận dụng kiến thức CSDL nâng cao để giải quyết một số vấn đề trong từng giai đoạn thiết kế: quan niệm, lô-gíc, vật lý
- Hiểu biết thêm một số loại CSDL khác: CSDL hướng đối tượng, CSDL trên Internet và XML , NoSQL, CSDL thông tin địa lý,...



Kiến thức đạt được

- Biết sử dụng mô hình thực thể kết hợp (ER) và Mô hình dữ liệu quan hệ.
- Hiểu được lý thuyết về dạng chuẩn và áp dụng để đánh giá chất lượng cho một lược đồ dữ liệu được phân tích, thiết kế.
- Biết thiết kế một lược đồ cơ sở dữ liệu ở mức vật lý hiệu quả phục vụ khai thác dữ liệu.
- Biết cài đặt một cơ sở dữ liệu trong một hệ quản trị cơ sở dữ liệu cụ thể MS SQL 200x.
- Biết thêm về một số loại CSDL khác.

Kiến thức đạt được





Nội dung

- Chương 1: Tổng quan về TK CSDL quan hệ
- Chương 2: Mô hình dữ liệu (ER, EER, RDM)
- Chương 3: Phụ thuộc hàm và dạng chuẩn
- Chương 4: Thiết kế dữ liệu mức quan niệm
- Chương 5: Thiết kế dữ liệu mức logic và vật lý
- Chương 6: Các phương pháp truy xuất dữ liệu
- Chương 7: Một số loại CSDL khác



4.0

Đánh giá

• LÝ THUYẾT

- Bài tập hằng tuần (BTTL, BTVN), bài tập trợ giảng và kiểm tra định kỳ:
- Thi lý thuyết cuối kỳ:
 - Thi tự luận/trắc nghiệm
 - Không sử dụng tài liệu
 - Thời gian : 90 - 120 phút

• THỰC HÀNH

- Đồ án, bài tập thiết kế CSDL:
- Lập trình, khai thác dữ liệu (HQT MS SQL, .NET):
- TH Nhóm SV (2-3 SV/nhóm).
- Làm và nộp sản phẩm/bài tập theo tuần.



4.0

Đánh giá

#	Nội dung	Loại	Hình thức	Tỉ lệ	GVPT
Test#01	Phân tích mô hình dữ liệu ở mức quan niệm	Hoạt động, Trợ giảng Thực hiện: Cá nhân (Quiz/Essay)	Kiểm tra sau khi học kiến thức liên quan	5%	HTHVy
Test#02	Phân tích mô hình dữ liệu mở mức logic	Hoạt động, Trợ giảng Thực hiện: Cá nhân (Quiz/Essay)	Kiểm tra sau khi học kiến thức liên quan	5%	HTHVy
Test#03	Phân tích dữ liệu ở mức vật lý	Hoạt động lý thuyết Thực hiện: Cá nhân (Quiz/Essay)	Kiểm tra sau khi học kiến thức liên quan	5%	NTMThư
Test#04	Bài thi lý thuyết theo yêu cầu.	Bài thi tổng kết LTCK Thực hiện: Cá nhân (Quiz/Essay)	Kiểm tra sau khi học kiến thức liên quan	40%	NTMThư
Homework /In Class	Thực hiện các yêu cầu bài tập về nhà/tại lớp LT	BTVN- LT Thực hiện: Cá nhân/Nhóm	Thực hiện và nộp bài trên hệ thống theo yêu cầu	5%	NTMThư
Test#05	Lập trình CSDL: + Viết SQL + Trigger, Store Procedure, Functions + Index, Partition	TH Thực hiện: theo Cá nhân	Thi trực tiếp qua Internet	10%	HTHVy, TGHồng
Project	- Phân tích dữ liệu đầy đủ theo các bước (quan niệm, logic, vật lý) cho một ứng dụng HTTT - Cài đặt một số tính năng yêu cầu	Đồ án TH: Thực hiện: theo Nhóm	Đồ án, và nộp và chấm vấn đáp	30%	HTHVy, TGHồng



4.0

Hình thức xử phạt gian lận

Phạm vi xem xét

- Các bài tập tuần, các bài kiểm tra định kỳ.
- Bài tập thực hành
- Đồ án môn học
- Báo cáo môn học
- Bài thi cuối kỳ

Mức độ xử phạt

- Vi phạm lần đầu: nhận 0 điểm của môn học.
- Vi phạm từ lần thứ hai (trong cùng năm học): nhận 0 điểm môn học. Tùy theo mức độ vi phạm sẽ xử lý bổ sung các hình thức kỷ luật từ mức cảnh cáo, đình chỉ một học kỳ cho đến buộc thôi học.

*****Tham chiếu qui chế xử phạt gian lận Khoa đã ban hành**



Qui định điểm thành phần/tổng kết

- Bất kỳ sao chép nào được phát hiện: 0 điểm
 - Bài thi lý thuyết
 - Bài thực hành
 - Bài tập tại lớp, quá trình
- **Một trong các cột điểm qui định (QT, TH, LT) bị 0 điểm thì điểm tổng kết là 0 điểm**
- Không giải quyết bất kỳ trường hợp nộp bài trễ hạn.



Tài liệu tham khảo

1. Jeff Hoffer, Ramesh Venkataraman, Heikki Topi, Modern Database System, 12th edition.
2. Carlos Coronel, Steven Morris, Database System: Design, Implementation, and Management, 12th edition.
3. Toby Teorey, Sam Lightstone, Tom Nadeau, *Database Modeling and Design - Logical Design*, 4th Edition
4. Sam Lightstone, Toby Teorey, Tom Nadeau, *Physical Database Design*, 4th Editio.
5. TapioLahdenmaki, MichaelLeach, *Relational Database Index Design and the Optimizers*, 2005
6. Toby Teorey, *Database Modeling & Design: The Fundamental Principles* 2nd Edition, Morgan Kaufmann Publishers, Inc. San Francisco, California.
7. Raghu Ramarkrishnan, Johannes Gehrke, *Database Management Systems* 3rd Edition, 2002.
8. Silberschatz, Korth, Sudarshan, *Database System Concepts*, 4th Edition , 2004
9. Elmasri Navathe, *Fundamentals of Database Systems* 4th Edition, 2006

