

Korea IT School 2022

# CƠ SỞ DỮ LIỆU

Unit 2: Mô hình dữ liệu

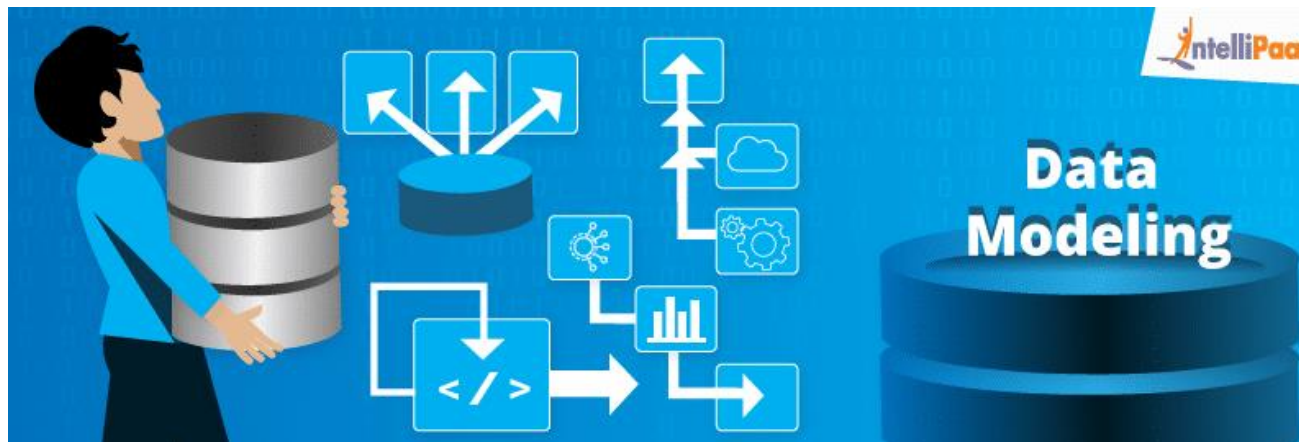
**multicampus**

Copyright by Multicampus Co., Ltd. All right reserved

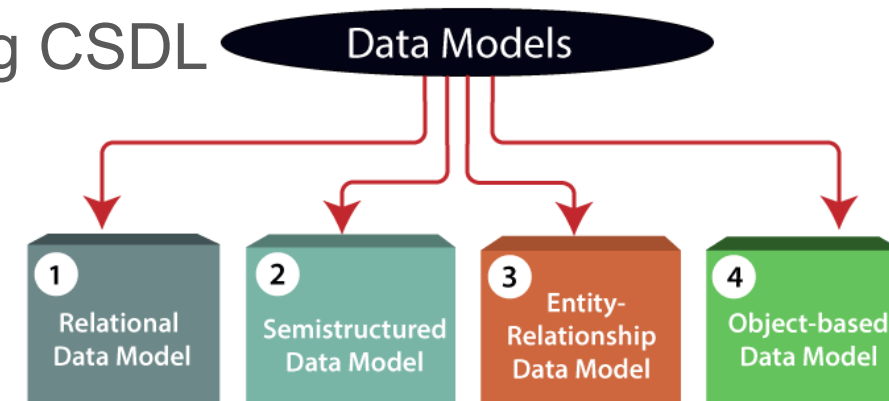
# Nội dung

1. Mô hình hóa dữ liệu và tầm quan trọng của mô hình dữ liệu
2. Các vấn đề trong xây dựng dữ liệu cơ bản
3. Các quy tắc nghiệp vụ nào có ảnh hưởng đến thiết kế CSDL
4. Sự phát triển của mô hình dữ liệu
5. Phân loại mô hình dữ liệu
6. Các mô hình dữ liệu thay thế mới nổi

- Các nhà thiết kế, lập trình viên và người dùng cuối nhìn dữ liệu theo nhiều cách khác nhau
- Sự nhìn nhận khác nhau về cùng loại dữ liệu dẫn đến thiết kế không phản đúng hoạt động của tổ chức.
- Mô hình dữ liệu làm giảm sự phức tạp của thiết kế dữ liệu
- Một vài mức độ trừu tượng dữ liệu giúp hòa giải các quan điểm khác nhau về cùng dữ liệu.



- Mô hình:
  - Sự trừu tượng hóa đối tượng hay sự vật của thế giới thực
  - Hữu ích trong việc hiểu sự phức tạp của thế giới thực
- Mô hình dữ liệu:
  - Sự biểu diễn đơn giản cho cấu trúc dữ liệu phức tạp của thế giới thực
  - Thường bằng đồ họa
- Mô hình hóa dữ liệu:
  - Quá trình tạo một mô hình dữ liệu để lưu trữ trong CSDL

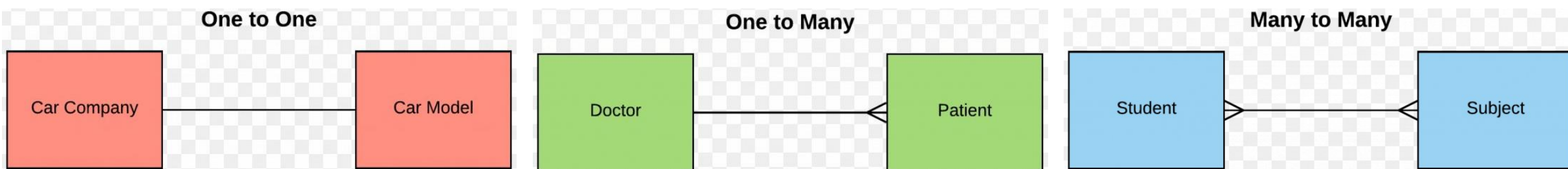




- Tạo điều kiện trao đổi giữa nhà thiết kế, lập trình viên, và người dùng cuối
- Người dùng cuối có cái nhìn và nhu cầu khác nhau về dữ liệu
- Mô hình dữ liệu sẽ tổ chức dữ liệu cho nhiều người dùng khác nhau.
- Mô hình dữ liệu là sự trừu tượng:
  - Không thể tạo ra CSDL tốt nếu không có mô hình thích hợp
  - Không thể xây dựng 1 ngôi nhà tốt mà không cần thiết kế



- Thành phần cơ bản của các mô hình dữ liệu là các thực thể, thuộc tính, mối quan hệ và ràng buộc.
  - Thực thể - Entity: bất kỳ điều gì về dữ liệu được thu thập và lưu trữ
  - Thuộc tính - Attribute: một đặc điểm, tính chất của một thực thể
  - Quan hệ - Relationship: mô tả một liên kết giữa các đối tượng
    - Quan hệ 1 – Nhiều (One-to-Many)
    - Quan hệ Nhiều – Nhiều (Many-to-Many)
    - Quan hệ 1 – 1 (One-to-One)
  - Ràng buộc – Constraint: sự giới hạn về dữ liệu



- Mô tả về chính sách, thủ tục hoặc nguyên tắc trong một tổ chức
  - Phi công không thể làm 10/24 tiếng;
  - Giáo sư không dạy quá 4 buổi/1 tuần
- Được sử dụng để xác định các đối tượng, thuộc tính, mối quan hệ và ràng buộc.
  - Một **khách hàng** có thể có nhiều **hóa đơn**
  - Một **hóa đơn** chỉ thuộc về 1 **khách hàng**
  - Một **khóa đào tạo** không thể được tổ chức nếu có ít hơn 10 hoặc nhiều hơn 30 **nhân viên**
- Mô tả đặc điểm dữ liệu mà tổ chức xem xét.
  - Sinh viên gồm: họ tên, ngày sinh và mã sinh viên xác định duy nhất 1 sinh viên



- Nguồn quy tắc nghiệp vụ:
  - Người quản lý công ty
  - Người hoạch định chính sách
  - Trưởng các phòng ban
  - Tài liệu văn bản: Quy định, Các tiêu chuẩn, Hướng dẫn hoạt động
  - Phỏng vấn trực tiếp người dùng cuối
- Quá trình xác định quy tắc nghiệp vụ giúp cho thiết kế dữ liệu:
  - Giúp chuẩn hóa quan điểm của công ty về dữ liệu.
  - Có thể là công cụ giao tiếp giữa người dùng và người thiết kế.
  - Cho phép nhà thiết kế hiểu bản chất, vai trò và phạm vi của dữ liệu.
  - Giúp nhà thiết kế hiểu quy trình nghiệp vụ.
  - Nhà thiết kế có thể đưa ra các ràng buộc, quy tắc phù hợp trong mối quan hệ và tạo ra mô hình dữ liệu chính xác.





- → Thành phần của mô hình dữ liệu
  - Danh từ → Thực thể
  - Động từ → Mối quan hệ của các thực thể
  - Mối quan hệ là hai chiều
  - Hai câu hỏi xác định kiểu mối quan hệ:
    - Có bao nhiêu trường hợp B có quan hệ với 1 trường hợp A?
    - Có bao nhiêu trường hợp A có quan hệ với 1 trường hợp B?
    - Ví dụ:
      - Có bao nhiêu sinh viên có thể tham gia vào 1 lớp học?
        - Trả lời: Nhiều sinh viên
      - Có bao nhiêu lớp học mà 1 sinh viên có thể tham gia?
        - Trả lời: Nhiều lớp học
  - Do đó, có mối quan hệ **nhiều – nhiều** giữa **Lớp học** và **Sinh viên**

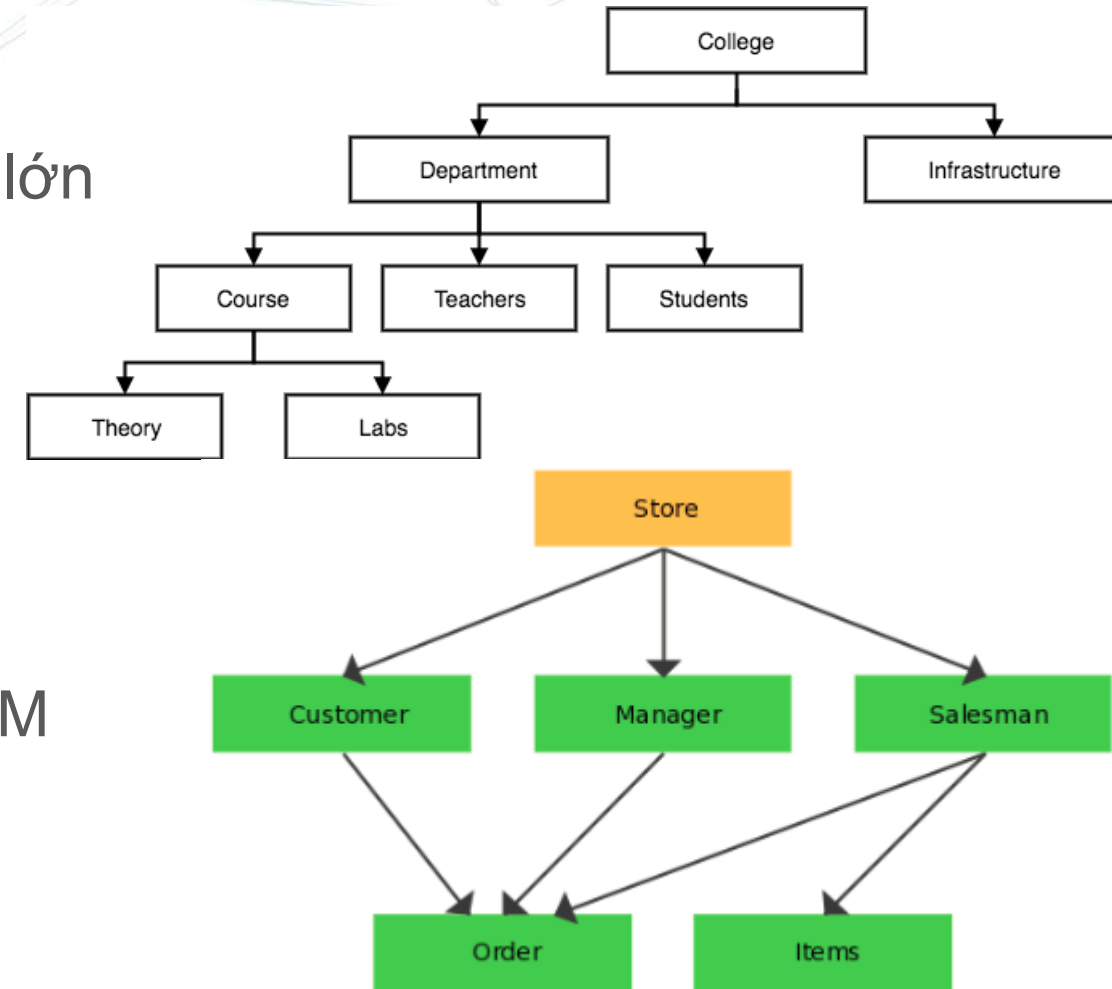
- Trong quá trình chuyển đổi các quy tắc nghiệp vụ thành các thành phần của mô hình dữ liệu, cần xác định:
  - Thực thể, Thuộc tính, Mối quan hệ và Ràng buộc
- Đặt tên các đối tượng là duy nhất và phân biệt được
- Tên mô tả đối tượng phải quen thuộc với người dùng trong môi trường sử dụng hàng ngày
- Cách đặt tên đúng:
  - Tạo điều kiện để giao tiếp giữa các bên (NSD, Thiết kế, Lập trình,...)
  - Thống nhất một quy tắc chung
    - Ví dụ: SINHVIEN(MaSinhVien, HoTen, NgaySinh)

# Sự phát triển của các mô hình dữ liệu

## EVOLUTION OF MAJOR DATA MODELS

GENERATION	TIME	DATA MODEL	EXAMPLES	COMMENTS
First	1960s–1970s	File system	VMS/VSAM	Used mainly on IBM mainframe systems Managed records, not relationships
Second	1970s	Hierarchical and network	IMS, ADABAS, IDS-II	Early database systems Navigational access
Third	Mid-1970s	Relational	DB2 Oracle MS SQL Server MySQL	Conceptual simplicity Entity relationship (ER) modeling and support for relational data modeling
Fourth	Mid-1980s	Object-oriented Object/relational (O/R)	Versant Objectivity/DB DB2 UDB Oracle 12c	Object/relational supports object data types Star Schema support for data warehousing Web databases become common
Fifth	Mid-1990s	XML Hybrid DBMS	dbXML Tamino DB2 UDB Oracle 12c MS SQL Server	Unstructured data support O/R model supports XML documents Hybrid DBMS adds object front end to relational databases Support large databases (terabyte size)
Emerging Models: NoSQL	Early 2000s to present	Key-value store Column store	SimpleDB (Amazon) BigTable (Google) Cassandra (Apache) MongoDB Riak	Distributed, highly scalable High performance, fault tolerant Very large storage (petabytes) Suited for sparse data Proprietary application programming interface (API)

- Mô hình phân cấp:
  - Phát triển năm 1960, để quản lý số lượng lớn dữ liệu cho các dự án
  - Cấu trúc logic được biểu diễn dạng cây từ trên xuống
  - Cấu trúc chứa các cấp hoặc phân đoạn
- Mô hình mạng:
  - Tập hợp các bản ghi trong mối quan hệ 1:M
  - Gồm 2 loại bản ghi: Sở hữu và thành viên
  - Cải thiện thực thi của CSDL
  - Áp dụng tiêu chuẩn dữ liệu





- Phát triển bởi E. F. Codd (IBM) năm 1970
- Nền tảng của mô hình quan hệ là toán quan hệ
- Bảng (Quan hệ)
  - Ma trận, bao gồm các giao nhau giữa hàng và cột
  - Mỗi hàng trong bảng gọi là 1 bộ (tuple)
  - Mỗi cột là 1 thuộc tính
- Hệ quản trị dữ liệu quan hệ(RDBMS)
  - Là một phần mềm QL các CSDLQH
  - Một số RDBMS:
    - MySQL, Oracle - Oracle
    - DB2 and Informix Dynamic Server - IBM
    - SQL Server, Access – Microsoft
    - ...

Table also called Relation

Primary Key Domain  
Ex: NOT NULL

© guru99.com

CustomerID	CustomerName	Status
1	Google	Active
2	Amazon	Active
3	Apple	Inactive

Tuple OR Row  
Total # of rows is Cardinality

Column OR Attributes  
Total # of column is Degree



- Sơ đồ quan hệ:
  - Biểu diễn các thực thể, thuộc tính và mối quan hệ
- Bảng lưu trữ tập hợp các thực thể có liên quan
  - Liên kết giữa 2 bảng cho phép kết nối 2 bảng

## A RELATIONAL DIAGRAM

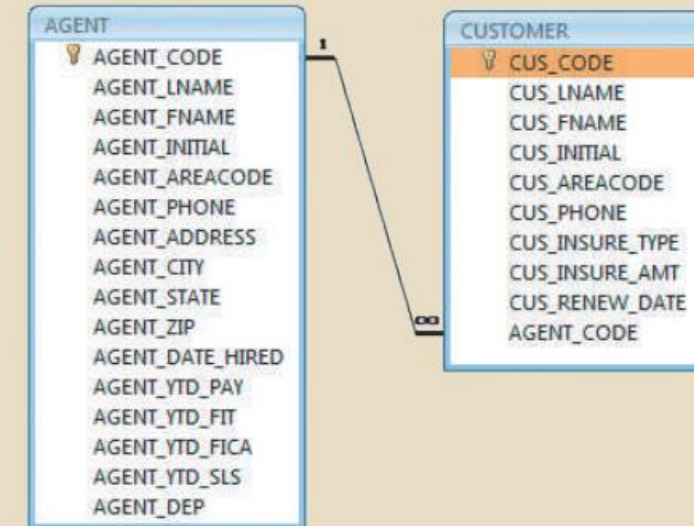


Table name: AGENT (first six attributes)

AGENT_CODE	AGENT_LNAME	AGENT_FNAME	AGENT_INITIAL	AGENT_AREACODE	AGENT_PHONE
501	Alby	Alex	B	713	228-1249
502	Hahn	Leah	F	615	882-1244
503	Okon	John	T	615	123-5589

Database name: Ch02\_InsureCo

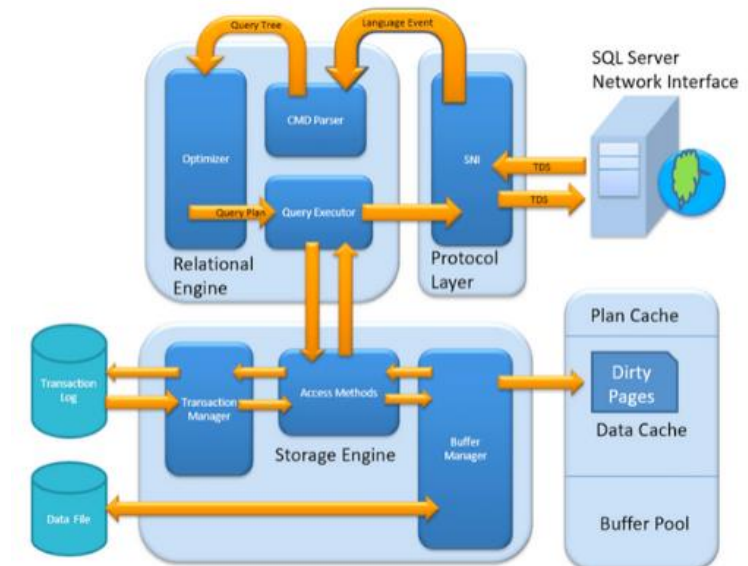
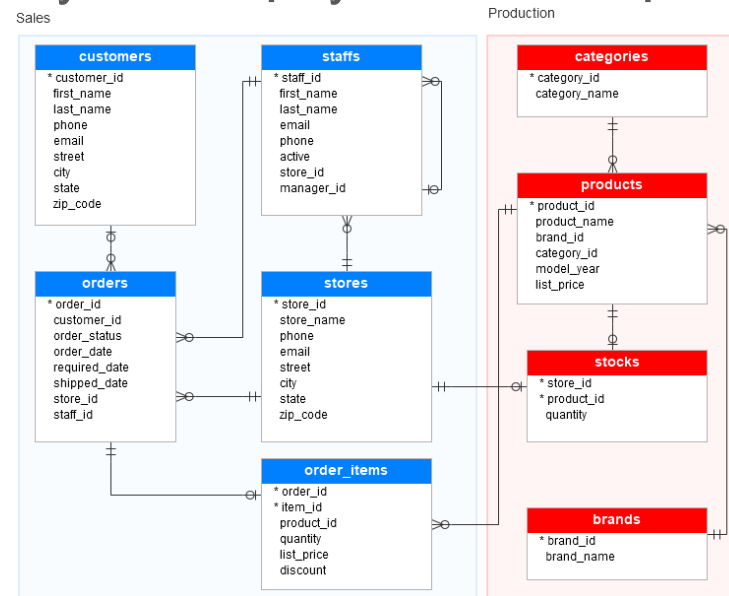
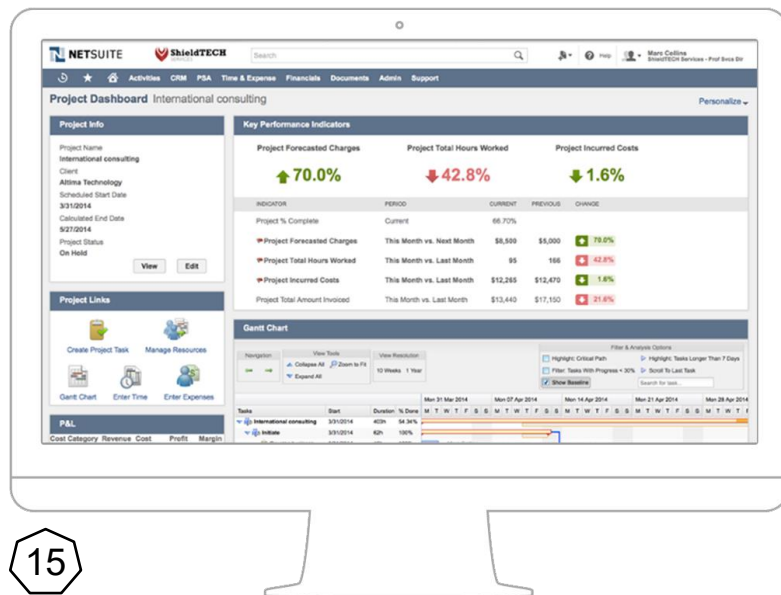
Link through AGENT\_CODE

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_FNAME	CUS_INITIAL	CUS_AREACODE	CUS_PHONE	CUS_INSURE_TYPE	CUS_INSURE_AMT	CUS_RENEW_DATE	AGENT_CODE
10010	Ramas	Alfred	A	615	844-2573	T1	100.00	05-Apr-2018	502
10011	Dunne	Leona	K	713	894-1238	T1	250.00	16-Jun-2018	501
10012	Smith	Kathy	W	615	894-2285	S2	150.00	29-Jan-2019	502
10013	Olowski	Paul	F	615	894-2180	S1	300.00	14-Oct-2018	502
10014	Orlando	Myron		615	222-1672	T1	100.00	28-Dec-2019	501
10015	O'Brian	Amy	B	713	442-3381	T2	850.00	22-Sep-2018	503
10016	Brown	James	G	615	297-1228	S1	120.00	25-Mar-2019	502
10017	Williams	George		615	290-2556	S1	250.00	17-Jul-2018	503
10018	Farriss	Anne	G	713	382-7185	T2	100.00	03-Dec-2018	501
10019	Smith	Olette	K	615	297-3809	S2	500.00	14-Mar-2019	503

# Mô hình quan hệ

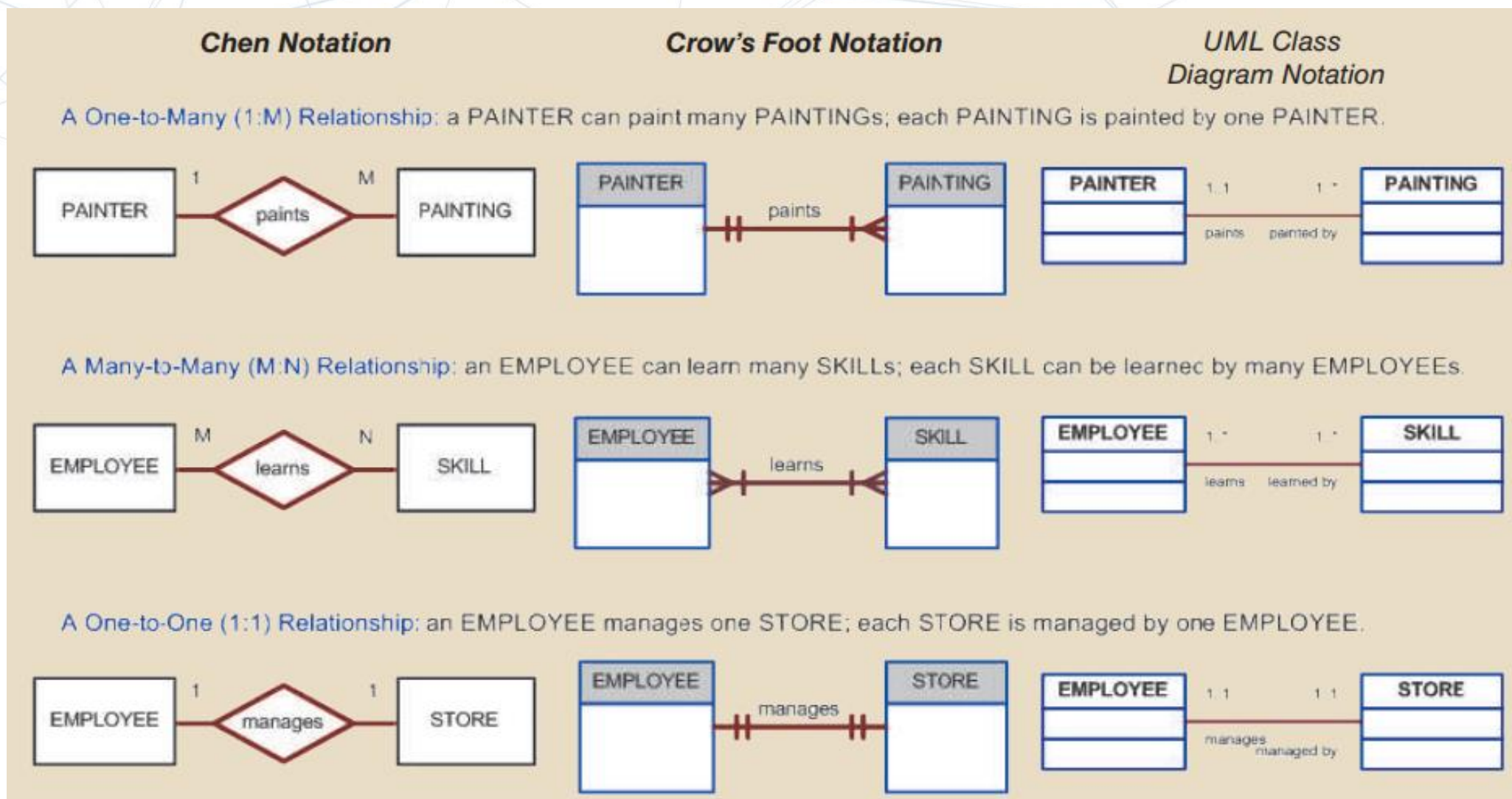
- Ứng dụng cơ sở dữ liệu quan hệ dựa trên SQL gồm:
  - Giao diện người dùng cuối
    - Cho phép người dùng tương tác với dữ liệu
  - Tập hợp các bảng được lưu trong CSDL
    - Mỗi bảng độc lập với bảng khác
  - Máy chạy SQL:
    - Thực hiện các câu truy vấn hoặc yêu cầu dữ liệu



- Đây là mô hình được chấp nhận rộng rãi cho mô hình dữ liệu
- Giới thiệu bởi Chen, 1976
- Biểu diễn dưới dạng đồ họa, thể hiện các thực thể và mối quan hệ.
- Sơ đồ quan hệ thực thể (Entity relationship diagram - ERD)
  - Sử dụng biểu tượng hình ảnh để mô hình các thành phần của CSDL
  - Thực thể được ánh xạ với 1 bảng dữ liệu
- Mỗi thực thể là 1 dòng trong bảng
- Tập thực thể là 1 bảng gồm các thực thể giống nhau
- Có 2 loại ký hiệu: Chen & Crow's foot

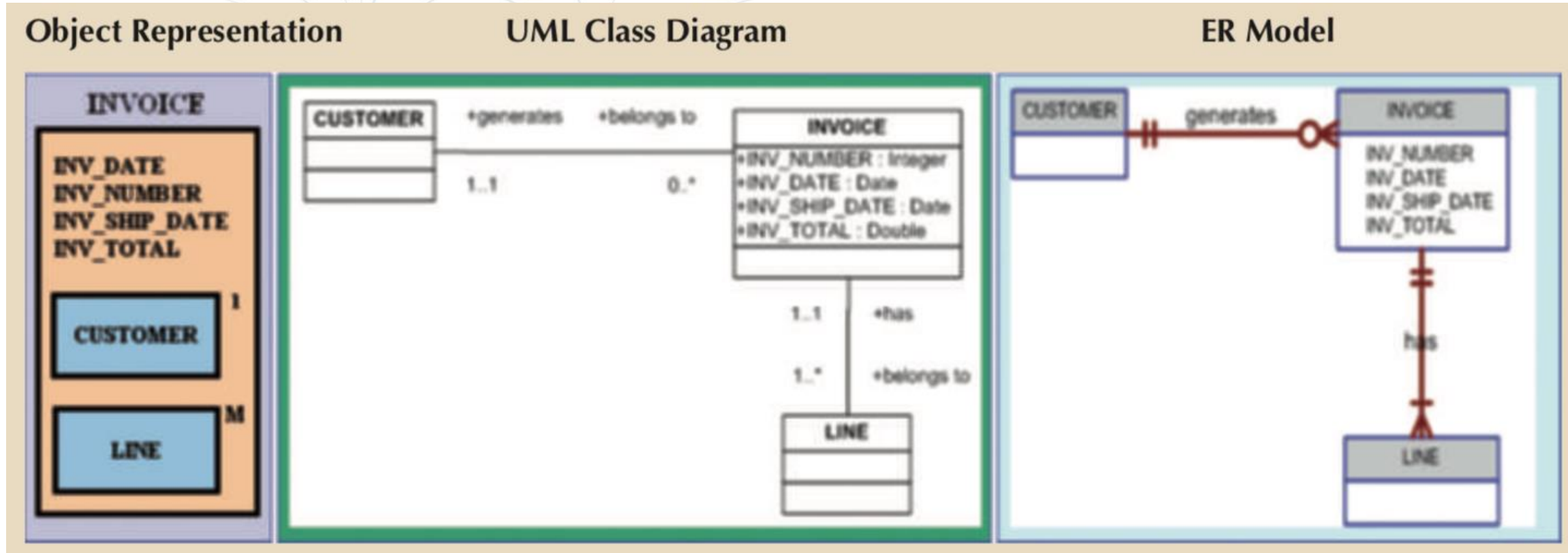


- Ký hiệu:



- Dữ liệu và mối quan hệ được chứa trong 1 cấu trúc duy nhất gọi là đối tượng
- Mô hình dữ liệu hướng đối tượng (object-oriented data model - OODM) là cơ sở cho hệ quản trị CSDL hướng đối tượng
- Đối tượng:
  - Bao gồm các thao tác
  - Chứa đựng cấu trúc riêng
  - Là sự trừu tượng của 1 thực thể trong thực tế
- UML dựa trên khái niệm hướng đối tượng để mô tả sơ đồ và kí hiệu sử dụng đồ họa.

- So sánh: Hướng đối tượng, UML và ERM



- Là sự mở rộng của mô hình dữ liệu quan hệ
  - Mô hình dữ liệu quan hệ thêm các tính năng hướng đối tượng
  - Tạo ra CSDL quan hệ có hỗ trợ thêm tính năng của hướng đối tượng
- XML:
  - Internet phát triển → nhu cầu trao đổi thông tin
  - Ngôn ngữ đánh dấu mở rộng (Extensible Markup Language –XML)
  - Là tiêu chuẩn cho trao đổi dữ liệu hiệu quả của dữ liệu có cấu trúc



```
<?xml version="1.0" encoding="utf-8"?>
```

```
<books>
```

```
<book isbn="978-1594489501">
```

```
<title> Lưới trời ai dệt </title>
```

```
<author> Nguyễn Tường Bách </author>
```

```
<publisher> NXB Lao động </publisher>
```

```
<price>24.27</price>
```

```
</book>
```

```
<book isbn="978-1594489587">
```

```
<title> Tôi tự học </title>
```

```
<author> Nguyễn Duy Cần </author>
```

```
<publisher> NXB Thanh niên </publisher>
```

```
<price>14.97</price>
```

```
</book>
```

```
<book isbn="978-0545010221">
```

```
<title> Tầm ván phóng dao </title>
```

```
<author> Mạc can </author>
```

```
<publisher> NXB Văn hoá </publisher>
```

```
<price>9.24</price>
```

```
</book>
```

```
</books>
```

## XML

```
<empinfo>
```

```
<employees>
```

```
<employee>
```

```
<name>James Kirk</name>
```

```
<age>40</age>
```

```
</employee>
```

```
<employee>
```

```
<name>Jean-Luc Picard</name>
```

```
<age>45</age>
```

```
</employee>
```

```
<employee>
```

```
<name>Wesley Crusher</name>
```

```
<age>27</age>
```

```
</employee>
```

```
</employees>
```

```
</empinfo>
```

## JSON

```
{ "empinfo" :
```

```
{
```

```
  "employees" : [
```

```
{
```

```
    "name" : "James Kirk",
```

```
    "age" : 40,
```

```
  },
```

```
{
```

```
    "name" : "Jean-Luc Picard",
```

```
    "age" : 45,
```

```
  },
```

```
{
```

```
    "name" : "Wesley Crusher",
```

```
    "age" : 27,
```

```
  }
```

```
]
```

```
}
```

```
}
```

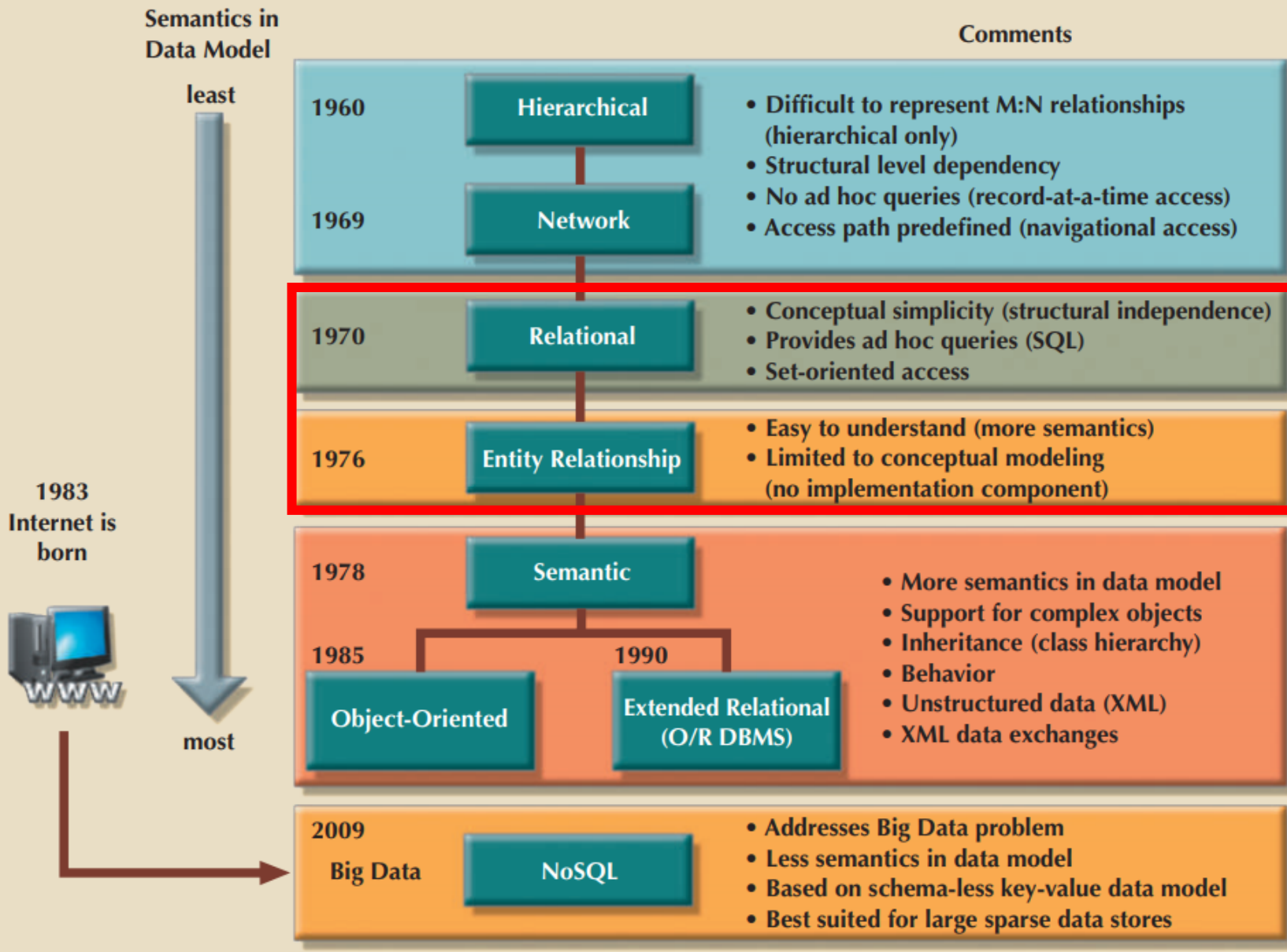
- Big Data:
  - Là một thuật ngữ cho việc xử lý một tập hợp dữ liệu rất lớn và phức tạp mà các ứng dụng xử lý dữ liệu truyền thống không xử lý được.
  - Thách thức như phân tích, thu thập, giám sát dữ liệu, tìm kiếm, chia sẻ, lưu trữ, truyền nhận, trực quan, truy vấn và tính riêng tư.
  - **Volume:** Khối lượng; **Velocity:** Tốc độ; **Variety:** Đa dạng
- NoSQL:
  - Không dựa trên mô hình quan hệ
  - Hỗ trợ kiến trúc CSDL phân tán
  - Cung cấp khả năng mở rộng, tính sẵn sàng, khả năng chịu lỗi
  - Hỗ trợ rất lớn dữ liệu thừa
  - Hướng tới hiệu suất thay vì tính nhất quán

- Mô hình Key – Value:
  - Có 2 phần tử dữ liệu: Khóa và Giá trị
    - Mỗi khóa có 1 giá trị tương ứng
- Dữ liệu thưa:
  - Số lượng lớn thuộc tính
  - Số lượng dữ liệu thực tế thấp

Car	
Key	Attributes
1	Make: Nissan Model: Pathfinder Color: Green Year: 2003
2	Make: Nissan Model: Pathfinder Color: Blue Color: Green Year: 2005 Transmission: Auto

Sparse Data												
Day	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
1-Jan	0	0.89	0	0	0	0	0	0	0	0	0	0.911
2-Jan	0	0	0	0	0	0	0	0	0	0	0	0.931
3-Jan	0	0.951	0	0	0	0	0	0	0	0	0	0.951
4-Jan	0.954	0.911	0	0	0	0	0	0	0	0	0	0.899
5-Jan	0	0	0	0	0	0	0	0	0	0	0	0.897
6-Jan	0	0.899	0	0	0	0	0	0	0	0	0	0.968
7-Jan	0.895	0.911	0	0	0	0	0	0	0	0	0	0.991
8-Jan	0.911	0.962	0	0	0	0	0	0	0	0	0	0.951
9-Jan	0	0.954	0	0	0	0	0	0	0	0	0	0.898
10-Jan	0.898	0.934	0	0	0	0	0	0	0	0	0	0.962

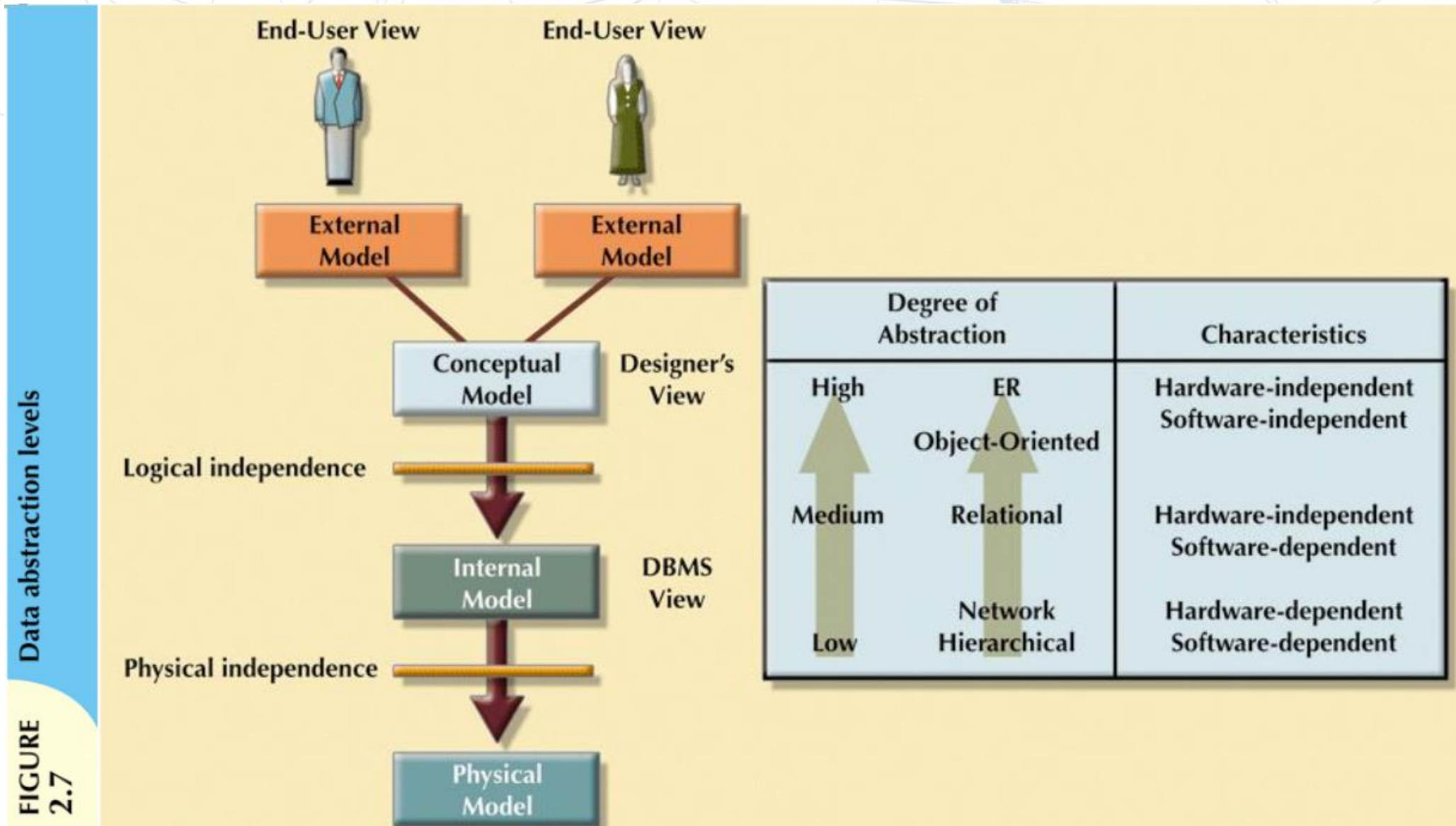
# Sự phát triển của cơ sở dữ liệu





- Người thiết kế CSDL bắt đầu từ quan điểm trừu tượng, sau đó đến chi tiết
  - Từ việc khái quát hóa các đối tượng, sau đó chi tiết các thuộc tính của đối tượng.
- Lập mô hình dữ liệu dựa trên mức độ trừu tượng hóa dữ liệu
  - Mô hình ngoài –The external model
  - Mô hình khái niệm –The concept model
  - Mô hình trong –The internal model

- Là cái nhìn của người dùng cuối về môi trường dữ liệu
- Biểu đồ thực thể quan hệ(ERD): thể hiện quan điểm ngoài
- Lược đồ bên ngoài: thể hiện chế độ xem bên ngoài
  - Thực thể, Quan hệ, Quy trình, Ràng buộc
- Dễ dàng xác định yêu cầu dữ liệu cụ thể để hỗ trợ hoạt động kinh doanh
- Tạo điều kiện cho người thiết kế bằng cách phản hồi về tính đầy đủ của mô hình dữ liệu
- Đảm bảo các ràng buộc bảo mật trong thiết kế
- Đơn giản hóa việc phát triển chương trình ứng dụng

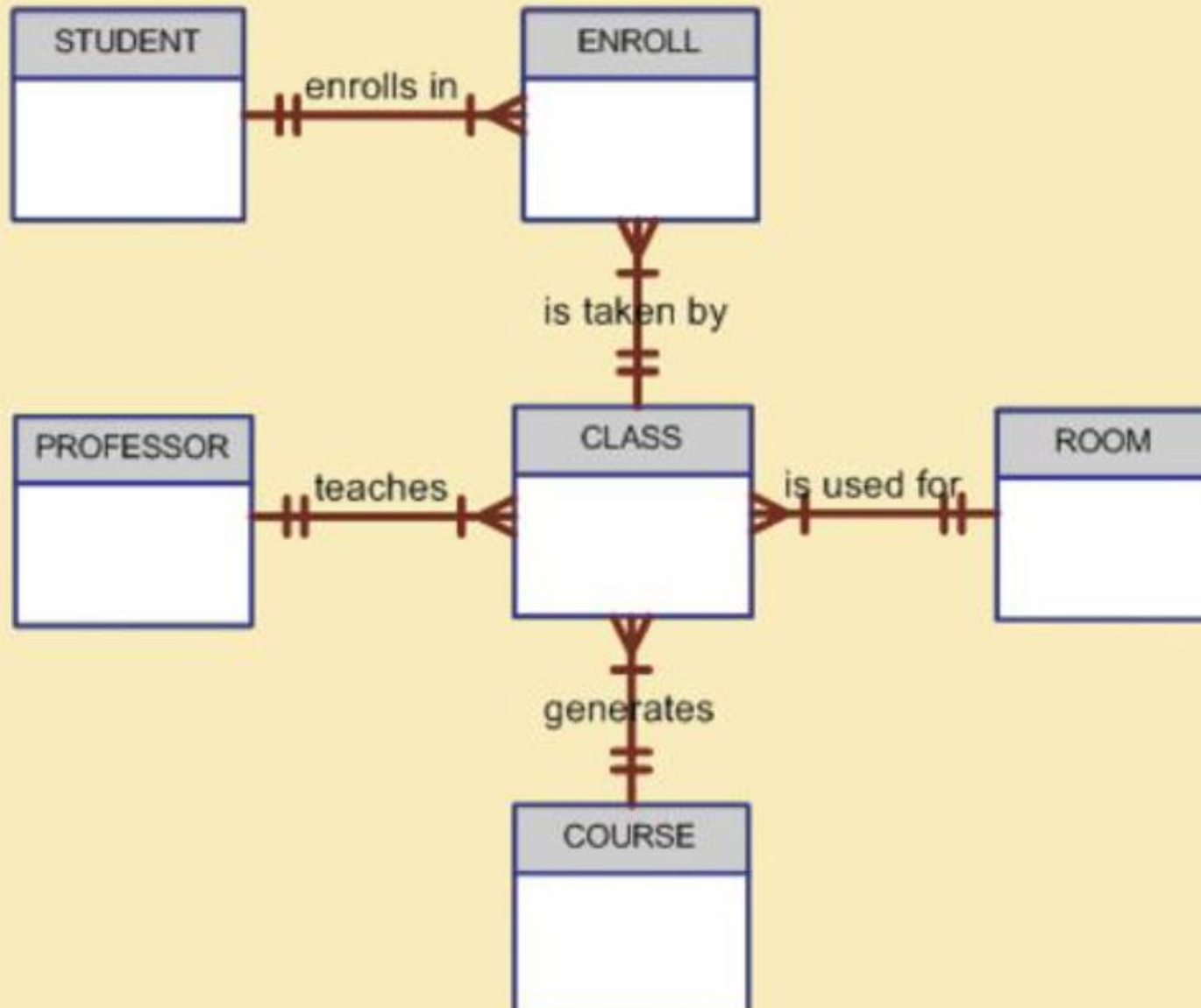


- Thể hiện chế độ xem tổng thể của cơ sở dữ liệu
- Tất cả các chế độ xem bên ngoài được tích hợp vào 1 chế độ xem tổng thể của cơ sở dữ liệu
- Mô hình thực thể quan hệ–ERD được sử dụng rộng rãi
  - Giao diện đồ họa của ERD thể hiện lược đồ khái niệm
- Cung cấp chế độ xem tổng thể tương đối dễ hiểu về môi trường cơ sở dữ liệu
- Độc lập với Phần mềm và Phần cứng:
  - Không phụ thuộc vào DBMS để triển khai mô hình
  - Không phụ thuộc vào phần cứng trong triển khai
  - Những thay đổi trong phần cứng hay phần mềm không ảnh hưởng đến thiết kế CSDL ở mức khái niệm

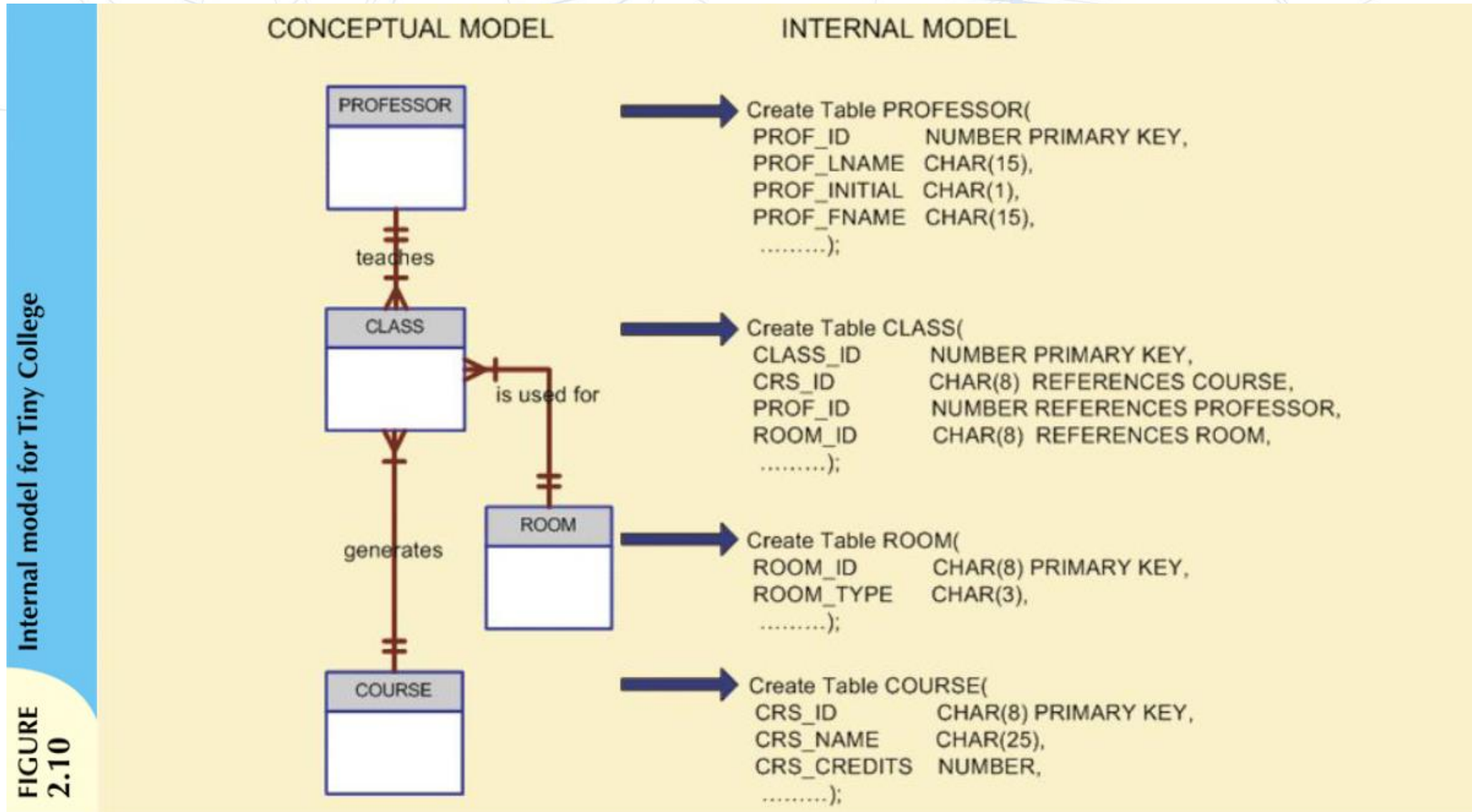


Conceptual model for Tiny College

FIGURE 2.9



- Biểu diễn dữ liệu mà DBMS có thể hiểu được
  - Ánh xạ mô hình khái niệm với DBMS
- Mô tả cụ thể về mô hình
- Phụ thuộc vào DBMS cụ thể
  - Thay đổi DBMS thì yêu cầu thay đổi mô hình bên trong
- Độc lập logic thay đổi mô hình trong cũng không ảnh hưởng đến mô hình khái niệm



Internal model for Tiny College

FIGURE 2.10

# Mô hình vật lý – The Physical Model

- Hoạt động ở mức độ thấp nhất
- Mô tả dữ liệu được lưu trên phương tiện lưu trữ
- Yêu cầu định nghĩa về lưu trữ vật lý và phương pháp truy cập dữ liệu
- Mô hình quan hệ hướng đến mức logic, không yêu cầu chi tiết mức vật lý
- Tính độc lập vật lý: những thay đổi trong mô hình vật lý không ảnh hưởng đến mô hình bên trong

TABLE  
2.4

Levels of Data Abstraction

MODEL	DEGREE OF ABSTRACTION	FOCUS	INDEPENDENT OF
External	High ↑ ↓ Low	End-user views	Hardware and software
Conceptual		Global view of data (database model independent)	Hardware and software
Internal		Specific database model	Hardware
Physical		Storage and access methods	Neither hardware nor software



3 Vs	entity relationship diagram (ERD)	object-oriented data model (OODM)	connectivity	internal model	relational diagram
American National Standards Institute (ANSI)	entity set	object-oriented database management system (OODBMS)	constraint	internal schema	relational model
attribute	extended relational data model (ERDM)	one-to-many (1:M or 1..*) relationship	Crow's Foot notation	logical design	relationship
Big Data	Extensible Markup Language (XML)	one-to-one (1:1 or 1..1) relationship	data definition language (DDL)	logical independence	schema
business rule	external model	physical independence	data manipulation language (DML)	MapReduce	segment
Chen notation	external schema	physical model	data model	many-to-many (M:N or *..*) relationship	semantic data model
class	Hadoop	relation	data modeling	method	software independence
class diagram	Hadoop Distributed File System (HDFS)	relational database management system (RDBMS)	data node	name node	subschema
class diagram notation	hardware independence	relational diagram	entity	network model	table
class hierarchy	hierarchical model		entity instance	NoSQL	tuple
client node	inheritance		entity occurrence	object	Unified Modeling Language (UML)
conceptual model			entity relationship (ER) model (ERM)	object/relational database management system (O/R DBMS)	
conceptual schema					

- Học viên trả lời và làm các bài tập trong tệp:
  - Chapter2\_Data\_Models.docx
- Đổi tên tệp thành STT\_HoVaTen\_Chapter2\_Data\_Models.docx
- Nộp bài lên hệ thống



# THANK YOU

**multicampus**

Copyright by Multicampus Co., Ltd. All right reserved