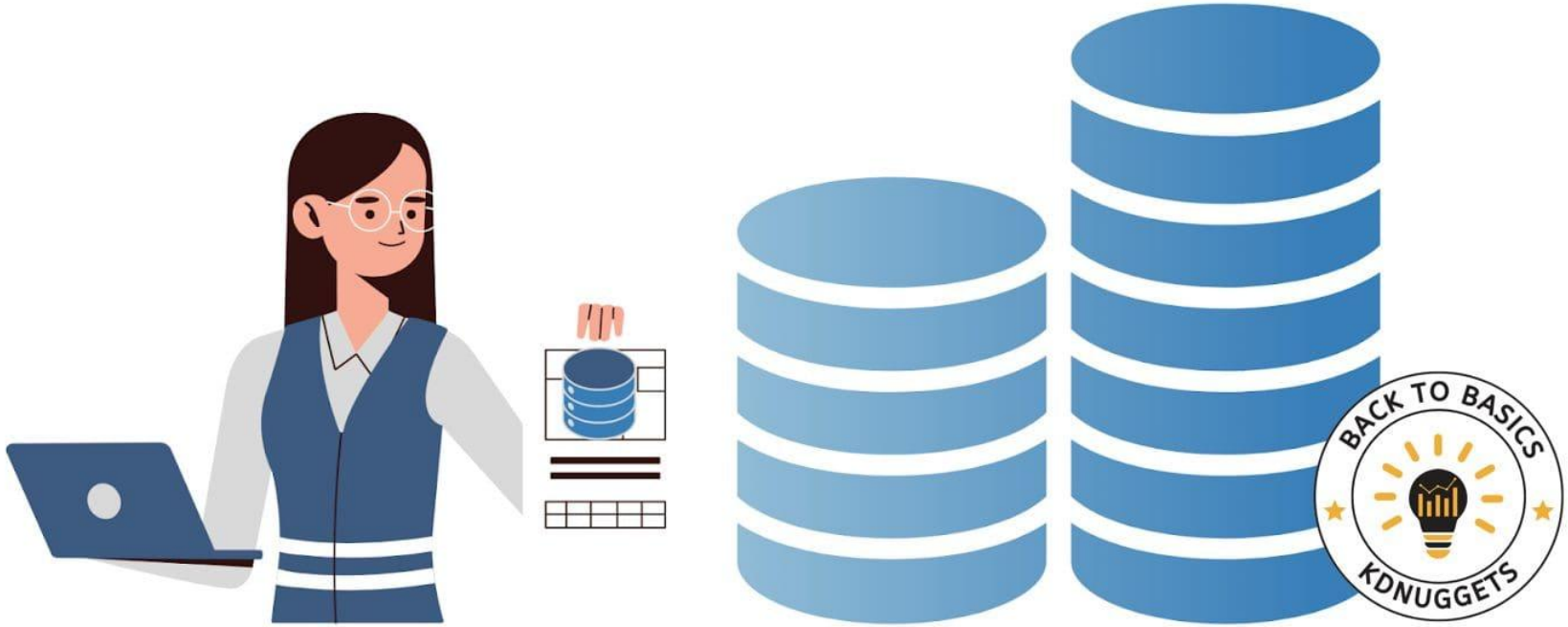
Intro to Databases



Data Models

Lecture 2

DATA MODELING

- Graphical, simple representation of data
- Must have:
  - Description of data structure
  - Set of enforceable rules to guarantee the integrity of data
  - Data manipulation methodology to support real world transformation
- Important!!

DATA MODEL BUILDING BLOCKS

Entity – Person, place, thing (noun) that is stored

Attribute – Characteristics of entity

Relationship –

- One to Many (1:M or 1..* or 1:∞)
- Many to Many (M:N or *..* or ∞:∞)
- One to One (1:1 or 1..1)

Constraints

BUSINESS RULES

Brief

Precise

Unambiguous

EXAMPLES OF BUSINESS RULES

Terms

- A manager is a person to whom 2 or more people report directly
- An order must have a buyer

Facts

- Each order can only have 1 discount
- Orders can contain multiple line items
- Customers must have a customer ID number

Constraints

- Total sale must not exceed $500
- Alcohol can only be purchased by people over 21.

MONTANA WIDGET COMPANY (MWC)

Sales Description

At MWC, we build the best widgets you could ask for! Our widgets are built in our state of the art facility efficiently and quickly. On any given day, we can ship out up to 1000 new widgets to local markets.

Our salespeople are working tirelessly with our customers to ensure that the best products are delivered on time and accurately. As a customer, you would be working with one assigned salesperson, who will get to know your unique needs. Each salesperson is limited to 10 customers, making sure that the personal aspect of the sales process is never lost.

You can either call, email or use our online e-store to create your order. Just pick the items and quantity you want, add them to your cart, enter your information and click submit. We will immediately get to work processing your widgets or building new ones (if we don't have them in stock). You'll get an email when your order is on its way!

Business Rules:

- A **customer** can generate many **invoices**
- Each **invoice** can only have one **customer**
- A 'cart' online represents an **invoice**
- An **invoice** can contain multiple **line items**
- Each **line item** is attached to only one **invoice**
- Each **line item** must contain a unique item and a quantity
- A **salesman** can have up to 10 **customers**
- Each **customer** can only have one **salesman** (at a time)

TABLE 2.1

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)

EVOLUTION OF DATA MODELS

- Schema – organization as seen by DBA
- SubSchema – as seen by application
- DML – Data Manipulation Language
- DDL – Data Definition Language

Definitions

Schema:

- A logical grouping of database objects, such as tables, indexes, views, and queries, that are related to each other.

Subschema:

- The portion of the database that interacts with application programs.

Definitions

DML:

- Data manipulation language (DML) The set of commands that allows an end user to manipulate the data in the database.

DDL:

- The language that allows a database administrator to define the database structure, schema, and subschema

The Relational Model (RM)

- The relational model was introduced in 1970
- In Relation (sometimes called a table) each row in a relation is called a **tuple**. Each column represents an **attribute**.
- Tables are related to each other through the sharing of a common attribute (a value in a column).

FIGURE 2.1 LINKING RELATIONAL TABLES

Table name: AGENT (first six attributes)

Database name: Ch02_InsureCo

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

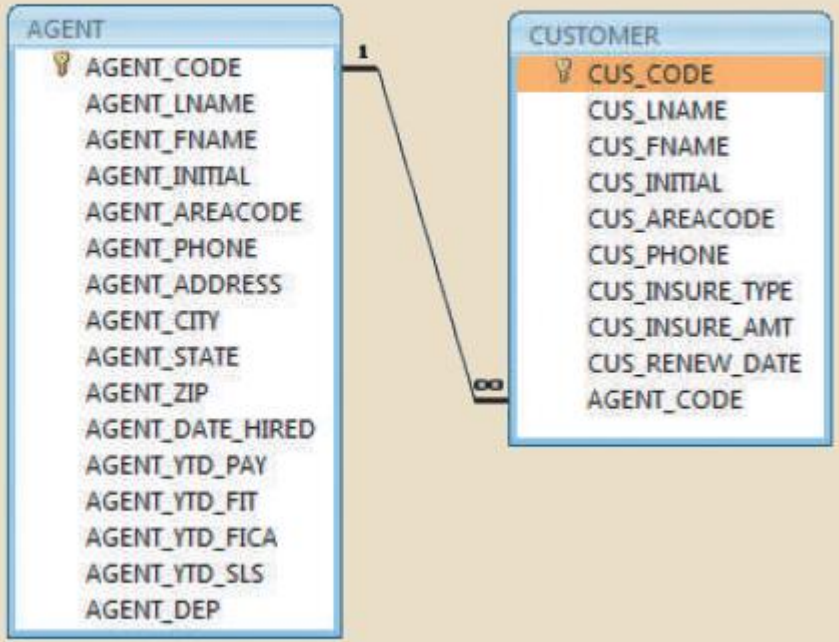
Link through AGENT_CODE

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

The Relational Model

A RELATIONAL DIAGRAM



Entity relationship (ER) Model (ERM)

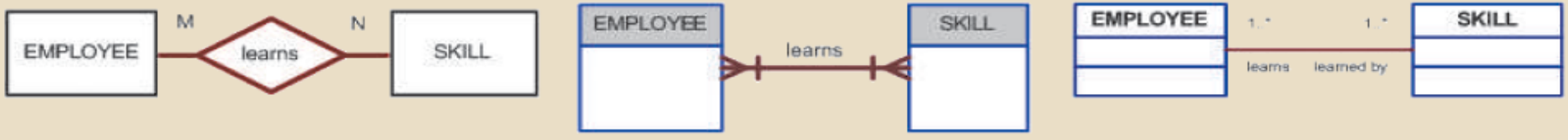
- Peter Chen first introduced the ER data model in 1976.
- normally represented in an entity relationship diagram (ERD).

RE 2.3 THE ER MODEL NOTATIONS

Chen Notation ***Crow's Foot Notation*** *UML Class Diagram Notation*

A One-to-Many (1:M) Relationship: a PAINTER can paint many PAINTINGs; each PAINTING is painted by one PAINTER.

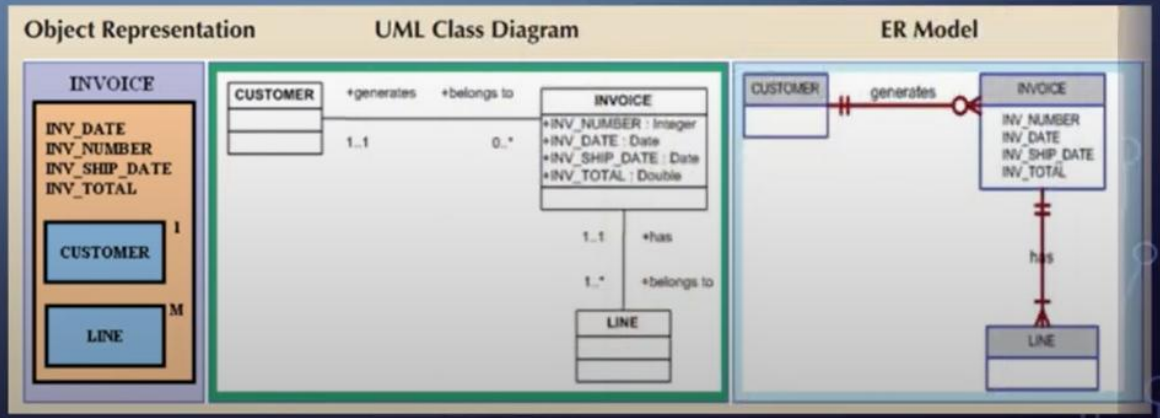
A Many-to-Many (M:N) Relationship: an EMPLOYEE can learn many SKILLs; each SKILL can be learnec by many EMPLOYEEs.



A One-to-One (1:1) Relationship: an EMPLOYEE manages one STORE; each STORE is managed by one EMPLOYEE.

OBJECT ORIENTED DATA MODEL (OODM)

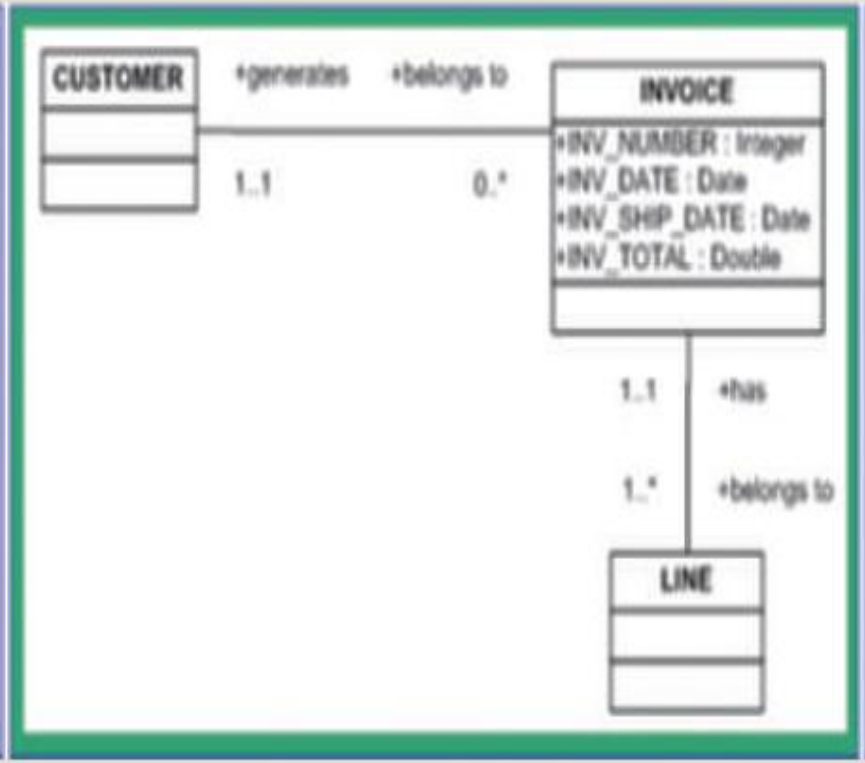
- Data Structure is based on objects (not relationships)
- Similar to classes : Object is a real world entity, Attributes are properties of the object
- Contains behaviors (methods, procedures)
- Class hierarchy / Inheritance
- Usually use a UML to describe



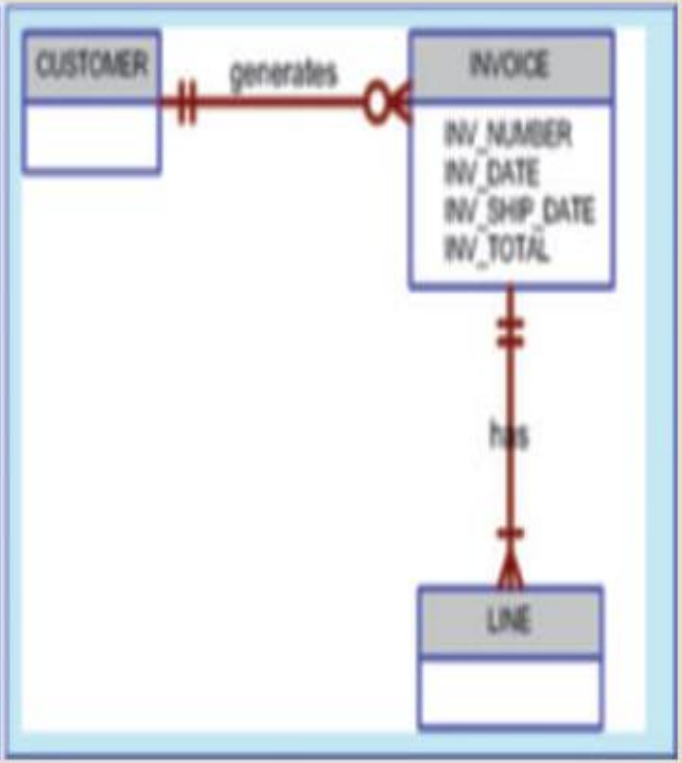
URE 2.4 A COMPARISON OF THE OO, UML, AND ER MODELS

Object Representation

UML Class Diagram



ER Model



EMERGING DATA MODELS (BIG DATA, NOSQL)

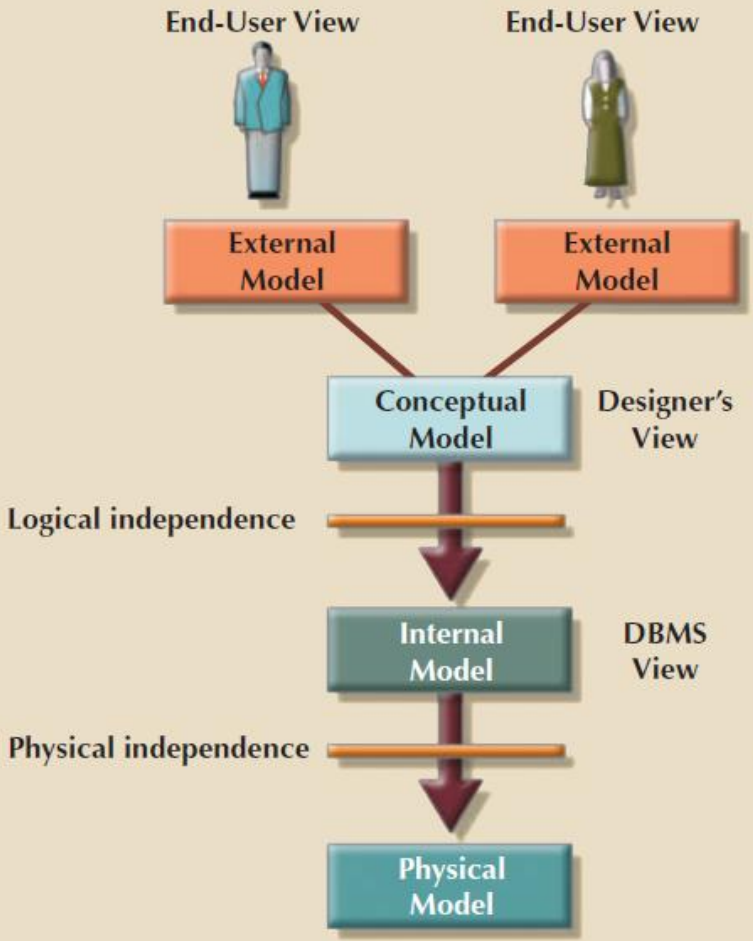
- 3 V's : Volume, Velocity, Variety
- When a relational database just isn't enough
- Hadoop, Hadoop Distributed File System, MapReduce
- NoSQL
  - Not relational based/Do not use SQL
  - High scalability, availability, fault tolerance, distributed database architecture
  - Large amounts of sparse data (lots of attributes, not many instances)
  - Performance based (vs. consistency)

What if:

- Blackberry responded to Apple's smartphone tech?
- MySpace responded to Facebook?
- Blockbuster responded to Netflix?
- Barnes and Noble responded to Amazon?

Degrees of Data Abstraction

FIGURE 2.6 DATA ABSTRACTION LEVELS

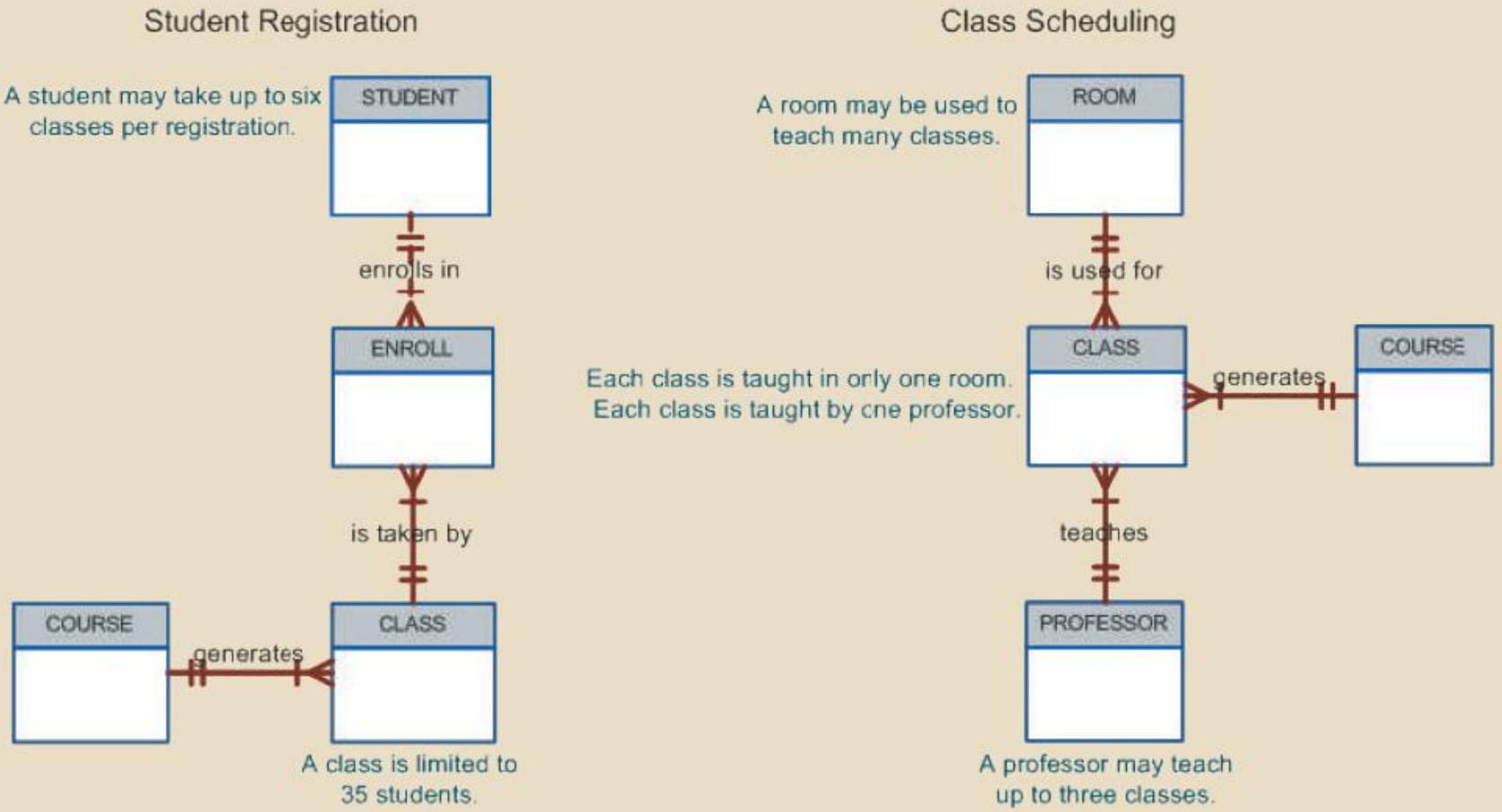


Degree of Abstraction		Characteristics
High	ER Object-Oriented	Hardware-independent Software-independent
Medium	Relational	Hardware-independent Software-dependent
Low	Network Hierarchical	Hardware-dependent Software-dependent

The External Model

- **External model:** is the end users' view an external model works with a data subset of the global database schema.
- **External schema** The specific representation of an external view. normally represented in an entity relationship diagram (ERD).

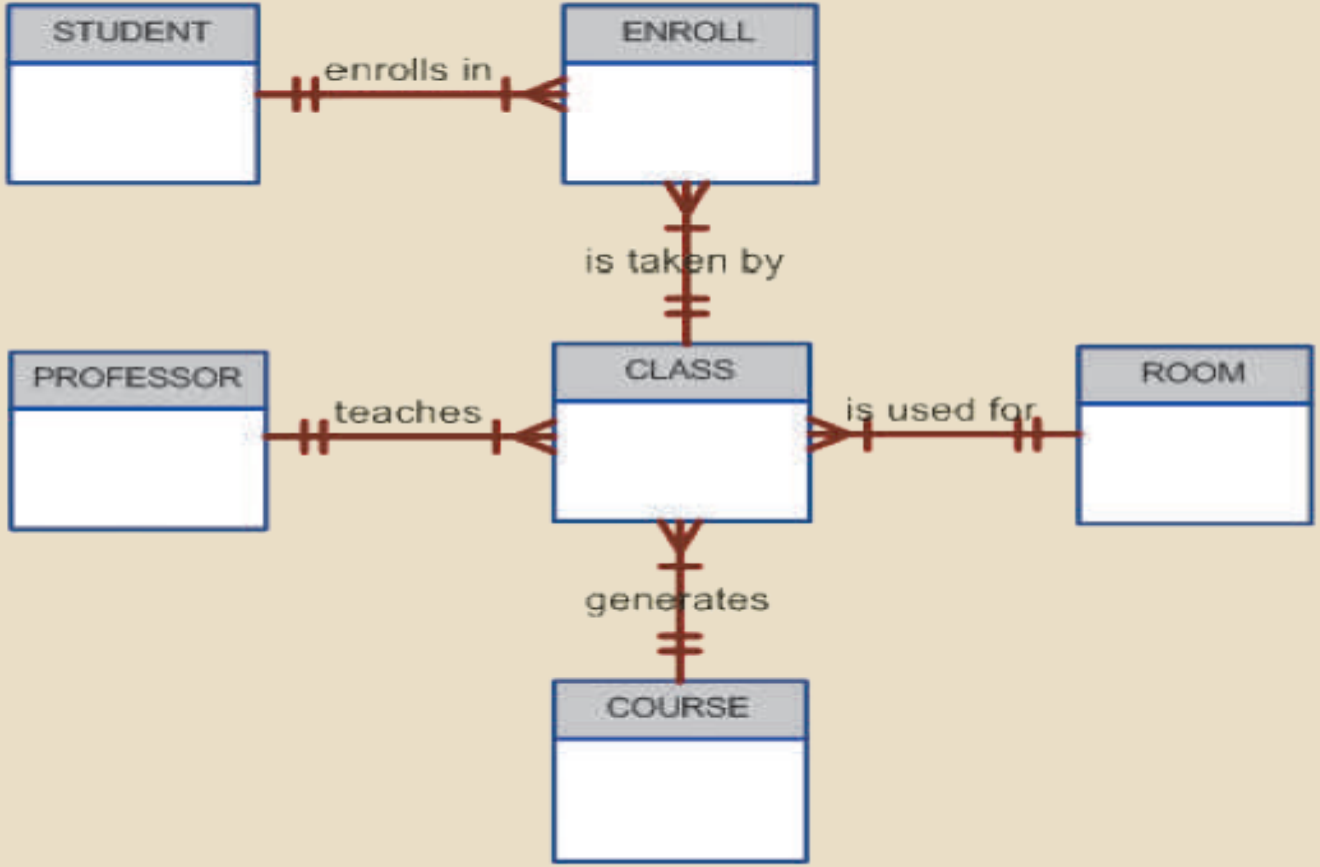
FIGURE 2.7 EXTERNAL MODELS FOR TINY COLLEGE



The Conceptual Model

- **Conceptual model:** The conceptual model provides a global view of an entire database and describes the main data objects, avoiding details.
- **Conceptual schema** A representation of the conceptual model, usually expressed graphically.

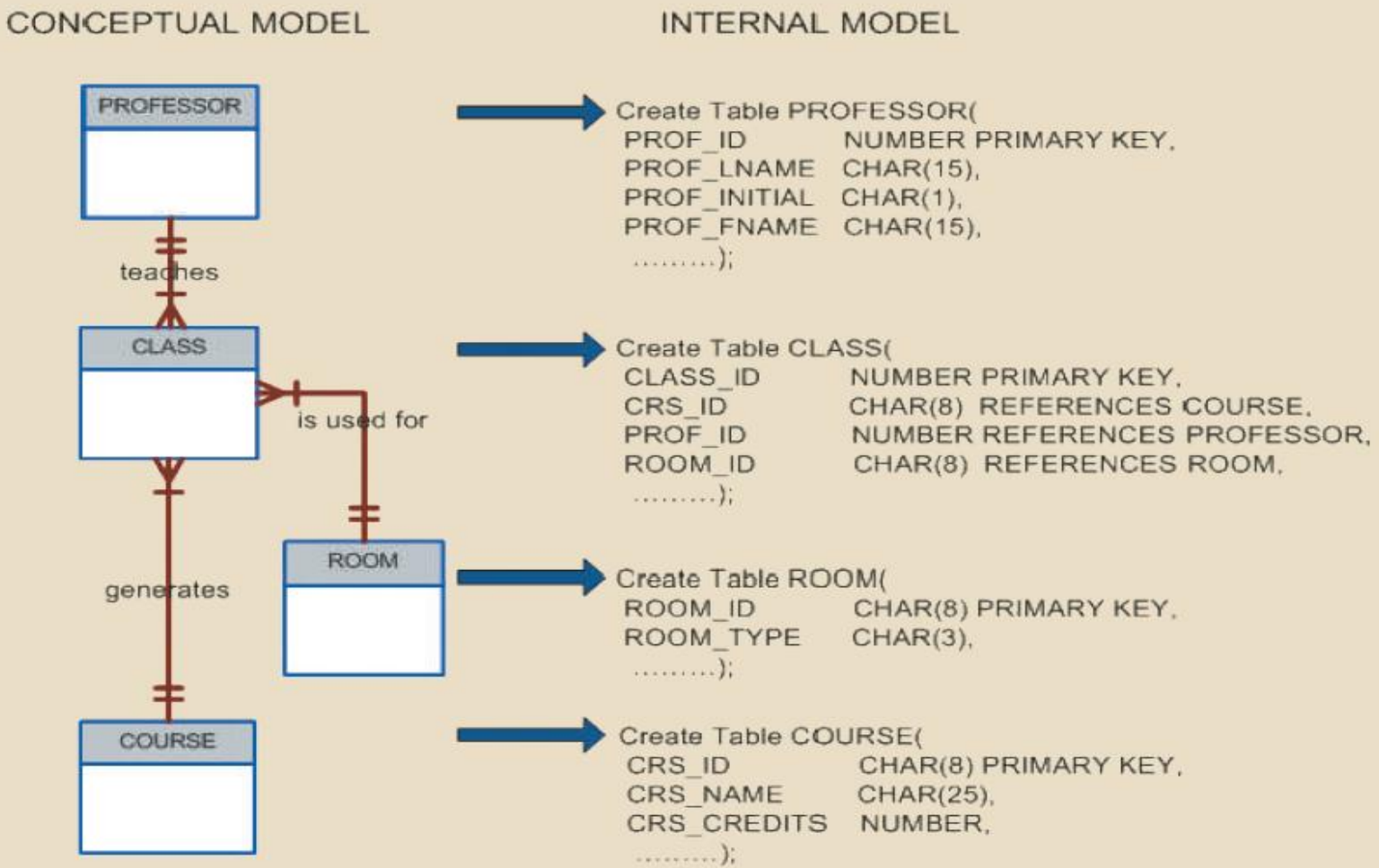
E 2.8 A CONCEPTUAL MODEL FOR TINY COLLEGE



The Internal Model

- **Internal model:** is the representation of the database as “seen” by the DBMS.
- **Internal schema** depicts a specific representation of an internal model.

FIGURE 2.9 INTERNAL MODEL FOR TINY COLLEGE



The Physical Model

- **Physical model:** operates at the lowest level of abstraction, describing the way data is saved on storage media such as magnetic, solid state, or optical media.