

Korea IT School 2022

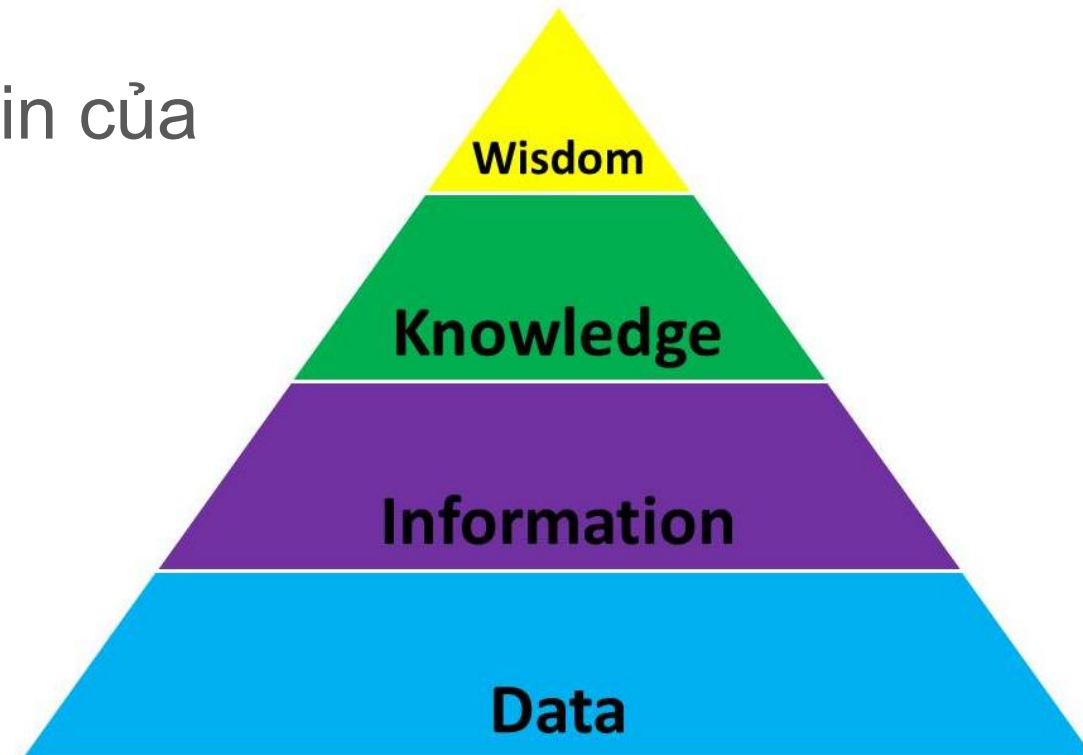
# CƠ SỞ DỮ LIỆU

Unit 1: Giới thiệu Hệ thống cơ sở dữ liệu

# Nội dung

1. Dữ liệu và Thông tin
2. CSDL là gì? Một số loại CSDL?
3. Tầm quan trọng của thiết kế dữ liệu
4. Các CSDL hiện đại được phát triển từ hệ thống tệp
5. Một số vấn đề trong quản lý hệ thống dữ liệu tệp tin
6. Các thành phần chính trong hệ CSDL
7. Các chức năng chính của Hệ quản trị CSDL

- Một quyết định tốt yêu cầu phải có thông tin tốt được lấy từ thực tế.
- Dữ liệu được quản lý hiệu quả nhất khi được lưu trong cơ sở dữ liệu (CSDL).
- CSDL được phát triển từ hệ thống tệp tin của máy tính.
- Sự hiểu biết về CSDL là rất quan trọng



- Từ lúc sinh ra → chết đi, chúng ta đã sản sinh ra rất nhiều dữ liệu
- Ứng dụng trong hầu hết các lĩnh vực đòi hỏi về quản trị dữ liệu:
  - Kinh doanh
  - Nghiên cứu
  - Quản trị
- **Làm sao có thể xử lý được nhiều dữ liệu? Làm sao có thể lưu trữ để sau đó lấy lại khi cần thiết?**
- Cơ sở dữ liệu giải quyết nhiều vấn đề gặp phải trong quản trị dữ liệu.



# Tại sao cần cơ sở dữ liệu?

FIGURE 1.1 THE PERVASIVE NATURE OF DATABASES

## A Day In Susan's Life

See how many databases she interacts with each day

Trước khi đi làm  
vào tài khoản  
Facebook & Twitter



Vào buổi trưa, cô  
ấy nhận đơn thuốc  
tại quầy thuốc



Sau giờ làm việc,  
cô ấy đến cửa hàng  
tạp hóa



Buổi tối, lên kế hoạch  
cho chuyến đi: mua vé  
và đặt phòng



Cuối ngày, thực  
hiện mua hàng trực  
tuyến



- Vây dữ liệu về bạn bè và nhóm được lưu trữ ở đâu?
- Like lưu trữ ở đâu và để làm gì?



- Dữ liệu về kho hàng được lưu trữ ở đâu?
- Dữ liệu nào về từng sản phẩm sẽ có trong dữ liệu kiểm kê?
- DL nào lưu lại khi KH mua hàng?



- Sản phẩm được lưu trữ ở đâu?
- Số lượng sản phẩm trong kho có được cập nhật lúc thanh toán không?
- Thanh toán bằng gì?



- Website du lịch lấy dữ liệu về khách sạn và hàng không ở đâu?
- Dữ liệu nào của KH sẽ được lưu trữ?
- DL KH sẽ lưu ở đâu?

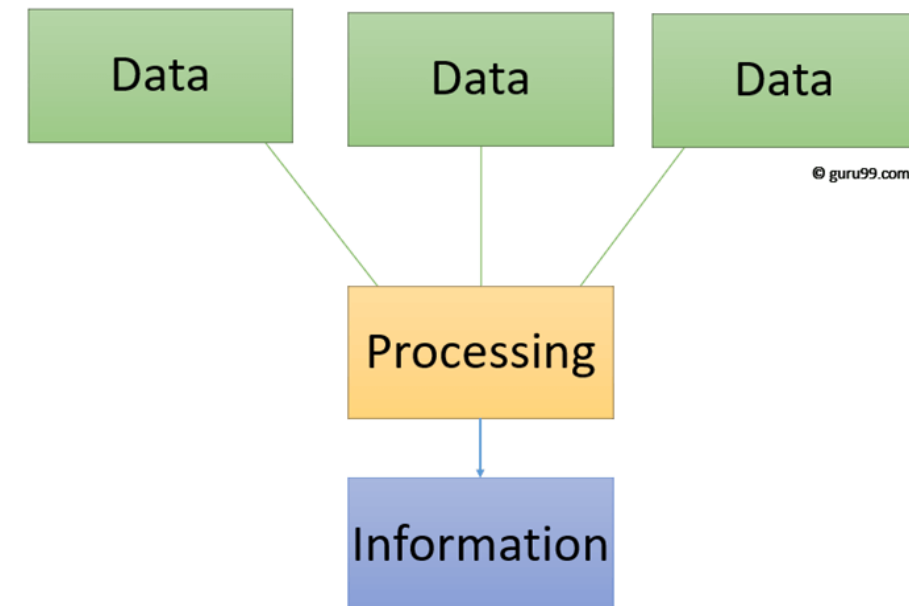


- DL về sản phẩm và kho hàng được lưu ở đâu?
- HT lấy DL từ đâu để "Gợi ý" cho khách hàng?
- Thông tin thẻ tín dụng được lưu ở đâu?



# Dữ liệu vs Thông tin

- Dữ liệu là các sự kiện thô được lưu trữ trên thiết bị lưu trữ
- Thông tin là kết quả của việc xử lý dữ liệu thô thành thể hiện có ý nghĩa.
- Thông tin cần ngữ cảnh để trình bày ý nghĩa.
- Dữ liệu thô cần định dạng để lưu trữ, xử lý và trình bày.
- Dữ liệu là nền tảng của thông tin và tri thức.



© guru99.com

- Thông tin được lấy từ xử lý dữ liệu.
- Thông tin để thể hiện ý nghĩa của dữ liệu.
- Thông tin chính xác, phù hợp, kịp thời là chìa khóa để đưa ra quyết định tốt.
- Quyết định tốt là chìa khóa cho sự tồn tại của tổ chức.
- Quản lý dữ liệu: tập trung vào **lưu trữ** và **tìm kiếm** dữ liệu.

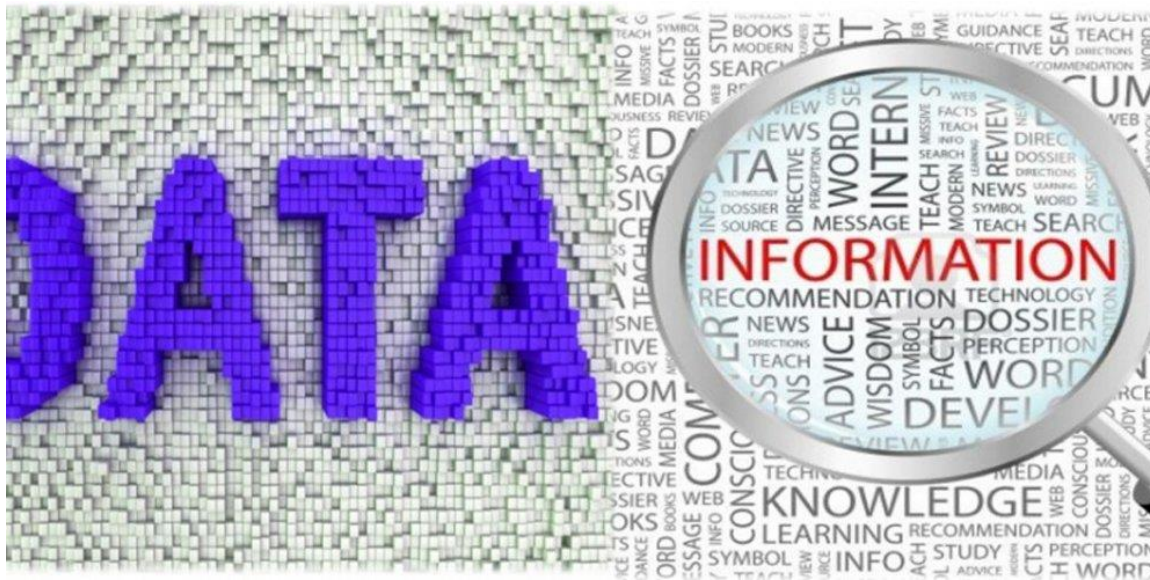




FIGURE 1.2 TRANSFORMING RAW DATA INTO INFORMATION

a) Data entry screen

**Middle Tennessee State University**

You are viewing the College website

Home Reports Activity Applications Summer Session Information About Us Contact Us

Jennings A. Jones College of Business

Home Manage Members Add Faculty Form

DO NOT append School ID (SIT) to Member ID

Member ID \* (Password will be initially sent to the email address below)

First name or initial \*

Mobile number \*

Last name \*

Chair Head: ☐ Yes ☐ No

Department \*

Area \*

Email \*

Hire Term \*

Member Default Status: Changing the status here changes only the default that is pulled into the upcoming teaching schedule. No change the historical status of members and to see your changes reflected in the various reports, all the teaching schedule preferences

Qualification:

☐ Participating ☐ Supporting

☐ Participates in the governance of the school

☐ Considered to be a long term member

High Degree:

Year Awarded:

Rank:

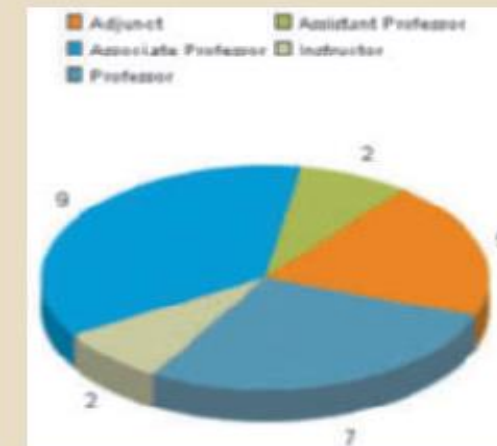
b) Raw data

ID	Last Name	First Name	Dept Code	Office	Email	Rank	Hire Year	Degree
1	Wright	George	MGMT	N135	gwright@mtsu.edu	Professor	2011	Ph.D.
2	Adams	John	PSY	N013	jadams@mtsu.edu	Professor	1994	Ph.D.
3	Jefferson	Thomas	ECOM	N013	tjefferson@mtsu.edu	Instructor	2002	M.S.A.
4	Madison	James	PSY	N035	jmadison@mtsu.edu	Associate Professor	1994	Ph.D.
5	Adams	John	ACCT	N401	jadams@mtsu.edu	Associate Professor	1995	Ph.D.
6	Adams	John	ACCT	N401	jadams@mtsu.edu	Associate Professor	1995	Ph.D.
7	Adams	John	ECOM	N013	jadams@mtsu.edu	Associate Professor	1995	Ph.D.
8	Van Horn	Maria	PSY	N035	mvhorn@mtsu.edu	Professor	1998	Ph.D.
9	Harris	William	MGMT	N135	wharris@mtsu.edu	Professor	1994	Ph.D.
10	Tyler	John	MGMT	N135	jtyler@mtsu.edu	Associate Professor	2008	Ed.D.
11	Pink	Cheryl	MGMT	N140	cpink@mtsu.edu	Associate Professor	2007	Ph.D.
12	Taylor	Zachary	ACCT	N401	ztaylor@mtsu.edu	Associate Professor	1996	Ph.D.
13	Fleming	Michael	JOB	N013	mefleming@mtsu.edu	Professor	1992	Ph.D.
14	Pierce	Franklin	MGMT	N135	fpierce@mtsu.edu	Instructor	2005	M.S.A.
15	Blackman	Yvonne	MGMT	N140	yblackman@mtsu.edu	Associate Professor	1996	D.B.A.
16	Lincoln	Wendy	MGMT	N135	wlincoln@mtsu.edu	Associate Professor	1996	Ph.D.
17	Johnson	Andrew	SVS	N035	ajohnson@mtsu.edu	Professor	1997	Ph.D.
18	Gent	Kate	MGMT	N120	kgent@mtsu.edu	Associate Professor	1999	D.B.A.
19	Rutherford	Hayes	ACCT	N401	hrutherford@mtsu.edu	Professor	1992	Ph.D.
20	Graham	Denise	ACCT	N401	dgraham@mtsu.edu	Associate Professor	2016	Ph.D.
21	Adair	Emily	ACCT	N401	eadair@mtsu.edu	Associate Professor	2017	J.D.
22	Owens	Robert	ACCT	N401	rowens@mtsu.edu	Associate Professor	1997	Ph.D.
23	Harris	Blair	BLA	N401	bharris@mtsu.edu	Associate Professor	2011	J.D.
24	McIntyre	Brian	SVS	N035	bmcintyre@mtsu.edu	Adjunct	1994	M.S.
25	Popovich	Henry	MGMT	N135	hpopovich@mtsu.edu	Associate Professor	2002	Ph.D.
26	Wiley	Lois	BOEN	N440	lwiley@mtsu.edu	Professor	1992	Ph.D.
27	Harding	William	MGMT	N135	wharding@mtsu.edu	Professor	1994	Ed.D.
28	Condit	Colin	ECOM	N013	ccondit@mtsu.edu	Professor	1995	Ph.D.
29	Hobbs	Lisa	MGMT	N013	lhobbs@mtsu.edu	Adjunct	1999	M.S.A.
30	Thorn	Buffy	ACCT	N401	bthorn@mtsu.edu	Professor	1971	Ed.D.
31	Adams	Robert	BOEN	N440	radams@mtsu.edu	Professor	2001	Ph.D.

c) Information in summary format

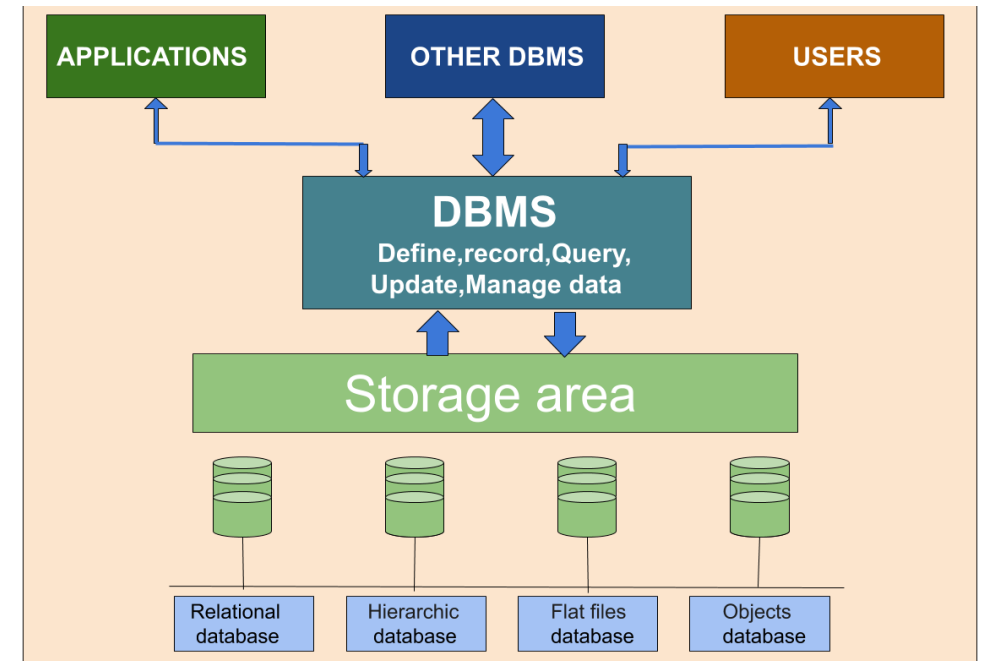
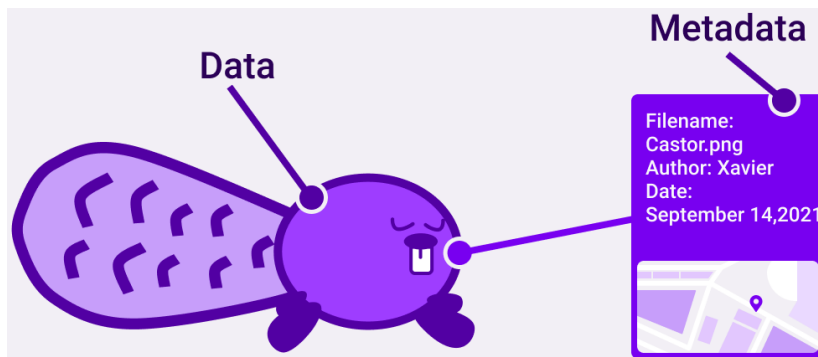
Rank	COUNT	%INF	TOT/COL	%COL. TOT.	%COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



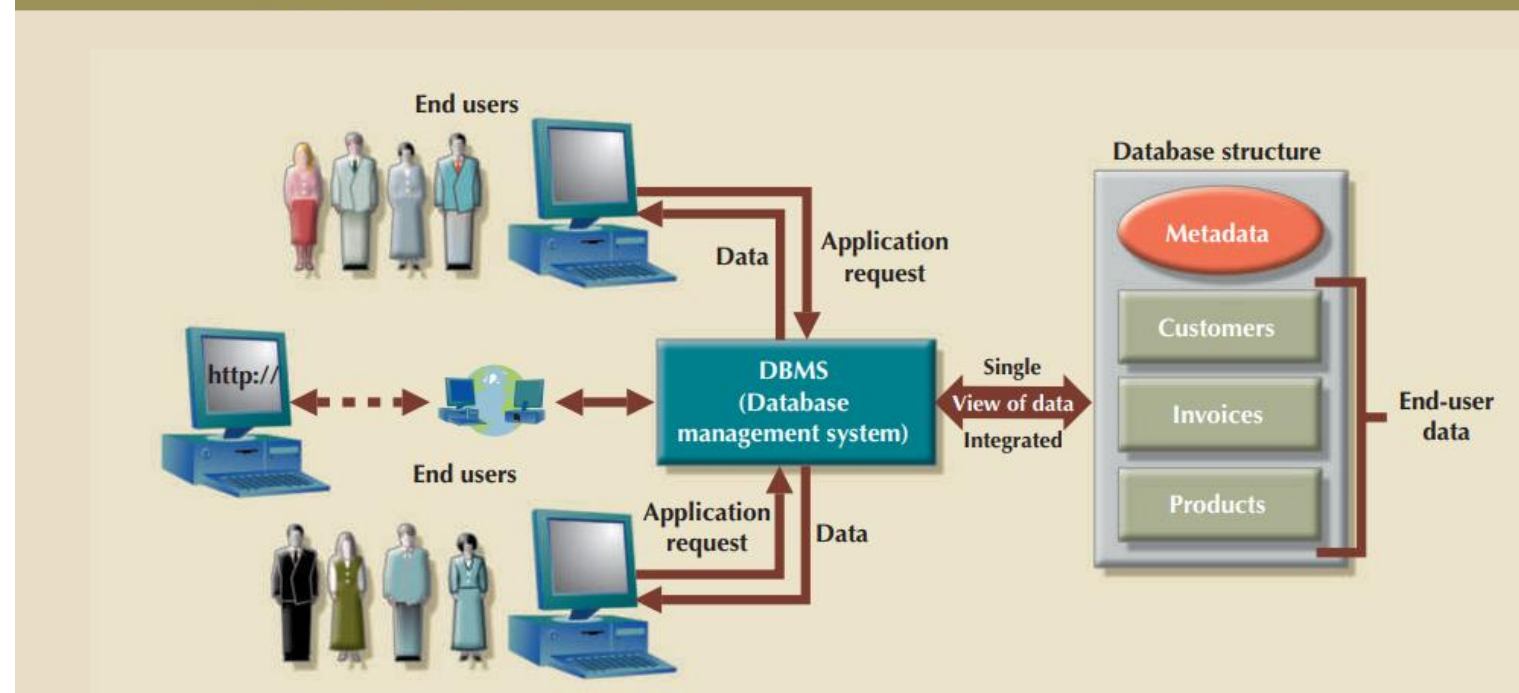


- CSDL: chia sẻ, tích hợp các cấu trúc máy tính để lưu trữ
  - Dữ liệu người dùng cuối: dữ liệu thô của người dùng cuối
  - Siêu dữ liệu: dữ liệu về dữ liệu, qua đó dữ liệu người dùng cuối được tích hợp và quản lý.
    - Cung cấp mô tả về các đặc tính dữ liệu và mối quan hệ trong dữ liệu.
    - Bổ sung và mở rộng giá trị dữ liệu.
- Hệ thống quản lý cơ sở dữ liệu (DBMS):
  - Là chương trình để quản lý cấu trúc và kiểm soát truy cập dữ liệu.



- Là trung gian giữa người dùng và cơ sở dữ liệu:
  - Lưu trữ cấu trúc CSDL giống như tập hợp các tệp
  - Có thể truy cập các tệp thông qua DBMS
- Cho phép chia sẻ dữ liệu
- Tích hợp nhiều khung nhìn của người dùng về dữ liệu

FIGURE 1.4 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE



- Cải thiện chia sẻ dữ liệu
- Cải thiện bảo mật dữ liệu
- Tích hợp dữ liệu tốt hơn
- Tối thiểu hóa sự xung đột dữ liệu
- Cải thiện truy cập dữ liệu
- Cải thiện việc ra quyết định
- Làm tăng hiệu suất của người dùng
- Dự phòng và Khôi phục



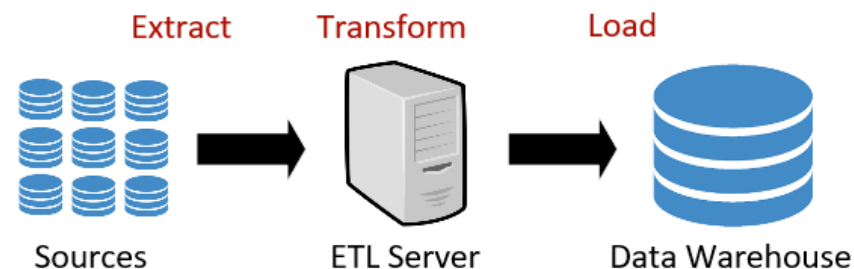
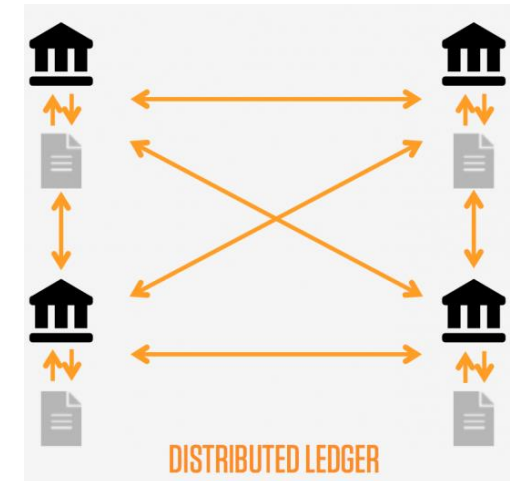
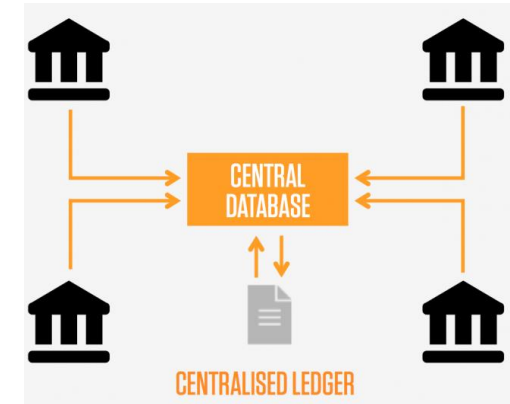
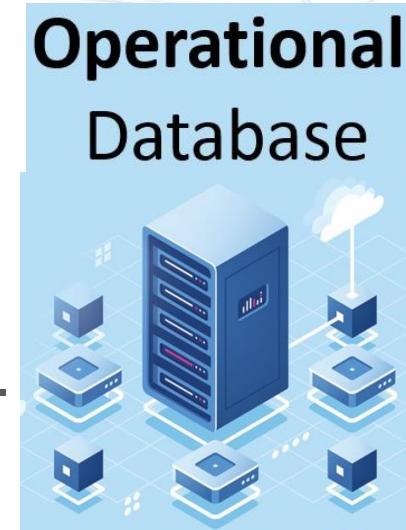


- CSDL có thể phân loại theo:
  - Số lượng người dùng: Cá nhân hay nhiều người dùng
  - Vị trí CSDL: tập trung hay phân tán
    - Hạ tầng vận hành: Hệ điều hành, mạng
    - CSDL đám mây: Microsoft Azure, Amazon AWS,...
  - Theo nhu cầu và mở rộng của sử dụng
- CSDL 1 người dùng:
  - Hỗ trợ 1 người dùng tại 1 thời điểm
  - CSDL cài trên PC cho 1 người dùng
- CSDL nhiều người dùng:
  - Hỗ trợ nhiều người dùng tại cùng 1 thời điểm
  - CSDL doanh nghiệp & nhóm

# Các loại Cơ sở dữ liệu

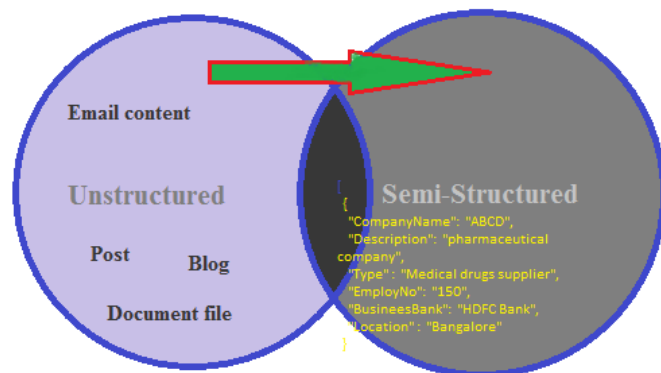
multicampus

- CSDL tập trung:
  - Đặt CSDL tại 1 trạm
- CSDL phân tán:
  - CSDL được phân tán tại một vài trạm khác nhau.
- CSDL hoạt động:
  - Hỗ trợ các hoạt động hàng ngày của công ty.
  - CSDL giao dịch hoặc sản xuất
- Kho dữ liệu:
  - Lưu trữ dữ liệu để sử dụng cho các quyết định.



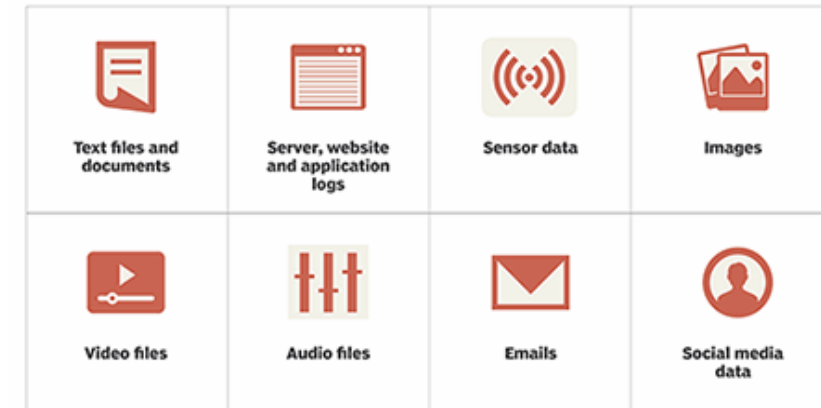
# Các loại Cơ sở dữ liệu

- Dữ liệu phi cấu trúc:
  - Tồn tại ở trạng thái nguyên bản, chưa định dạng.
- Dữ liệu có cấu trúc:
  - Là kết quả của định dạng theo yêu cầu.
- Dữ liệu bán cấu trúc:
  - Là dữ liệu đã được xử lý 1 phần.
- XML:
  - Trình bày phần tử dữ liệu trong định dạng văn bản.



```
<!-- Schema Components -->
<xs:complexType name="baseComponent">
  <xs:complexContent> [29 lines]
</xs:complexType>
<xs:complexType name="componentWithFacets">
  <xs:complexContent>
    <xs:extension base="baseComponent"> [3 lines]
  </xs:complexContent>
</xs:complexType>
<xs:element name="schema">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="baseComponent">
        <xs:attribute name="type" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
</xs:schema>
```

## Unstructured data types



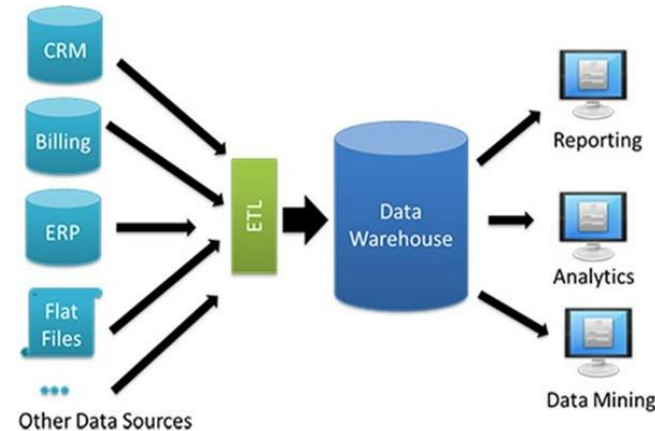
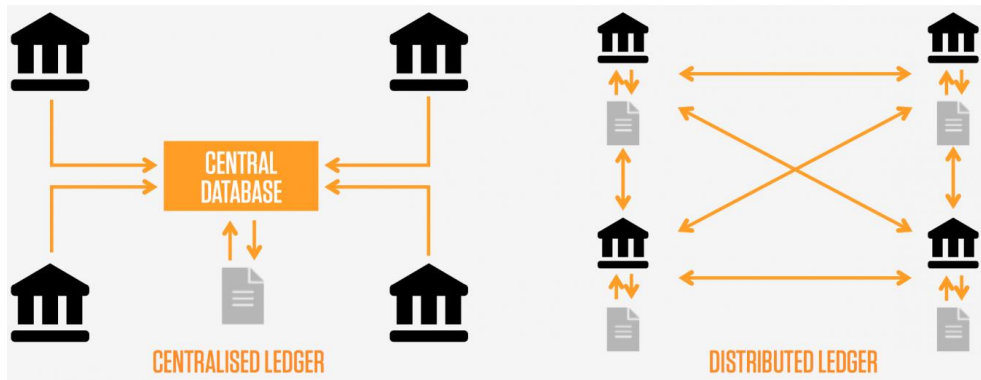
CUSTOMER

CUSTOMER_ID	LAST_NAME	FIRST_NAME	STREET	CITY	ZIP_CODE	COUNTRY
10302	Boucher	Leo	54, rue Royale	Nantes	44000	France
11244	Smith	Laurent	8489 Strong St	Las Vegas	83030	USA
11405	Han	James	636 St Kilda Road	Sydney	3004	Australia
11993	Mueller	Tomas	Berliner Weg 15	Tamm	71732	Germany
12111	Carter	Nataly	5 Tomahawk	Los Angeles	90006	USA
14121	Cortez	Nola	Av. Grande, 86	Madrid	28034	Spain
14400	Brown	Frank	165 S 7th St	Chester	33134	USA
14578	Wilson	Sarah	Seestreet #6101	Emory	1734	USA
14622	Jones	John	71 San Diego Ave	Arlington	69004	USA



## TYPES OF DATABASES

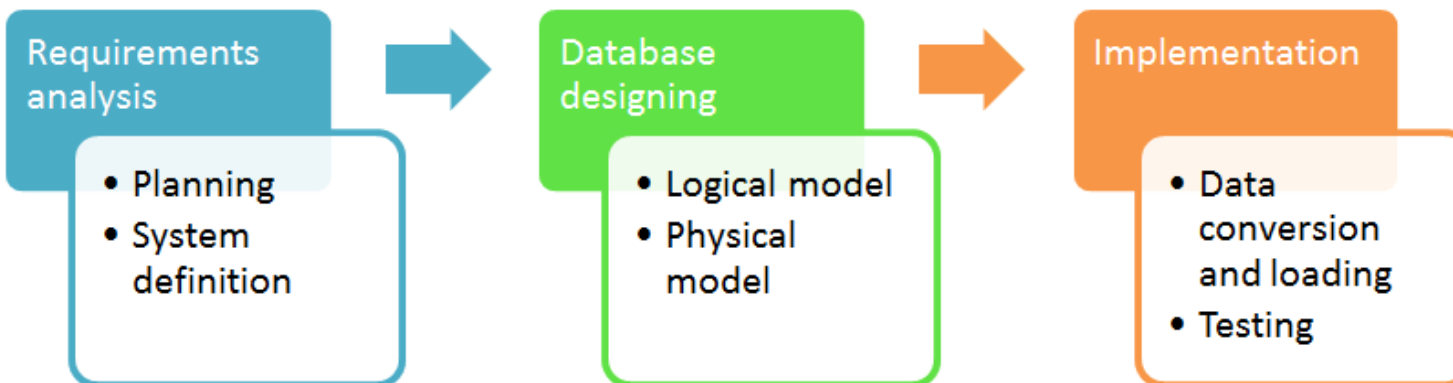
PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER		CENTRALIZED	DISTRIBUTED	OPERATIONAL	ANALYTICAL	
		WORKGROUP	ENTERPRISE					
MS Access	X	X		X		X		
MS SQL Server	X*	X	X	X	X	X	X	X
IBM DB2	X*	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X*	X	X	X	X	X	X	X



```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" h
<items>
  <item available
    <name>Mocha
    <type>Coffee</type>
    <photo>photos/candles.jpg<
```

# Tại sao thiết kế dữ liệu lại quan trọng?

- Thiết kế cơ sở dữ liệu tập trung vào thiết kế **cấu trúc** cơ sở dữ liệu dùng cho dữ liệu **người dùng cuối**.
  - Người thiết kế phải xác định được việc sử dụng CSDL dự kiến
- Thiết kế CSDL tốt:
  - Quản lý dữ liệu trở nên **dễ dàng**
  - Tạo ra thông tin **chính xác** và **có giá trị**
- Thiết kế CSDL chưa tốt:
  - Gây ra những khó khăn dễ dẫn đến những sai lầm



# Tại sao thiết kế dữ liệu lại quan trọng?

Tại sao Enum lại để trống dòng 9 và 10?

Lập danh sách nhân viên theo thứ tự ABC như thế nào?

Làm thế nào để đếm số lượng nhân viên có chứng chỉ Basic Database Manipulation?

Có phải chứng chỉ Basic Database Manipulation giống Basic DB Manipulation không?

Nếu nhân viên thêm 1 chứng chỉ nữa thì làm thế nào?

another column?

ID	Enum	Name	Title	HireDate	Skill1	Skill1Date	Skill2	Skill2Date	Skill3	Skill3Date
1	02345	Brian Dates	DBA	2/14/1995	Basic Database Management	2/14/2002	Advanced Database Management	2/14/2005	Basic Web Design	8/9/2009
2	08273	Marco Blizz	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	05373	Franklin Johnson, Jr.	Purchasing Agent	3/15/2002	Advanced Spreadsheets	5/20/2011				
5	13687	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	08382	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	13883	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	13832	Lee, Megan	Programmer	9/29/2013						



# Tại sao thiết kế dữ liệu lại quan trọng?

- FIGURE 1.6 EMPLOYEE SKILLS CERTIFICATION IN A GOOD DESIGN

Table name: EMPLOYEE

Employee_ID	Employee_FName	Employee_LName	Employee_HireDate	Employee_Title
02345	Brian	Oates	2/14/1999	DBA
03373	Franklin	Johnson	3/15/2006	Purchasing Agent
04893	Patricia	Richards	6/11/2008	DBA
06234	Jasmine	Patel	8/10/2009	Programmer
08273	Marco	Bienz	7/28/2010	Analyst
09002	Wade	Gather	5/20/2014	Clerk
09283	Juan	Chavez	7/4/2014	Clerk
09382	Susan	Mathis	8/2/2014	Database Programmer
10282	Amanda	Richardson	4/11/2015	Clerk
13383	Raymond	Matthews	3/12/2016	Programmer
13567	Robert	Almond	9/30/2016	Analyst
13932	Megan	Lee	9/29/2017	Programmer
14311	Lee	Duong	9/1/2018	Programmer

Table name: CERTIFIED

Employee_ID	Skill_ID	Certified_Date
02345	100	2/14/2004
02345	110	8/9/2005
02345	180	2/14/2007
03373	120	6/20/2013
04893	180	6/11/2008
04893	220	9/20/2014
06234	110	8/10/2009
06234	200	8/10/2009
06234	210	1/29/2014
08273	110	3/8/2011
08273	190	8/19/2014
09002	110	5/16/2015
09002	120	5/16/2015
09382	140	8/2/2014
09382	210	8/2/2014
09382	220	5/1/2015
13383	170	3/12/2016
13567	130	9/30/2016
13567	140	5/23/2017
14311	110	9/1/2018

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.

- Biết những hạn chế của tập → Hiểu cách CSDL sẽ khắc phục
- Lý do cần hiểu xử lý tập:
  - Sự phức tạp trong thiết kế CSDL trở nên dễ hiểu hơn
  - Tránh gặp phải các vấn đề với HQTCSDL
  - Chuyển đổi dữ liệu từ tập sang CSDL
- Hệ thống tập: bao gồm tập các tập, mỗi tập chứa nội dung riêng
  - Được tổ chức theo mục đích mong muốn
- Nội dung các tập có liên quan với nhau về mặt logic

# Sự phát triển của xử lý dữ liệu tệp

- Hệ thống tệp thủ công:
  - Tủ hồ sơ → Tập hồ sơ → Tờ giấy ghi chép
  - Dành cho tổ chức có ít dữ liệu
- Tệp tin máy tính:
  - Chuyển đổi dữ liệu từ dạng thủ công → cấu trúc tệp trên máy tính
    - Viết các phần mềm để xử lý dữ liệu
    - Thiết kế các chương trình ứng dụng
  - Đầu tiên, tệp tin máy tính được lưu như dạng thủ công
  - Đòi hỏi có những thuật ngữ chuyên biệt để giao tiếp
    - Dữ liệu: dữ liệu thô như: ngày sinh, số điện thoại, họ tên,...
    - Trường: 1 hoặc 1 nhóm ký tự có nghĩa, dùng để xác định và lưu trữ dữ liệu
    - Bản ghi: 1 hoặc nhiều trường dùng để mô tả 1 đối tượng dữ liệu.
    - Tệp: gồm tập các bản ghi có liên quan. VD tệp chứa danh sách SV lớp TT32

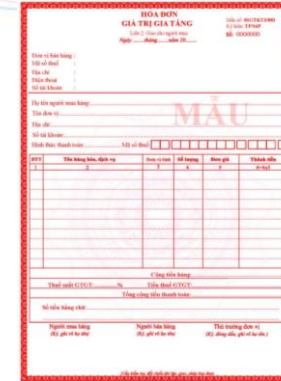
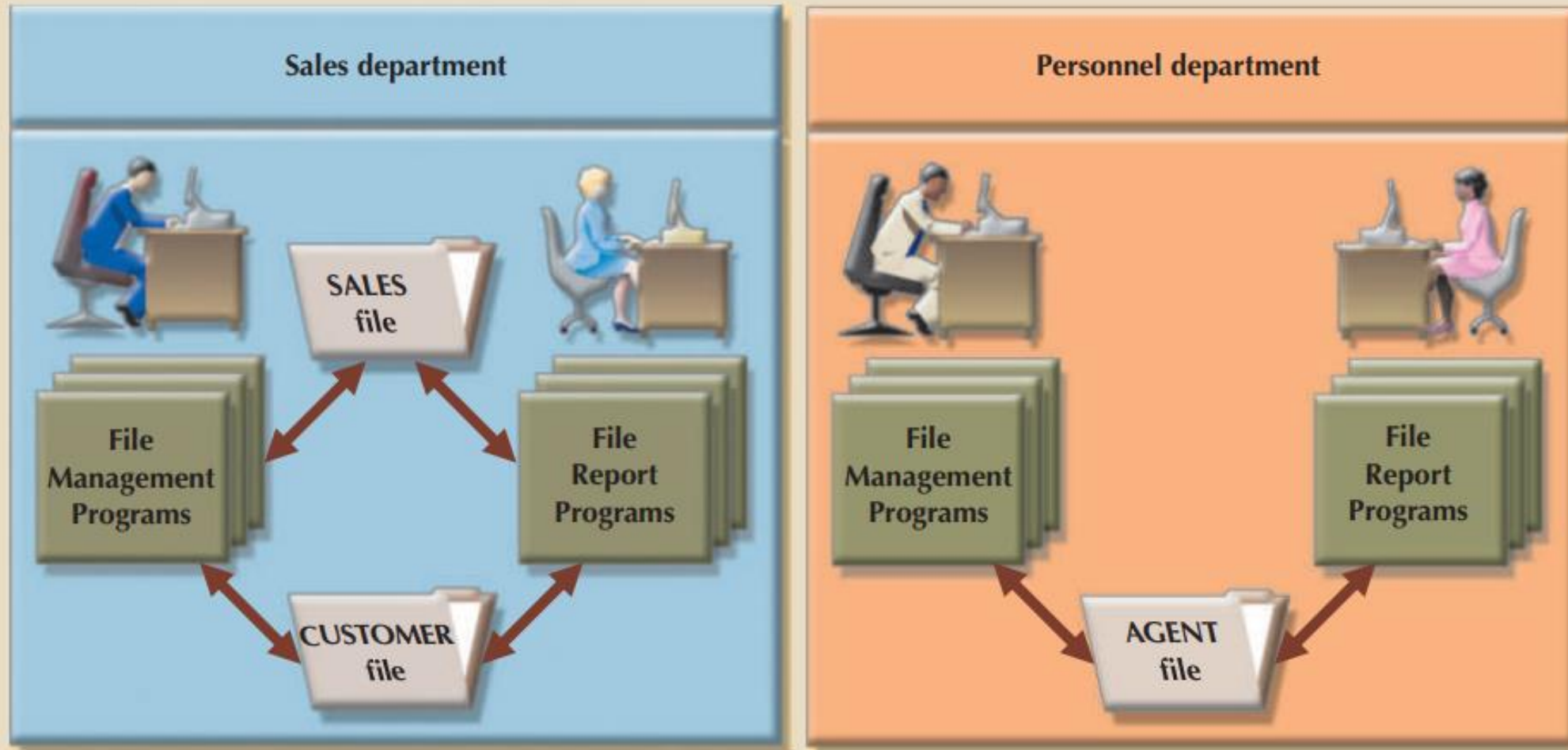


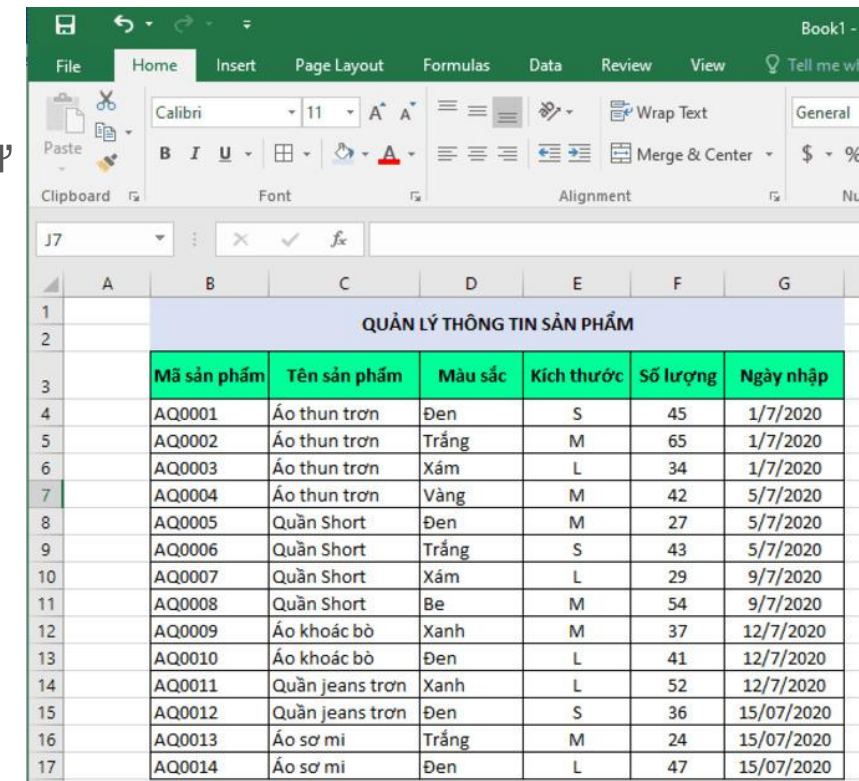


FIGURE 1.9 A SIMPLE FILE SYSTEM



# Sự phát triển của xử lý dữ liệu tệp

- Sự ra đời dạng tệp hỗ trợ tính toán
  - Dẫn đến mọi người đều sử dụng để tăng năng suất → xảy ra vấn đề tương tự như thủ công
- Microsoft Excel:
  - Được sử dụng rộng rãi trong người dùng kinh doanh
  - Sử dụng các hàm để tính toán nhanh, nhưng vẫn lưu như thủ công
  - Khi người dùng thành thạo bảng tính thì có xu hướng lạm dụng bảng tính thay CSDL



The screenshot shows the Microsoft Excel interface with a table titled "QUẢN LÝ THÔNG TIN SẢN PHẨM". The table has 7 columns: Mã sản phẩm, Tên sản phẩm, Màu sắc, Kích thước, Số lượng, and Ngày nhập. The data is organized into rows, with the first row (row 3) serving as the header. The table lists 14 products, including various types of shirts and shorts, with their respective colors, sizes, quantities, and input dates.

	A	B	C	D	E	F	G
1		QUẢN LÝ THÔNG TIN SẢN PHẨM					
2							
3		Mã sản phẩm	Tên sản phẩm	Màu sắc	Kích thước	Số lượng	Ngày nhập
4		AQ0001	Áo thun trơn	Đen	S	45	1/7/2020
5		AQ0002	Áo thun trơn	Trắng	M	65	1/7/2020
6		AQ0003	Áo thun trơn	Xám	L	34	1/7/2020
7		AQ0004	Áo thun trơn	Vàng	M	42	5/7/2020
8		AQ0005	Quần Short	Đen	M	27	5/7/2020
9		AQ0006	Quần Short	Trắng	S	43	5/7/2020
10		AQ0007	Quần Short	Xám	L	29	9/7/2020
11		AQ0008	Quần Short	Be	M	54	9/7/2020
12		AQ0009	Áo khoác bò	Xanh	M	37	12/7/2020
13		AQ0010	Áo khoác bò	Đen	L	41	12/7/2020
14		AQ0011	Quần jeans trơn	Xanh	L	52	12/7/2020
15		AQ0012	Quần jeans trơn	Đen	S	36	15/07/2020
16		AQ0013	Áo sơ mi	Trắng	M	24	15/07/2020
17		AQ0014	Áo sơ mi	Đen	L	47	15/07/2020

- Tệp tin là sự cải tiến trong hệ thống thủ công trên máy tính
- Đòi hỏi phải có thêm chương trình khi có yêu cầu mới
  - Truy vấn đặc biệt là không thể
  - Khó khăn trong thay đổi cấu trúc
- Khó khăn trong vấn đề bảo mật
  - Thường bị bỏ qua khi sử dụng tệp để lưu trữ



- **Độc lập về Dữ liệu và Cấu trúc:**
  - **Phụ thuộc cấu trúc:** truy cập vào một tập phụ thuộc vào cấu trúc riêng của nó
    - Tất cả chương trình phải sửa đổi cho phù hợp với tập mới
  - **Độc lập cấu trúc:** thay đổi cấu trúc tập không ảnh hưởng đến truy nhập dữ liệu
  - **Phụ thuộc dữ liệu:** các thay đổi truy cập dữ liệu khi các đặc tính lưu trữ dữ liệu thay đổi
  - **Độc lập dữ liệu:** các đặc tính lưu trữ dữ liệu không ảnh hưởng đến truy nhập dữ liệu

- **Độc lập về Dữ liệu và Cấu trúc:**
  - Ý nghĩa thực tế của sự phụ thuộc dữ liệu là khác nhau giữa định dạng logic và vật lý
  - Định dạng dữ liệu logic: cách con người xem dữ liệu
  - Định dạng dữ liệu vật lý: cách máy tính làm việc với dữ liệu
  - Mỗi chương trình gồm:
    - Dòng chỉ định mở loại tệp tin cụ thể
    - Đặc tả bản ghi
    - Định nghĩa trường

- **Sự dư thừa dữ liệu:**

- Cấu trúc của tệp khó kết hợp dữ liệu từ nhiều nguồn
  - Vấn đề an ninh
- Lưu trữ dữ liệu ở nhiều nơi khác nhau
  - Lưu trữ riêng rẽ, không có sự chia sẻ
- Thiếu tính nhất quán vì lưu trữ dữ liệu ở nhiều nơi khác nhau
- Dư thừa dữ liệu: dữ liệu được lưu trữ không cần thiết ở nhiều nơi khác nhau
- Không nhất quán dữ liệu: do các phiên bản khác nhau và xung đột của cùng một dữ liệu xảy ra ở các địa điểm khác nhau
- Bất thường dữ liệu: khi tất cả thay đổi trong dữ liệu dư thừa không được thực hiện đúng
  - Bất thường khi: Chèn, Sửa, Xóa



- Phụ thuộc cấu trúc:
  - Truy cập 1 tệp phụ thuộc vào cấu trúc của tệp
  - Chương trình phải thay đổi khi cấu trúc tệp thay đổi
- Độc lập cấu trúc:
  - Sự thay đổi cấu trúc của tệp không làm ảnh hưởng đến truy cập
- Kiểu dữ liệu:
  - Định nghĩa loại giá trị có thể sử dụng và lưu trữ
  - Dựa vào kiểu dữ liệu để xác định thao tác có thể thực hiện
- Phụ thuộc dữ liệu:
  - Biểu diễn và thao tác dữ liệu phụ thuộc lưu trữ dữ liệu vật lý
- Độc lập dữ liệu:
  - Việc truy cập dữ liệu không bị ảnh hưởng bởi thay đổi trong lưu trữ.

- Ý nghĩa của sự phụ thuộc dữ liệu là sự khác nhau giữa định dạng logic và định dạng vật lý.
  - Định dạng logic: là cách con người nhìn dữ liệu
  - Định dạng vật lý: là cách máy tính phải làm việc với dữ liệu
- Mỗi chương trình sẽ gồm:
  - Dòng chỉ định mở loại tệp cụ thể
  - Đặc tả, định nghĩa các trường

- Cấu trúc hệ thống tệp tin khó kết hợp dữ liệu từ nhiều nguồn.
  - Có thể vi phạm vấn đề an ninh
- Cách thức tổ chức dẫn đến lưu trữ cùng 1 dữ liệu ở nhiều địa điểm khác nhau.
  - Có thể trở thành 1 “ốc đảo” thông tin
  - Tạo ra sự không nhất quán dữ liệu
- Sự dư thừa dữ liệu:
  - Dữ liệu được lưu trữ giống nhau ở nhiều nơi.
- Sự không nhất quán:
  - Sự khác nhau và xung đột giữa các phiên bản, nơi khác nhau
- Sự bất thường:
  - Update, Insert, Delete

Students

IDSt	LastName	IDProf	Prof	Grade
1	Mueller	3	Schmid	5
2	Meier	2	Borner	4
3	Tobler	1	Bernasconi	6

Startsituation

Result after normalisation

Students

ID	LastName
1	Mueller
2	Meier
3	Tobler

Professors

IDProf	Professor
1	Bernasconi
2	Borner
3	Schmid

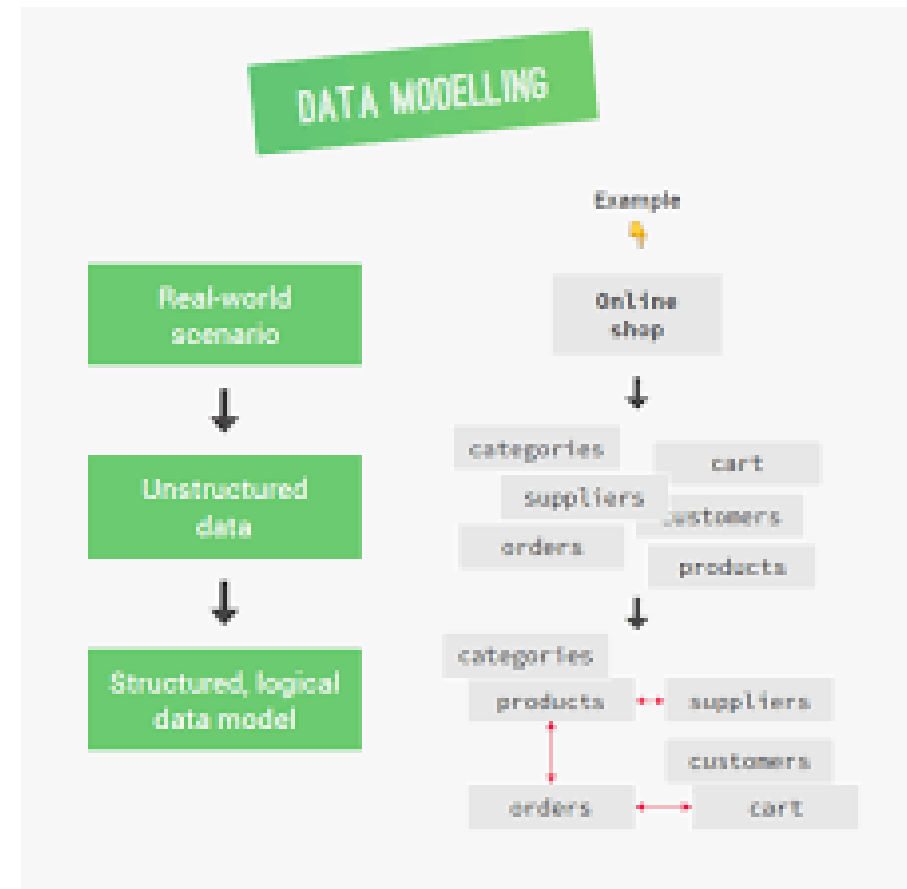
Grades

IDStIDProf	Grade	
1	3	5
2	2	4
3	1	6

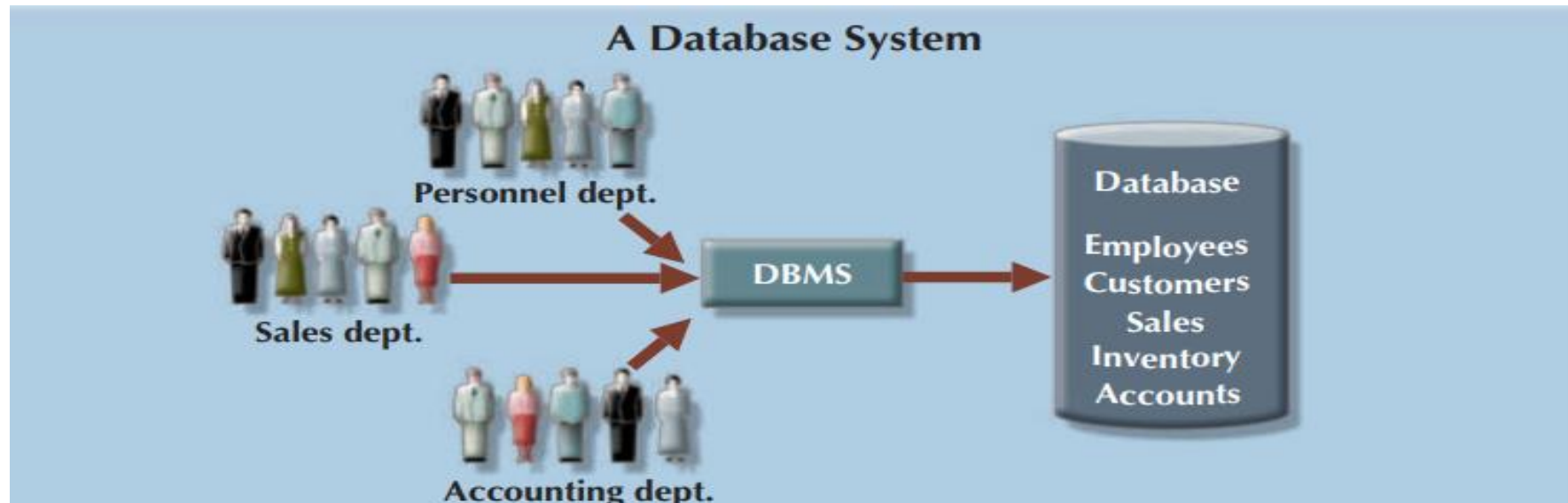


# Thiếu kỹ năng mô hình hóa và thiết kế dữ liệu multicampus

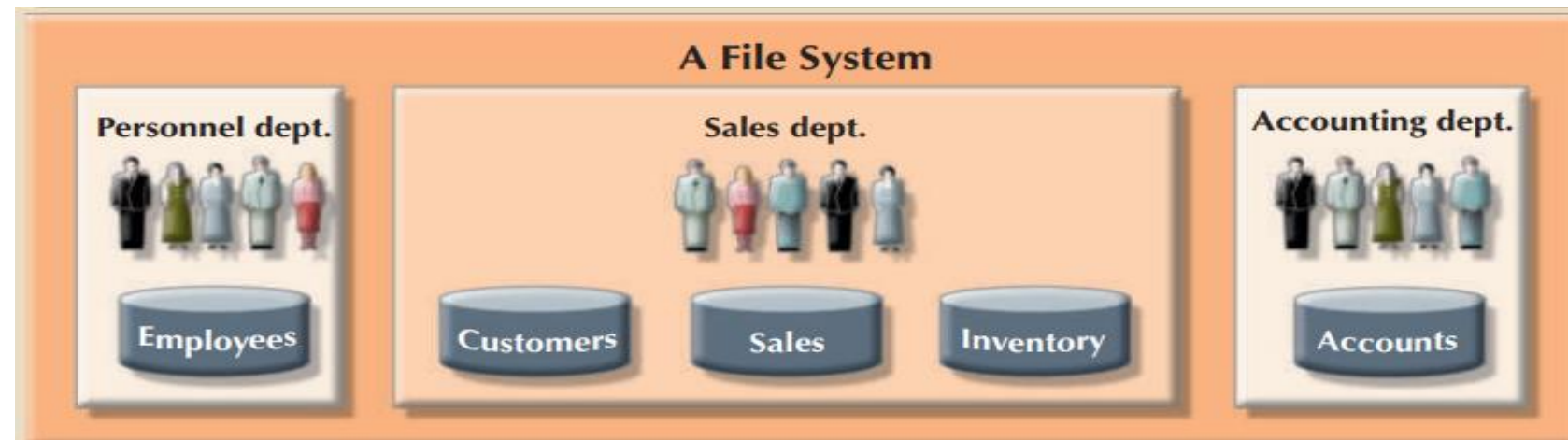
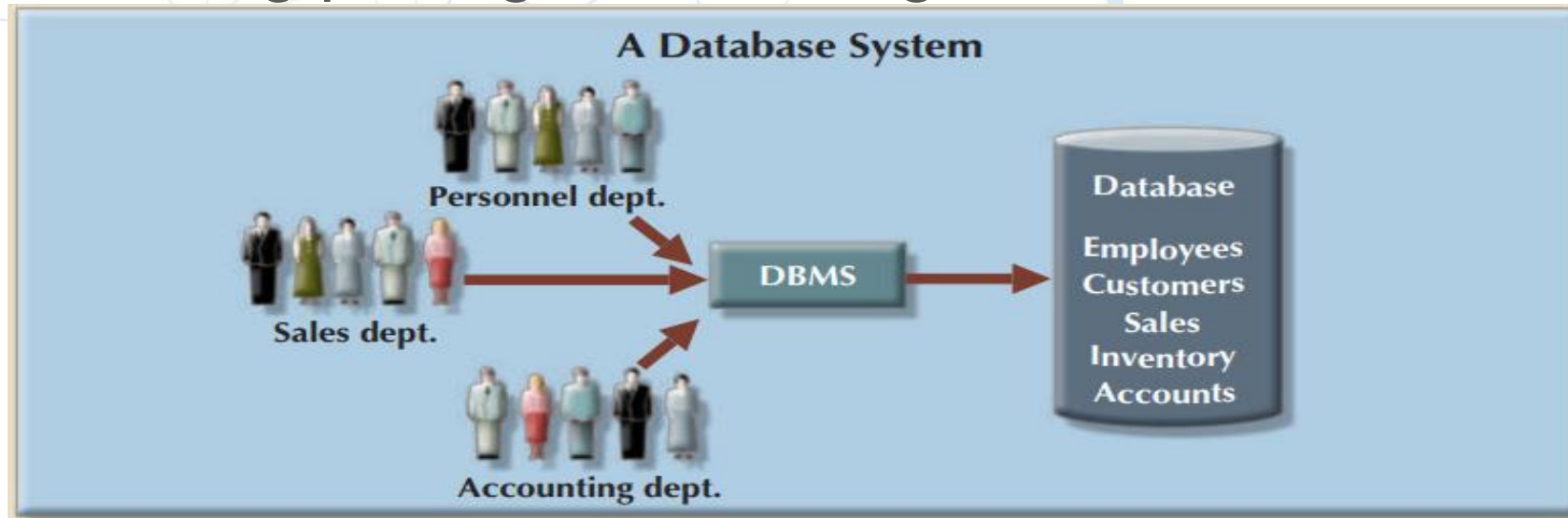
- Hầu hết người dùng đều thiếu kỹ năng về thiết kế CSDL
  - Mặc dù sử dụng nhiều công cụ cá nhân (Excel)
- Kỹ năng về mô hình hóa dữ liệu:
  - Trong quá trình thiết kế
- Mô hình dữ liệu tốt:
  - Sẽ tạo điều kiện kết nối giữa nhà thiết kế, người sử dụng và nhà phát triển.



- Hệ thống cơ sở dữ liệu bao gồm các dữ liệu liên quan được lưu trữ trong một nơi chứa dữ liệu duy nhất về mặt logic
  - Có thể phân tán về mặt vật lý tại nhiều nơi
  - HQTCSDL loại bỏ hầu hết các vấn đề của tệp tin
  - Lưu trữ cấu trúc dữ liệu, quan hệ giữa các cấu trúc và cách truy cập
    - Định nghĩa, lưu trữ, quản lý cách truy cập và các thành phần

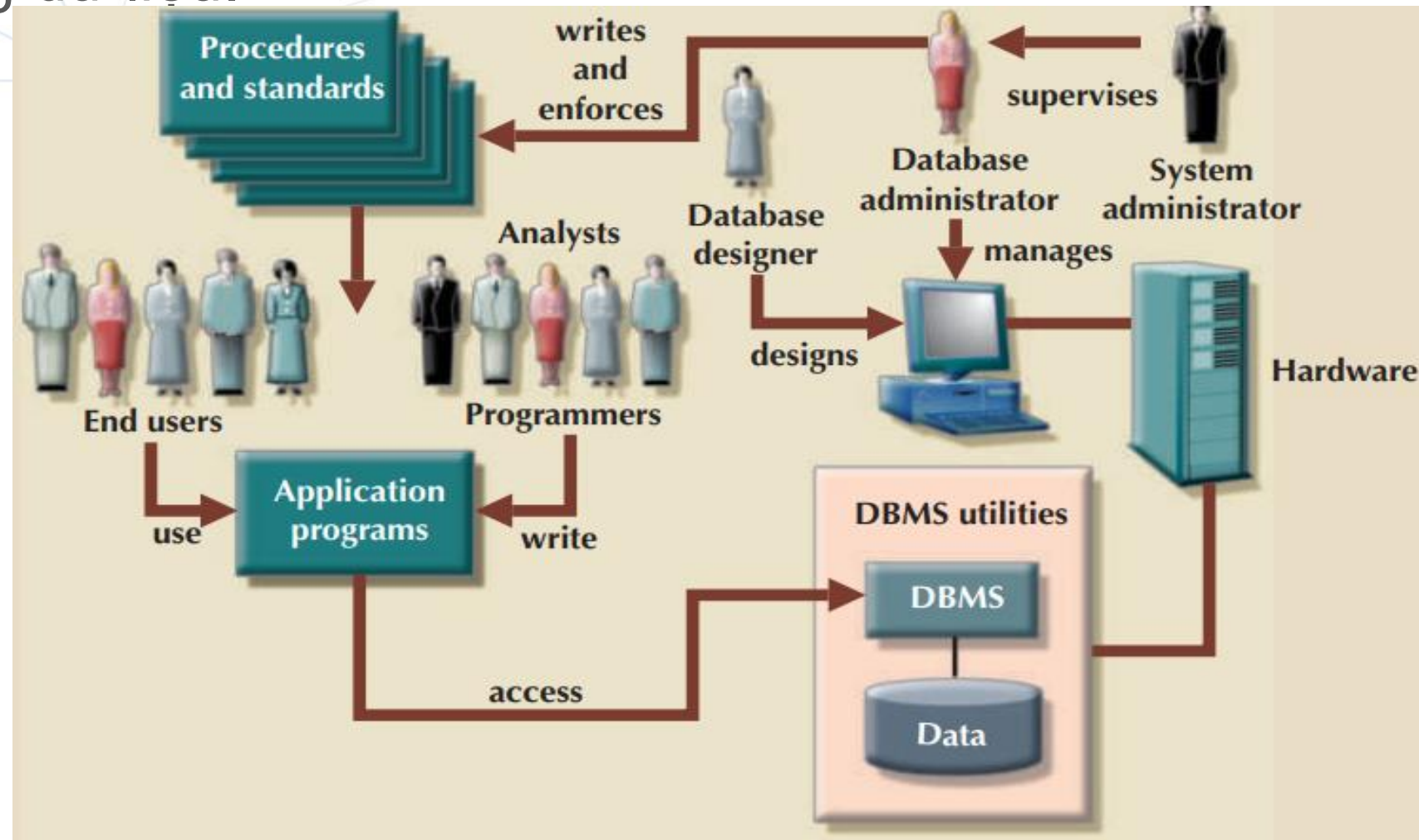


- Sự tương phản giữa hệ thống cơ sở dữ liệu và hệ thống tệp



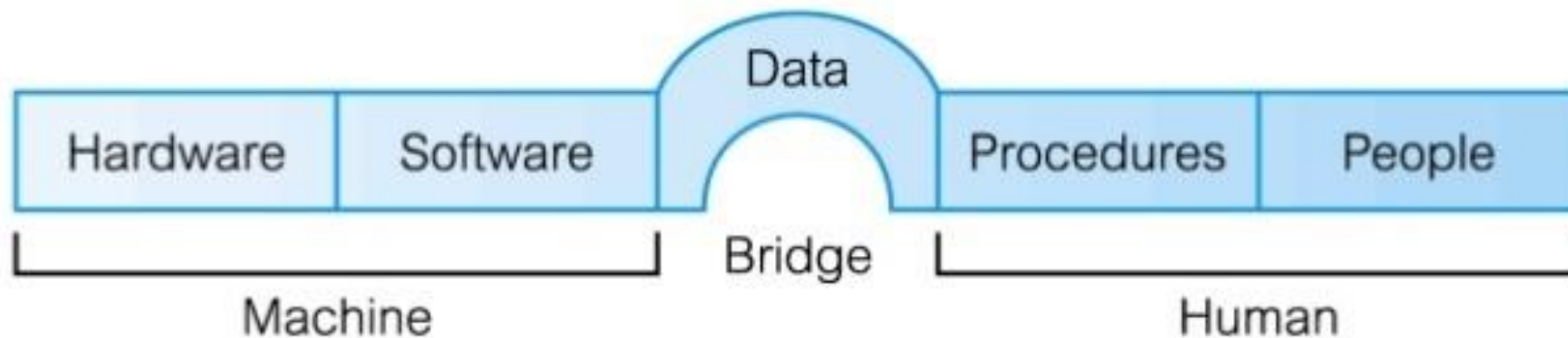


- Hệ thống CSDL: định nghĩa và điều chỉnh việc thu thập lưu trữ, quản lý, sử dụng dữ liệu.
- Gồm :
  - Phần cứng,
  - Phần mềm,
  - Con người,
  - Quy tắc,
  - Dữ liệu



- **Phần cứng:**
  - Tất cả các thiết bị vật lý của hệ thống
- **Phần mềm:**
  - Gồm 3 loại phần mềm: Hệ điều hành, Phần mềm DBMS, Chương trình ứng dụng.
- **Con người:** tất cả người tham gia vào Hệ thống CSDL
  - Quản trị dữ liệu và Hệ thống
  - Người thiết kế CSDL
  - Lập trình và phân tích hệ thống
  - Người dùng cuối
- **Thủ tục:**
  - Hướng dẫn và quy tắc điều chỉnh việc thiết kế và sử dụng hệ thống cơ sở dữ liệu
- **Dữ liệu:**
  - Tập hợp những sự kiện được lưu trữ trong CSDL

- Hệ thống cơ sở dữ liệu được tạo ra và quản lý ở các cấp độ phức tạp khác nhau và tiêu chuẩn khác nhau.
  - Hệ thống CSDL: ở cửa hàng bán lẻ và ở hệ thống bảo hiểm quốc gia
- Giải pháp CSDL phải đem lại hiệu quả về mặt chi phí cũng như chiến lược của tổ chức.
- Công nghệ cơ sở dữ liệu được sử dụng sẽ ảnh hưởng đến lựa chọn hệ thống cơ sở dữ liệu



- Quản lý lưu trữ dữ liệu:
  - Tạo và quản lý cấu trúc phức tạp cho việc lưu trữ
  - Điều hành hiệu quả việc thực thi của cơ sở dữ liệu
  - HQTCSDL lưu trữ CSDL trong nhiều tệp
- Chuyển đổi và trình bày dữ liệu
  - Chuyển đổi dữ liệu nhập vào phù hợp với cấu trúc dữ liệu
  - Biến đổi dữ liệu dạng vật lý sang dạng người dùng có thể hiểu
- Quản lý bảo mật:
  - Tạo hệ thống bảo mật để thực thi quyền riêng tư dữ liệu và người dùng
  - Quy tắc bảo mật xác định người dùng truy cập và mục nào có thể truy cập.



- Kiểm soát nhiều người dùng truy cập:
  - Đảm bảo việc truy cập đồng thời không ảnh hưởng đến tính toàn vẹn
- Quản lý dự phòng và khôi phục:
  - Đảm bảo tính an toàn và toàn vẹn dữ liệu
  - Có thể khôi phục dữ liệu khi gặp sự cố
- Quản lý tính toàn vẹn dữ liệu:
  - Thúc đẩy và thực thi tính toàn vẹn:
    - Giảm thiểu dữ liệu thừa
    - Tối đa sự nhất quán
  - Lưu trữ mối quan hệ dữ liệu để thực thi tính toàn vẹn
  - Tính toàn vẹn rất quan trọng trong CSDL hướng giao dịch

- Quản lý từ điển dữ liệu:
  - DBMS lưu trữ định nghĩa của các phần tử và mối quan hệ trong từ điển dữ liệu.
  - Sử dụng từ điển dữ liệu để tìm kiếm các cấu trúc và mối quan hệ → giảm sự phức tạp trong chương trình.
  - Mọi thay đổi liên quan đến CSDL đều được ghi lại trong từ điển → không cần sửa đổi trong chương trình.
  - DBMS cung cấp sự trừu tượng dữ liệu và loại bỏ sự phụ thuộc cấu trúc và dữ liệu từ chương trình.
  - Ví dụ: hình dưới minh họa cách MS SQL Server định nghĩa dữ liệu cho bảng CUSTOMER.

# Chức năng của HQTCS DL

Microsoft SQL Server Management Studio Express

File Edit View Table Designer Tools Window Help

New Query

Object Explorer

LA004\SQLEXPRESS (SQL Server 9.0.3042)

- Databases
  - System Databases
  - RobCor
    - Database Diagrams
    - Tables
      - System Tables
      - dbo.AGENT
      - dbo.AGENT\_V2
      - dbo.CUSTOMER
      - Columns
      - Keys
      - Constraints
      - Triggers
      - Indexes
      - Statistics
      - dbo.CUSTOMER\_V2
    - Views
    - Synonyms
    - Programmability
    - Security
  - Security
  - Server Objects
  - Replication
  - Management

Table - dbo.CUSTOMER Summary

Column Name	Data Type	Allow Nulls
C_NAME	varchar(20)	<input checked="" type="checkbox"/>
C_PHONE	varchar(12)	<input checked="" type="checkbox"/>
C_ADDRESS	varchar(30)	<input checked="" type="checkbox"/>
C_ZIP	varchar(5)	<input checked="" type="checkbox"/>
A_NAME	varchar(20)	<input checked="" type="checkbox"/>
A_PHONE	varchar(12)	<input checked="" type="checkbox"/>
TP	varchar(2)	<input checked="" type="checkbox"/>
AMT	numeric(6, 2)	<input checked="" type="checkbox"/>
REN	datetime	<input checked="" type="checkbox"/>

Column Properties

(General)

(Name)	C_NAME
Allow Nulls	Yes
Data Type	varchar
Default Value or Binding	
Length	20

Table Designer

Collation	<database default>
Computed Column Specification	
Condensed Data Type	varchar(20)
Description	
Deterministic	Yes
DTS-published	No
Full-text Specification	No
Has Non-SQL Server Subscriber	No
Identity Specification	No

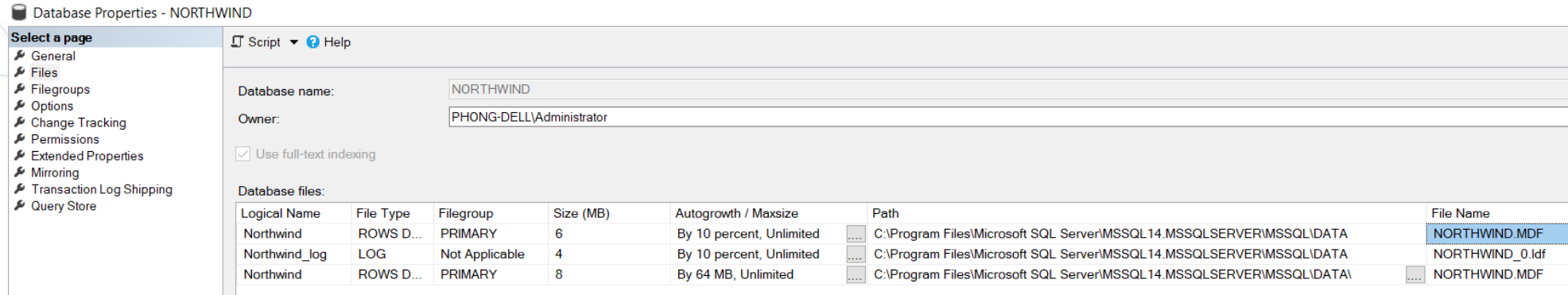
Metadata

- Quản lý lưu trữ dữ liệu:
  - DBMS tạo và quản lý cấu trúc phức tạp cần thiết cho việc lưu trữ → giảm nhiệm vụ về định nghĩa và lập trình cho các đặc tính dữ liệu mức vật lý.
  - DBMS hiện đại còn lưu trữ định nghĩa mẫu nhập liệu, báo cáo, quy tắc xác thực dữ liệu,...
  - Điều chỉnh hiệu suất CSDL liên quan đến các hoạt động làm cho CSDL thực hiện hiệu quả hơn trong lưu trữ và truy cập.
  - DBMS lưu trữ CSDL trong nhiều tệp vật lý

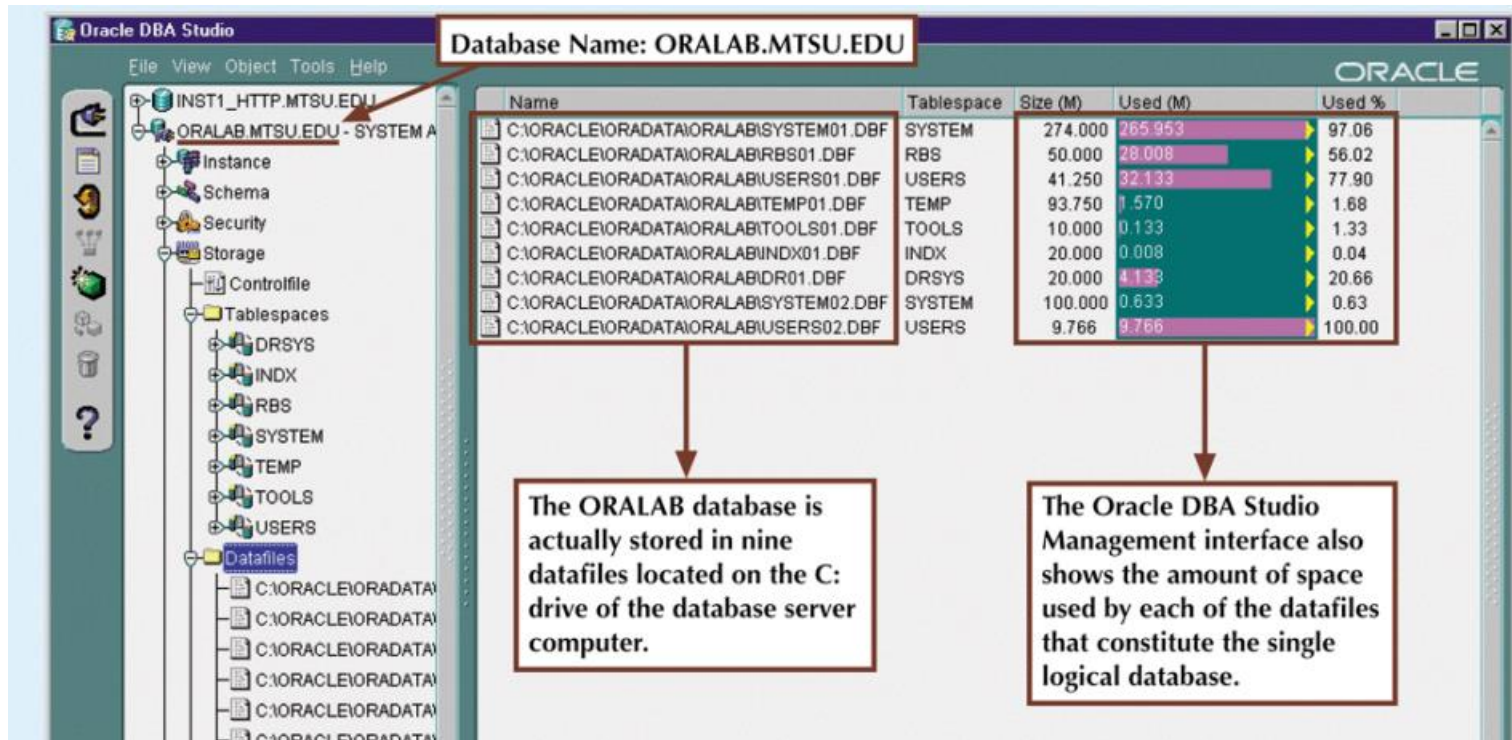


# Chức năng của HQTCSDL

- Ví dụ: CSDL trên MS SQL Server



- Ví dụ:  
CSDL  
trên  
Oracle



- Trình bày và chuyển đổi dữ liệu:
  - DBMS chuyển đổi dữ liệu cho phù hợp với cấu trúc vật lý
  - DBMS chuyển dữ liệu vật lý phù hợp với mong muốn người dùng
  - VD: “July 11, 2017” → “11/07/2017” → “07/11/2017”
- Quản lý bảo mật:
  - DBMS tạo hệ thống bảo mật với người dùng và dữ liệu
  - Quy tắc bảo mật xác định quy tắc mà người dùng có thể truy cập CSDL, các phần tử có thể được truy cập,...
- Kiểm soát nhiều người truy cập:
  - DBMS sử dụng thuật toán đảm bảo khi truy cập đồng thời không ảnh hưởng đến tính toàn vẹn dữ liệu.

- Quản lý dự phòng và khôi phục dữ liệu:
  - DBMS cung cấp cách dự phòng và khôi phục dữ liệu đảm bảo dữ liệu được an toàn và toàn vẹn.
- Quản lý tính toàn vẹn dữ liệu:
  - DBMS yêu cầu và thực thi các quy tắc toàn vẹn để giảm thiểu dư thừa, tối đa hóa dữ liệu.
  - Các mối quan hệ dữ liệu được lưu trong từ điển dữ liệu dùng để thực thi tính toàn vẹn dữ liệu.
  - Tính toàn vẹn dữ liệu là đặc biệt quan trọng trong Hệ thống CSDL giao dịch.



- Các ngôn ngữ truy cập cơ sở dữ liệu và giao diện lập trình ứng dụng
  - DBMS cung cấp cách truy cập bằng ngôn ngữ truy vấn
  - Ngôn ngữ truy vấn là ngôn ngữ phi thủ tục: chỉ ra cái gì cần làm mà không cần chỉ ra cách làm như thế nào.
  - SQL (Structured Query Language – Ngôn ngữ truy vấn cấu trúc):
    - Là ngôn ngữ chuẩn được đa số DBMS hỗ trợ: Oracle, SQL Server, DB2,...
- Giao diện giao tiếp CSDL:
  - DBMS hiện nay cung cấp cho người dùng truy cập bằng nhiều môi trường mạng khác nhau.
    - Sinh ra câu truy vấn trên màn hình của trình duyệt Web.
    - Tự động đưa ra các báo cáo định trước trên Web
    - Có thể kết nối với bên thứ 3 để phân phối thông tin qua email hoặc ứng dụng khác



- HT CSDL cung cấp một nền tảng mà các quy định và tiêu chuẩn nghiêm ngặt được áp dụng.
  - Vai trò của con người chuyển từ lập trình sang quản lý nguồn nhân lực của tổ chức.
  - HT CSDL cho phép sử dụng nhiều dữ liệu hơn.
  - Cấu trúc dữ liệu được tạo ra trong CSDL và mối quan hệ của chúng xác định tính hiệu quả.
- Nhược điểm của HT CSDL:
  - Tăng chi phí: đòi hỏi phần cứng, phần mềm,...
  - Sự phức tạp trong quản lý: sự phức tạp trong công nghệ, vận hành đúng cách,...
  - Sự duy trì: Sự cập nhật thường xuyên, áp dụng biện pháp bảo mật
  - Phụ thuộc nhà cung cấp: đầu tư nhiều → miễn cưỡng thay đổi
  - Chu kỳ thay thế và nâng cấp

# Các công việc liên quan đến CSDL

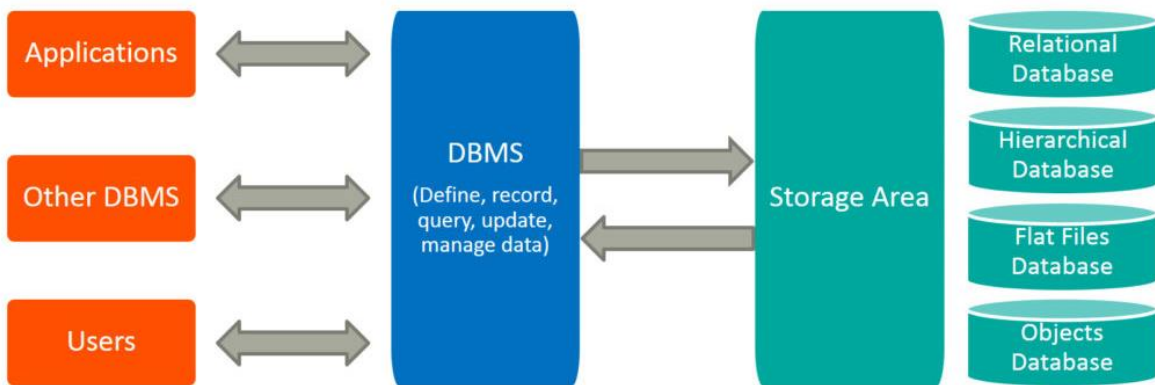
Công việc	Mô tả	Kỹ năng
Phát triển	Tạo và duy trì ứng dụng dựa trên CSDL	Lập trình, cơ bản về CSDL, SQL
Thiết kế	Thiết kế và duy trì CSDL	Thiết kế hệ thống, CSDL, SQL
Quản trị	Quản lý và duy trì HQTCSDL và CSDL	Cơ bản về CSDL, SQL, quản lý dữ liệu
Phân tích	Xây dựng các báo cáo hỗ trợ quyết định	SQL, tối ưu truy vấn, Kho dữ liệu
Kiến trúc sư	Thiết kế và cài đặt CSDL (khái niệm, logic, vật lý)	Cơ bản về HQTCSDL, mô hình dữ liệu, SQL, kiến thức về phần cứng
Tư vấn	Hỗ trợ các tổ chức sử dụng CSDL để cải thiện các quy trình kinh doanh	Cơ bản về CSDL, mô hình dữ liệu, thiết kế CSDL, SQL, HQTCSDL, phần cứng, công nghệ,...
Bảo mật	Cài đặt các chính sách bảo mật cho quản trị dữ liệu	Cơ bản về HQTCSDL, cài đặt CSDL, SQL, công nghệ về bảo mật dữ liệu,..
Kiến trúc sư dữ liệu điện toán đám mây	Thiết kế và cài đặt hạ tầng cho hệ thống CSDL đám mây	Công nghệ Internet, lưu trữ đám mây, bảo mật, CSDL lớn,....
Khoa học dữ liệu	Phân tích lượng lớn dữ liệu để tạo ra thông tin	Phân tích dữ liệu, thống kê, toán học, SQL, lập trình, khai phá dữ liệu, học máy, trực quan hóa dữ liệu,...

- Dữ liệu là dữ liệu thô
- Thông tin là kết quả của quá trình xử lý dữ liệu để biết ý nghĩa của nó
- Thông tin chính xác, phù hợp và kịp thời là chìa khóa để đưa ra quyết định tốt
- Dữ liệu thường được lưu trữ trong một cơ sở dữ liệu
- DBMS triển khai và quản lý nội dung của CSDL
- Siêu dữ liệu (Metadata): là dữ liệu của dữ liệu
- Thiết kế cơ sở dữ liệu xác định cấu trúc cơ sở dữ liệu
  - CSDL được thiết kế tốt sẽ tạo ra quản lý dữ liệu tốt và sinh ra thông tin có giá trị
  - Thiết kế CSDL kém sẽ dẫn đến các quyết định tồi

- CSDL được phát triển từ tệp tin thủ công (giấy) và máy tính
- Trong hệ thống tệp tin thì dữ liệu được lưu trữ trong tệp độc lập
  - Mỗi một tệp có 1 chương trình quản lý riêng
- Một số hạn chế của hệ thống tệp tin:
  - Yêu cầu mở rộng chương trình
  - Quản trị hệ thống phức tạp và khó khăn
  - Thay đổi cấu trúc đã tồn tại khó
  - Thiếu tính năng bảo mật và an toàn
  - Tệp tin độc lập có thể chứa dữ liệu dư thừa



- Hệ quản trị CSDL (DBMS) được phát triển để khắc phục những hạn chế của hệ thống tệp tin
- DBMS thể hiện CSDL với người dùng như 1 kho duy nhất
  - Thúc đẩy chia sẻ dữ liệu
  - Loại bỏ sự độc lập thông tin
- DBMS thực hiện tính toàn vẹn, loại bỏ sự dư thừa và đảm bảo an toàn dữ liệu.



ORACLE<sup>®</sup>  
DATABASE

MySQL<sup>®</sup>

mongoDB

Microsoft<sup>®</sup>  
SQL Server<sup>®</sup>

MariaDB

PostgreSQL

Microsoft<sup>®</sup>  
Access

IBM DB2

ad hoc query  
 analytical database  
 business intelligence  
 centralized database  
 cloud database  
 data  
 data anomaly  
 data dependence  
 data dictionary  
 data inconsistency  
 data independence  
 data integrity  
 data management  
 data processing (DP)  
 specialist  
 data quality  
 data redundancy  
 data type  
 data warehouse  
 database  
 database design  
 database management  
 system (DBMS)

database system  
 desktop database  
 discipline-specific database  
 distributed database  
 enterprise database  
 Extensible Markup  
 Language (XML)  
 field  
 file  
 general-purpose database  
 information  
 islands of information  
 knowledge  
 logical data format  
 metadata  
 multiuser database  
 NoSQL  
 online analytical processing  
 (OLAP)  
 online transaction  
 processing (OLTP)  
 database  
 operational database

performance tuning  
 physical data format  
 production database  
 query  
 query language  
 query result set  
 record  
 semistructured data  
 single-user database  
 social media  
 structural dependence  
 structural independence  
 structured data  
 Structured Query Language  
 (SQL)  
 transactional database  
 unstructured data  
 workgroup database  
 XML database

- Học viên trả lời và làm các bài tập trong tệp:
  - Chapter1\_Database\_System.docx
- Đổi tên tệp thành STT\_HoVaTen\_Chapter1\_Database\_System.docx
- Nộp bài lên hệ thống



# THANK YOU

**multicampus**

Copyright by Multicampus Co., Ltd. All right reserved