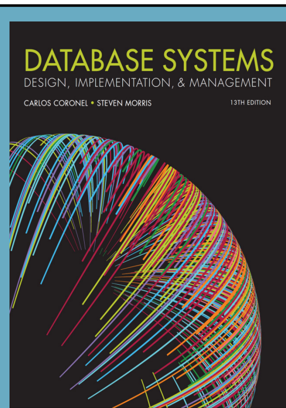


Chapter 2
Data Models

Learning Objectives

- After completing this chapter, you will be able to:
  - Discuss data modeling and why data models are important
  - Describe the basic data-modeling building blocks
  - Define what business rules are and how they influence database design
  - Understand how the major data models evolved
  - List emerging alternative data models and the needs they fulfill
  - Explain how data models can be classified by their level of abstraction

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Data Modeling and Data Models

- Data modeling: creating a specific data model for a determined problem domain
  - Data model: simple representation of complex real-world data structures
    - Useful for supporting a specific problem domain
  - Model: abstraction of a more complex real-world object or event

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Importance of Data Models

- The importance of data modeling cannot be overstated
  - Facilitates communication
  - Gives various views of the database
  - Organizes data for various users
  - Provides an abstraction for the creation of good a database

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Data Model Basic Building Blocks

- Entity: person, place, thing, or event about which data will be collected and stored
  - Attribute: characteristic of an entity
  - Relationship: association among entities
    - One-to-many (1:M OR 1..*)
    - Many-to-many (M:N or *..*)
    - One-to-one (1:1 OR 1..1)
  - Constraint: restriction placed on data
    - Ensures data integrity

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Business Rules

- Brief, precise, and unambiguous description of a policy, procedure, or principle
  - Create and enforce actions within that organization’s environment
  - Establish entities, relationships, and constraints

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Discovering Business Rules (1 of 2)

- Sources of business rules
  - Company managers
  - Policy makers
  - Department managers
  - Written documentation
  - Direct interviews with end users

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Discovering Business Rules (2 of 2)

- Reasons for identifying and documenting business rules
  - Standardize company's view of data
  - Facilitate communications tool between users and designers
  - Assist designers
    - Understand the nature, role, scope of data, and business processes
    - Develop appropriate relationship participation rules and constraints
    - Create an accurate data model

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Business Rules

- Examples of business rules are:
  - ✓An invoice contains one or more invoice lines.
  - ✓Each invoice line is associated with a single invoice.
  - ✓A store employs many employees.
  - ✓Each employee is employed by only one store.
  - ✓A college has many departments.
  - ✓Each department belongs to a single college. (This business rule reflects a university that has multiple colleges such as Business, Liberal Arts, Education, Engineering, etc.)

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Translating Business Rules into Data Model Components

- Business rules set the stage for the proper identification of entities, attributes, relationships, and constraints
  - Nouns translate into entities
  - Verbs translate into relationships among entities
- Relationships are bidirectional
  - Questions to identify the relationship type
    - How many instances of B are related to one instance of A?
    - How many instances of A are related to one instance of B?
  - For example, you can access the relationship between student and class
    - How many classes can one student enroll in? Answer: many classes
    - How many students can enroll in one class? Answer: many students

    Therefore, the relationship between student and class is many-to-many (M:N)

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Naming Conventions

- Entity name requirements
  - Be descriptive of the objects in the business environment
  - Use terminology that is familiar to the users
- Attribute name
  - Required to be descriptive of the data represented by the attribute
- Proper naming
  - Facilitates communication between parties
  - Promotes self-documentation

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Entity Relationship Model (1 of 2)

- Graphical representation of entities and their relationships in a database structure
  - Entity relationship diagram (ERD): uses graphic representations to model database components
  - Entity instance or entity occurrence: rows in the relational table
  - Attributes: describe particular characteristics
  - Connectivity: term used to label the relationship types

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

An example of a simple relational database

Table name: PAINTER

PAINTER_NUM	PAINTER_LNAME	PAINTER_FNAME	PAINTER_INITIAL
123	Ross	Georgette	P
126	Itero	Julio	G

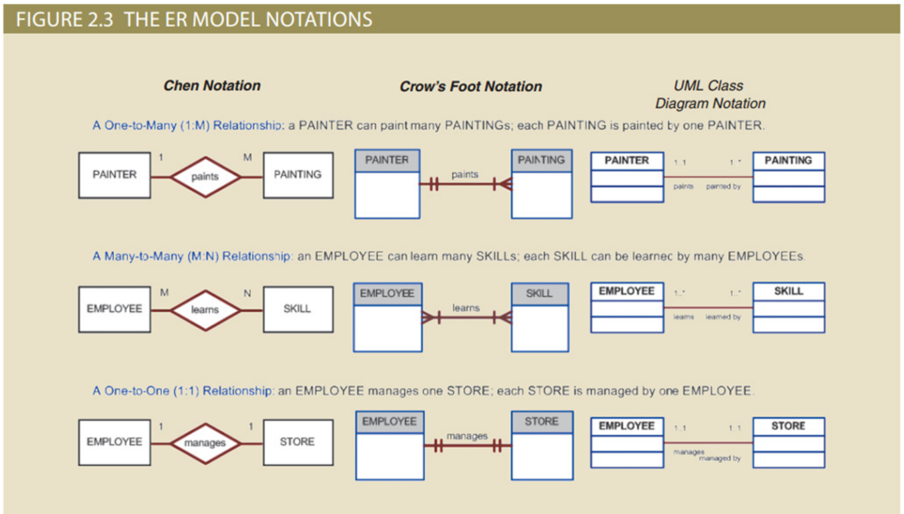
Table name: PAINTING

PAINTING_NUM	PAINTING_TITLE	PAINTER_NUM
1338	Dawn Thunder	123
1339	Vanilla Roses To Nowhere	123
1340	Tired Flounders	126
1341	Hasty Exit	123
1342	Plastic Paradise	126

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Entity Relationship Model (2 of 2)

FIGURE 2.3 THE ER MODEL NOTATIONS



© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Entity Relationship Model

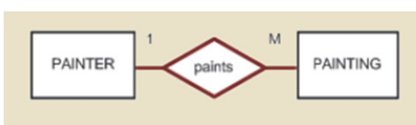
- Entity: person, place, thing, or event about which data will be collected and stored
- Entity name is usually written in capital letters and in singular form: PAINTER rather than PAINTERS, or EMPLOYEE rather than EMPLOYEES.
- Each entity is described by a set of attributes
- Attribute: characteristic of an entity such as a last name, phone number etc.
- Relationship: association among data or entities
  - One-to-many (1:M OR 1..*)
  - Many-to-many (M:N or *..*)
  - One-to-one (1:1 OR 1..1)
- Constraint: restriction placed on data
  - Ensures data integrity

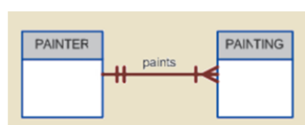
CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Entity Relationship Model

- ERD modelers use the term **connectivity** to label the types of relationships.
- Relationships are represented by a diamond connected to the related entities through a relationship line on the Chen model and the name of the relationship inside the diamond, or on the line connection line in the Crows foot notation.





CENGAGE

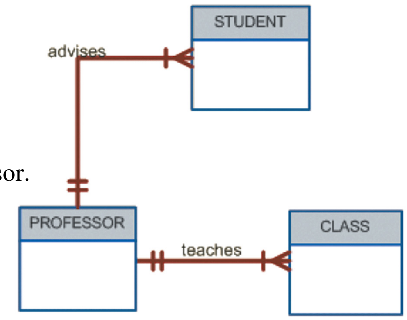
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Test your knowledge (teamwork)

- A professor in CUD is required to teach classes and also have office hours where he can advice students.
- Define the business rules and draw a simple Crow's foot Entity Relationship diagram indicating the relationships

The business rules may be written as follows:

- A professor can teach many classes.
- Each class is taught by one professor.
- A professor can advise many students.
- Each student is advised by one professor.



© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Test your knowledge (teamwork)

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

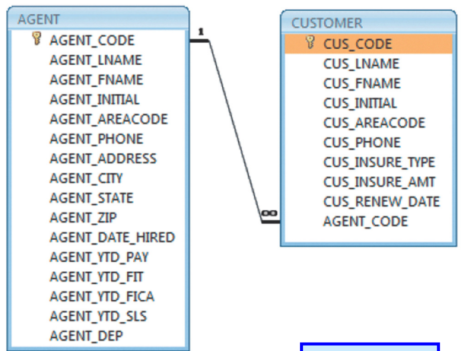
Link through AGENT_CODE

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

- Draw a relational diagram that govern the relationship between AGENT and CUSTOMER
- Write the business rule(s) that govern the relationship between AGENT and CUSTOMER
- Create the basic Crow's Foot and Chen's Model ERD

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.



- An AGENT can have many CUSTOMERs.
- Each CUSTOMER has only one AGENT
- Given these business rules, you can conclude that there is a 1:M relationship between AGENT and CUSTOMER.

Chen model

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Practice Question (Homework)

Table name: PAINTER **Database name: Ch02_UBA**

PAINTER_NUM	PAINTER_LNAME	PAINTER_FNAME	PAINTER_INITIAL
10014	Artiste	Josephine	P
10015	Itero	Julio	G
10016	McDonald	Theresa	

Table name: GALLERY

GALRY_NUMBER	GALRY_NAME	GALRY_WEB
18	Painter Place	www.painterplace.com
23	Art 's Us	www.artsus.com
24	Art Wonders	www.artwonders.com

Table name: PAINTING

PAINTING_NUM	PAINTING_TITLE	PAINTER_NUM	GALRY_NUMBER
20018	Dawn Thunder	10016	18
20023	Vanilla Roses To Nowhere	10015	18
20041	Tired Flounders	10016	23
20042	Hasty Exit	10015	24
20045	Plastic Paradise	10015	18
21003	Database Sunshine	10014	24
21987	Hierarchical Paths	10014	24
25108	File Systems Folly	10014	23

- United Broke Artists (UBA) is a broker for not-so-famous painters. UBA maintains a small network database to track painters, paintings, and galleries. A painting is painted by a particular artist, and that painting is exhibited in a particular gallery. A gallery can exhibit many paintings, but each painting can be exhibited in only one gallery. Similarly, a painting is painted by a single painter, but each painter can paint many paintings.

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Practice Question (Homework)

- Draw a relational diagram model for UBA
- Write the business rules for the UBA model
- Create the basic Crow's Foot ERD

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Object-Oriented Data Model (1 of 3)

- Both data and its relationships are contained in a single structure known as an object
  - Object-oriented database management system(OODBMS): based on OODM
- Object: contains data and their relationships with operations that are performed on it
  - Basic building block for autonomous structures
  - Abstraction of real-world entity
- Attribute: describes the properties of an object

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

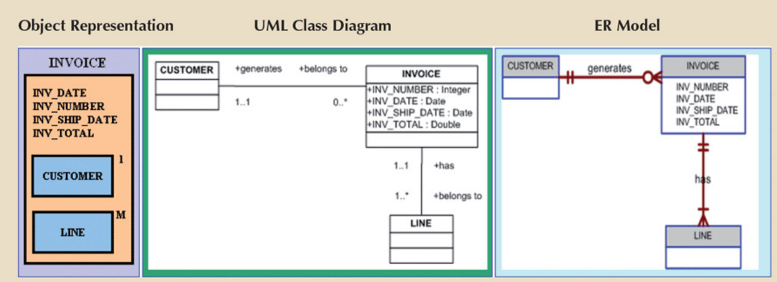
The Object-Oriented Data Model (2 of 3)

- Class: collection of similar objects with shared structure and behavior organized in a class hierarchy
- Class hierarchy: resembles a tree in which each class has only one parent
- Inheritance: object inherits methods and attributes of classes above it
- Unified Modeling Language (UML): describes sets of diagrams and symbols to graphically model a system

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Object-Oriented Data Model (3 of 3)

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



© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Big Data

- Over the last few years, a new wave of data has "emerged" to the limelight. Such data have always existed but did not receive the attention that is receiving today.
- These data are characterized for being
  - ✓ high volume (petabyte size and beyond),
  - ✓ high frequency (data are generated almost constantly),
  - ✓ and mostly semi-structured.
- These data come from multiple and varied sources such as web site logs, web site posts in social sites, and machine generated information (GPS, sensors, etc.)
- Such data have been accumulated over the years and companies are now awakening to the fact that it contains a lot of hidden information that could help the day-to-day business (such as browsing patterns, purchasing preferences, behavior patterns, etc.)
- The need to manage and leverage this data has triggered a phenomenon labeled "Big Data".

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Emerging Data Models: Big Data and NoSQL

- Goals of Big Data
  - Find new and better ways to manage large amounts of web and sensor-generated data
  - Provide high performance at a reasonable cost
- Characteristics of Big Data
  - Volume
  - Velocity
  - Variety

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Characteristics of Big Data – 3Vs

- The term “3 Vs” refers to the 3 basic characteristics of Big Data databases, they are:

Volume: Refers to the amounts of data being stored.

With the adoption and growth of the Internet and social media, companies have multiplied the ways to reach customers. Over the years, and with the benefit of technological advances, data for millions of e-transactions were being stored daily on company databases.

Furthermore, organizations are using multiple technologies to interact with end users and those technologies are generating mountains of data. This ever-growing volume of data quickly reached petabytes in size and it's still growing.

CENGAGE © 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Characteristics of Big Data – 3Vs

Velocity: Refers not only to the speed with which data grows but also to the need to process these data quickly in order to generate information and insight.

With the advent of the Internet and social media, business responses times have shrunk considerably. Organizations need not only to store large volumes of quickly accumulating data, but also need to process such data quickly.

The velocity of data growth is also due to the increase in the number of different data streams from which data is being piped to the organization (via the web, e-commerce, Tweets, Facebook posts, emails, sensors, GPS, and so on).

Variety: Refers to the fact that the data being collected comes in multiple different data formats.

A great portion of these data comes in formats not suitable to be handled by the typical operational databases based on the relational model.

CENGAGE © 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Emerging Data Models: Big Data and NoSQL

- Challenges of Big Data
  - Volume doesn’t allow usage of conventional structures
  - Expensive
  - OLAP tools proved inconsistent dealing with unstructured data
- New technologies of Big Data
  - Hadoop
  - Hadoop Distributed File System (HDFS)
  - MapReduce
  - NoSQL

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Hadoop and Mapreduce

- Hadoop is a Java based, open source, high speed, fault-tolerant distributed storage and computational framework.
- Hadoop uses low-cost hardware to create clusters of thousands of computer nodes to store and process data. Hadoop originated from Google's work on distributed file systems and parallel processing and is currently supported by the Apache Software Foundation.
- Hadoop has several modules, but the two main components are Hadoop Distributed File System (HDFS) and MapReduce.

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Hadoop and Mapreduce

- Hadoop Distributed File System (HDFS) is a highly distributed, fault-tolerant file storage system designed to manage large amounts of data at high speeds. In order to achieve high throughput, HDFS uses the write-once, read many model. This means that once the data is written, it cannot be modified.
- MapReduce is an open source application programming interface (API) that provides fast data analytics services.
- MapReduce distributes the processing of the data among thousands of nodes in parallel. MapReduce works with structured and nonstructured data.
- The MapReduce framework provides two main functions, Map and Reduce. In general terms, the Map function takes a job and divides it into smaller units of work; the Reduce function collects all the output results generated from the nodes and integrates them into a single result set.

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

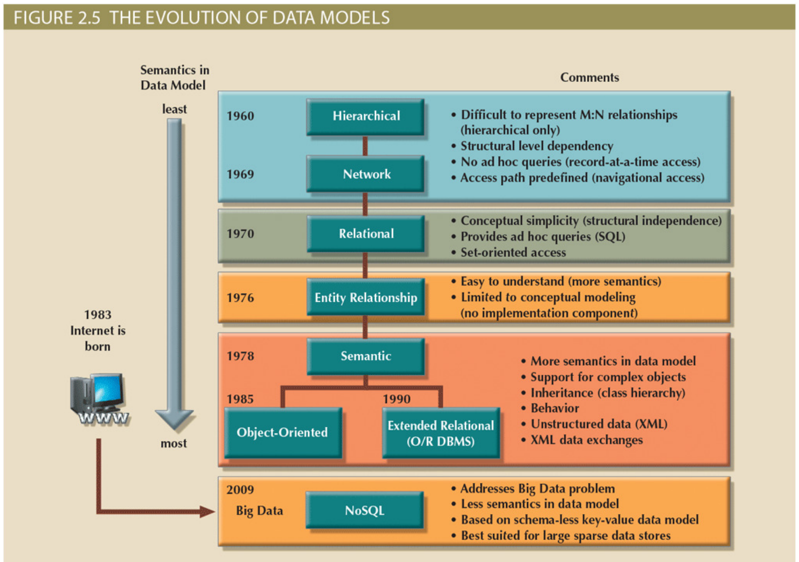
Emerging Data Models: Big Data and NoSQL

- NoSQL databases
  - Not based on the relational model
  - Support distributed database architectures
  - Provide high scalability, high availability, and fault tolerance
  - Support large amounts of sparse data
  - Geared toward performance rather than transaction consistency
  - Provides a broad umbrella for data storage and manipulation

CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Data Models: A Summary

FIGURE 2.5 THE EVOLUTION OF DATA MODELS

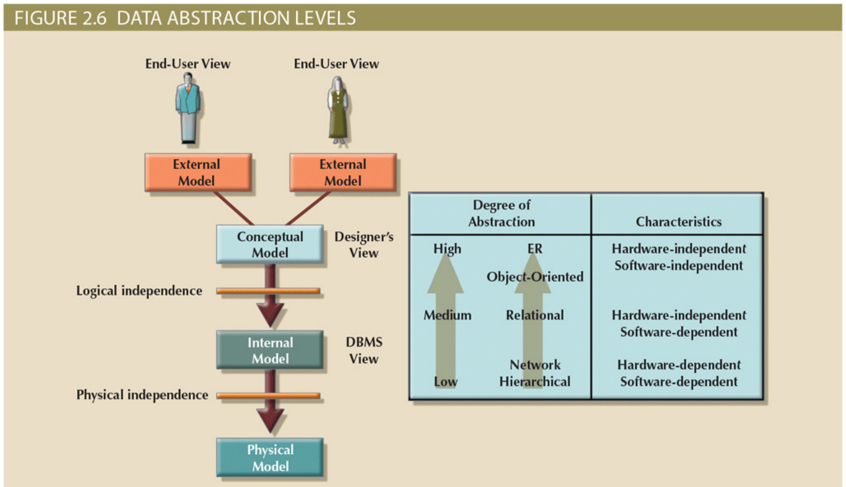


CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

Degrees of Data Abstraction

Consider construction details of an Architecture, Interior Designer, Structural Engineer, Contractor.

FIGURE 2.6 DATA ABSTRACTION LEVELS



CENGAGE
© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

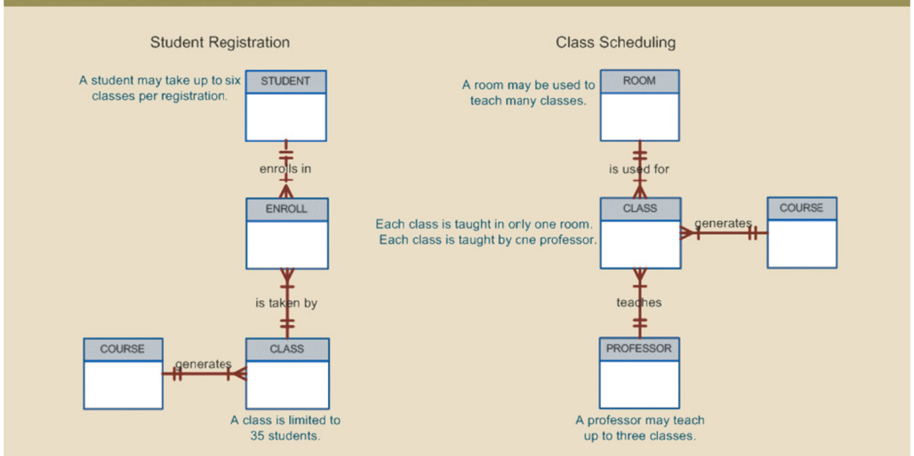
The External Model (1 of 2)

- End users' view of the data environment
  - People who use the application programs to manipulate the data and generate information
- ER diagrams are used to represent the external views
  - External schema: specific representation of an external view

CENGAGE © 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The External Model (2 of 2)

FIGURE 2.7 EXTERNAL MODELS FOR TINY COLLEGE



CENGAGE © 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

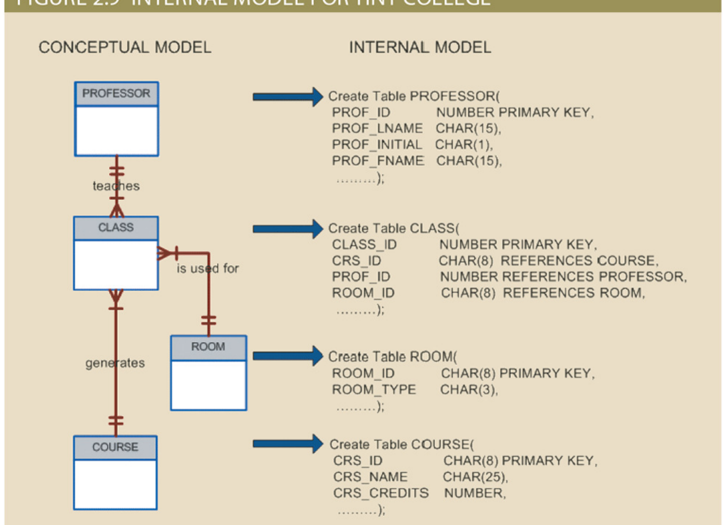
The Conceptual Model (1 of 2)

- Represents a global view of the entire database by the entire organization
  - Conceptual schema: basis for the identification and high-level description of the main data objects
  - Logical design: task of creating a conceptual data model
- Conceptual model advantages
  - Macro-level view of data environment
  - Software and hardware independent

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Conceptual Model (2 of 2)

FIGURE 2.9 INTERNAL MODEL FOR TINY COLLEGE



© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Internal Model

- Representing database as seen by the DBMS mapping conceptual model to the DBMS
  - Internal schema: specific representation of an internal model, using the database constructs supported by the chosen database
  - Logical independence: changing internal model without affecting the conceptual model
  - Hardware independent: unaffected by the type of computer on which the software is installed

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Physical Model (1 of 2)

- Operates at lowest level of abstraction
  - Describes the way data are saved on storage media such as magnetic, solid state, or optical media
- Requires the definition of physical storage and data access methods
  - Software and hardware dependent
- Relational model aimed at logical level
  - Does not require physical-level details
- Physical independence: changes in physical model do not affect internal model

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.

The Physical Model (2 of 2)

Table 2.4 Levels of Data Abstraction			
Model	Degree of Abstraction	Focus	Independent of
External	High	End-user views	Hardware and software
Conceptual	Medium-High	Global view of data (database model independent)	Hardware and software
Internal	Medium-Low	Specific database model	Hardware
Physical	Low	Storage and access methods	Neither hardware nor software

CENGAGE

© 2019 Cengage. May not be copied, scanned, or duplicated, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website for classroom use.