Chapter 2

Data Model

Database Systems:
Design, Implementation, and Management,
Sixth Edition, Rob and Coronel

In this chapter, you will learn:

- Why data models are important
- About the basic data-modeling building blocks
- What business rules are and how they affect database design
- How the major data models evolved, and their advantages and disadvantages
- How data models can be classified by level of abstraction

The Importance of Data Models

- Data model
  - Relatively simple representation, usually graphical, of complex real-world data structures
  - Communications tool to facilitate interaction among the designer, the applications programmer, and the end user
- Good database design uses an appropriate data model as its foundation

Importance of Data Modeling

- End-users have different views and needs for data
- Data model organizes data for various users

Data Model Basic Building Blocks

- Entity is anything about which data are to be collected and stored
- Attribute is a characteristic of an entity
- Relationship describes an association among (two or more) entities
  - **One-to-many (1:M) relationship**
  - **Many-to-many (M:N or M:M) relationship**
  - **One-to-one (1:1) relationship**

Business Rules

- Brief, precise, and unambiguous description of a policy, procedure, or principle within a specific organization's environment
- Apply to *any* organization that stores and uses data to generate information
- Description of operations that help to create and enforce actions within that organization's environment

Business Rules (continued)

- Must be rendered in writing
- Must be kept up to date
- Sometimes are external to the organization
- Must be easy to understand and widely disseminated
- Describe characteristics of the data *as viewed by the company*

Sources of Business Rules

- Company managers
- Policy makers
- Department managers
- Written documentation
  - Procedures
  - Standards
  - Operations manuals
- Direct interviews with end users

Importance of Business Rules

- Promote creation of an accurate data model
- Standardize company's view of data
- Constitute a communications tool between users and designers
- Allow designer to understand the nature, role, and scope of data
- Allow designer to understand business processes
- Allow designer to develop appropriate relationship participation rules and constraints

The Evolution of Data Models

- Hierarchical
- Network
- Relational
- Entity relationship
- Object oriented

The Hierarchical Model—Evolution

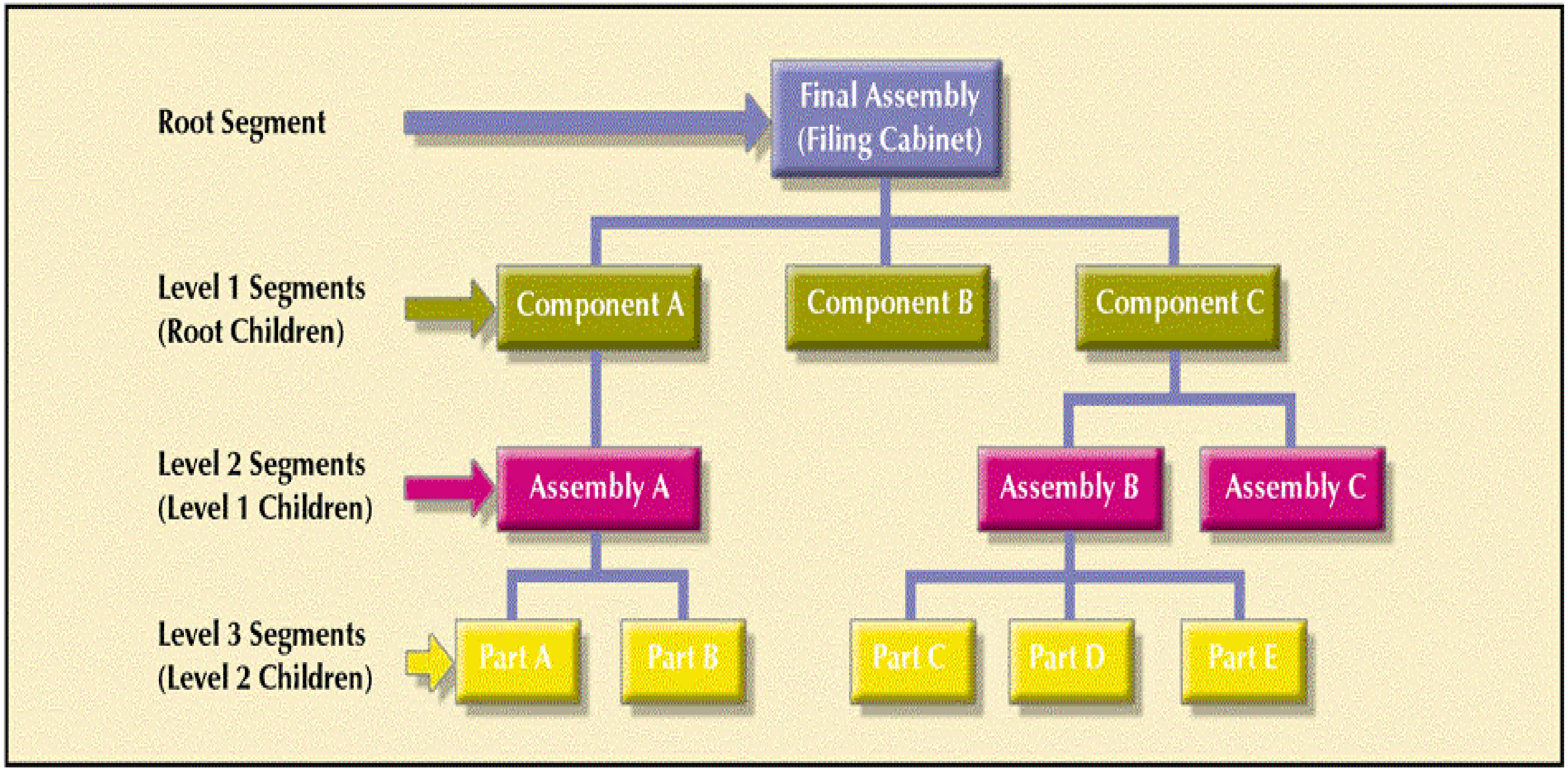
- GUAM (Generalized Update Access Method)
  - Based on the recognition that the many smaller parts would come together as components of still larger components
- Information Management System (IMS)
  - World's leading mainframe hierarchical database system in the 1970s and early 1980s

The Hierarchical Model—Characteristics

- Basic concepts form the basis for subsequent database development
- Limitations lead to a different way of looking at database design
- Basic concepts show up in current data models
- Best understood by examining manufacturing process

A Hierarchical Structure

FIGURE 2.1 A HIERARCHICAL STRUCTURE



Hierarchical Structure—Characteristics

- Each parent can have many children
- Each child has only one parent
- Tree is defined by path that traces parent segments to child segments, beginning from the left
- Hierarchical path
  - Ordered sequencing of segments tracing hierarchical structure
- Preorder traversal or hierarchic sequence
  - "Left-list" path

The Hierarchical Model

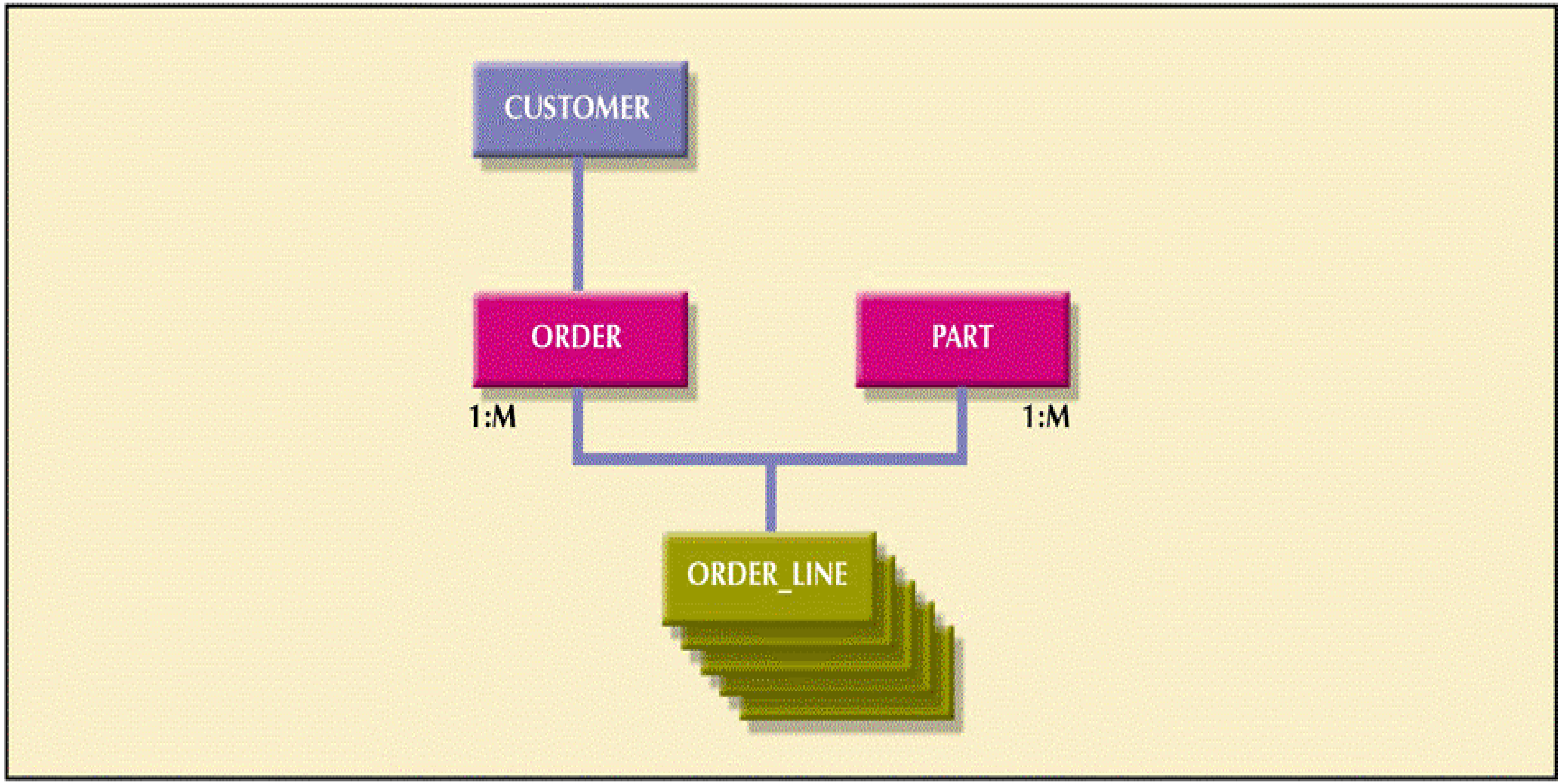
- Advantages
  - Conceptual simplicity
  - Database security
  - Data independence
  - Database integrity
  - Efficiency

The Hierarchical Model (continued)

- Disadvantages
  - Complex implementation
  - Difficult to manage
  - Lacks structural independence
  - Complex applications programming and use
  - Implementation limitations
  - Lack of standards

Child with Multiple Parents

FIGURE 2.2 CHILD WITH MULTIPLE PARENTS



The Network Model

- Created to
  - Represent complex data relationships more effectively
  - Improve database performance
  - Impose a database standard
- Conference on Data Systems Languages (CODASYL)
- American National Standards Institute (ANSI)
- Database Task Group (DBTG)

Crucial Database Components

- Schema
  - Conceptual organization of entire database as viewed by the database administrator
- Subschema
  - Defines database portion “seen” by the application programs that actually produce the desired information from data contained within the database
- Data Management Language (DML)
  - Define data characteristics and data structure in order to manipulate the data

Data Management Language Components

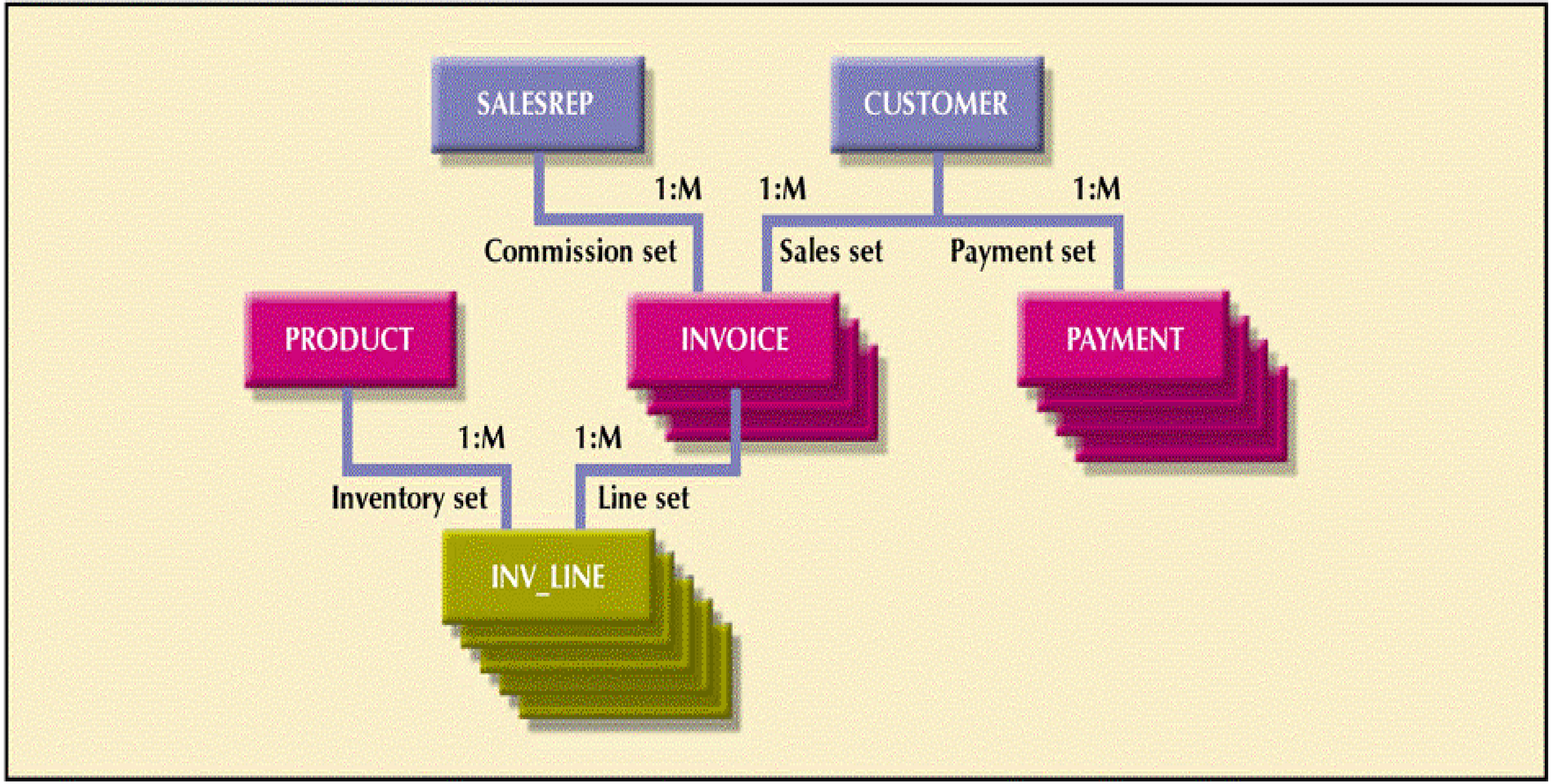
- Schema Data Definition Language (DDL)
  - Enables database administrator to define schema components
- Subschema DDL
  - Allows application programs to define database components that will be used
- DML
  - Manipulates database contents

Network Model—Basic Structure

- Resembles hierarchical model
- Collection of records in 1:M relationships
- Set
  - Relationship
  - Composed of at least two record types
    - Owner
      - Equivalent to the hierarchical model's parent
    - Member
      - Equivalent to the hierarchical model's child

A Network Data Model

FIGURE 2.3 A NETWORK DATA MODEL



The Network Data Model

- Advantages
  - Conceptual simplicity
  - Handles more relationship types
  - Data access flexibility
  - Promotes database integrity
  - Data independence
  - Conformance to standards

The Network Data Model (continued)

- Disadvantages
  - System complexity
  - Lack of structural independence

The Relational Model

- Developed by Codd (IBM) in 1970
- Considered ingenious but impractical in 1970
- Conceptually simple
- Computers lacked power to implement the relational model
- Today, microcomputers can run sophisticated relational database software

The Relational Model—Basic Structure

- Relational Database Management System (RDBMS)
- Performs same basic functions provided by hierarchical and network DBMS systems, plus other functions
- Most important advantage of the RDBMS is its ability to let the user/designer operate in a human logical environment

The Relational Model—Basic Structure (continued)

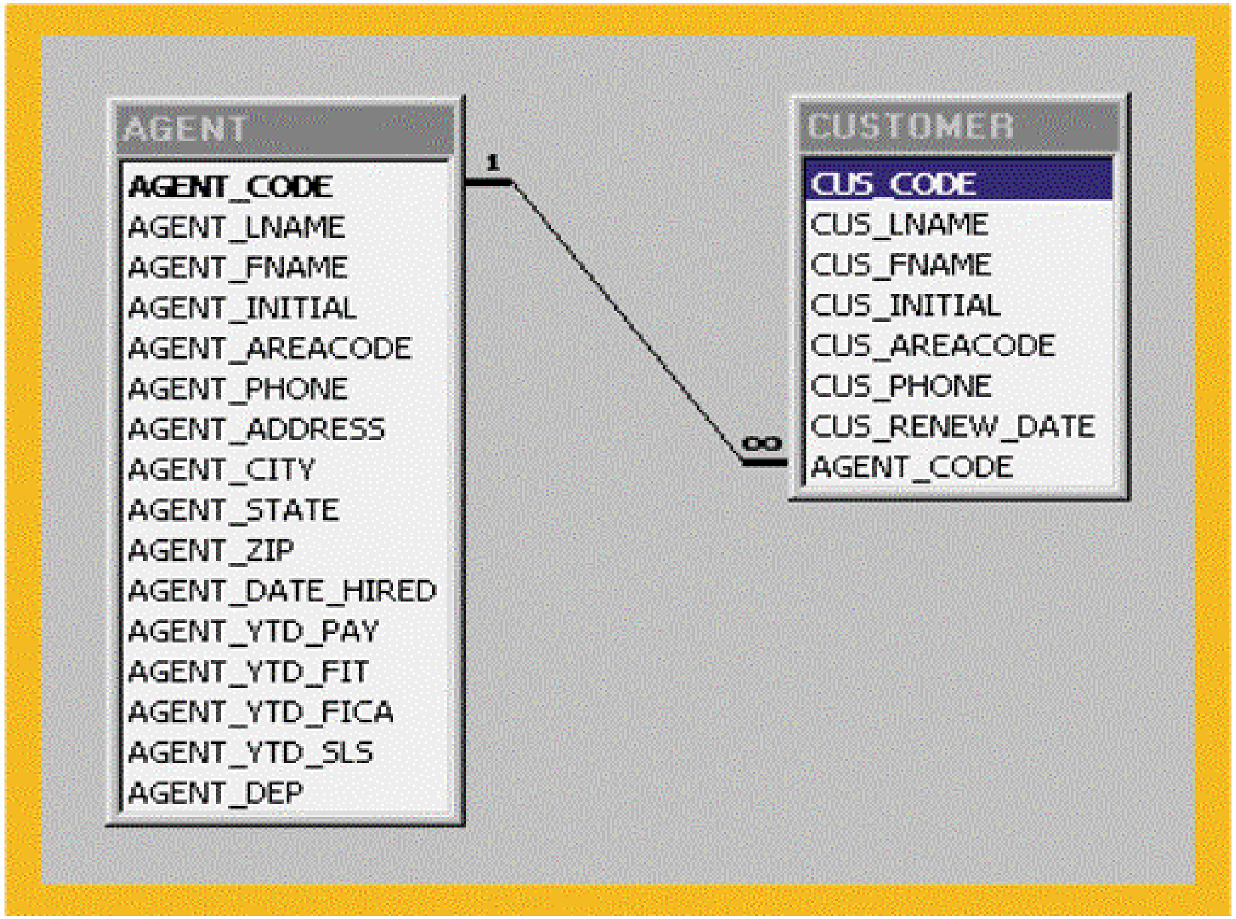
- Table (relations)
  - Matrix consisting of a series of row/column intersections
  - Related to each other by sharing a common entity characteristic
- Relational schema
  - Visual representation of relational database's entities, attributes within those entities, and relationships between those entities

Relational Table

- Stores a collection of related entities
  - Resembles a file
- Relational table is purely logical structure
  - How data are physically stored in the database is of no concern to the user or the designer
  - This property became the source of a real database revolution

A Relational Schema

FIGURE 2.5 A RELATIONAL SCHEMA



Linking Relational Tables

FIGURE 2.4 LINKING RELATIONAL TABLES

Database name: Ch02_InsureCo   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

Link through AGENT_CODE

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_FNAME	CUS_INITIAL	CUS_AREACODE	CUS_PHONE	CUS_RENEW_DATE	AGENT_CODE
10010	Ramas	Alfred	A	615	844-2573	05-Apr-2004	502
10011	Dunne	Leona	K	713	894-1238	16-Jun-2004	501
10012	Smith	Kathy	W	615	894-2285	29-Jan-2005	502
10013	Olowski	Paul	F	615	894-2180	14-Oct-2004	502
10014	Orlando	Myron		615	222-1672	28-Dec-2004	501
10015	O'Brian	Amy	B	713	442-3381	22-Sep-2004	503
10016	Brown	James	G	615	297-1228	25-Mar-2004	502
10017	Williams	George		615	290-2556	17-Jul-2004	503
10018	Farriss	Anne	G	713	382-7185	03-Dec-2004	501
10019	Smith	Olette	K	615	297-3809	14-Mar-2004	503

The Relational Model

- Advantages
  - Structural independence
  - Improved conceptual simplicity
  - Easier database design, implementation, management, and use
  - Ad hoc query capability
  - Powerful database management system

The Relational Model (continued)

- Disadvantages
  - Substantial hardware and system software overhead
  - Can facilitate poor design and implementation
  - May promote “islands of information” problems

The Entity Relationship Model

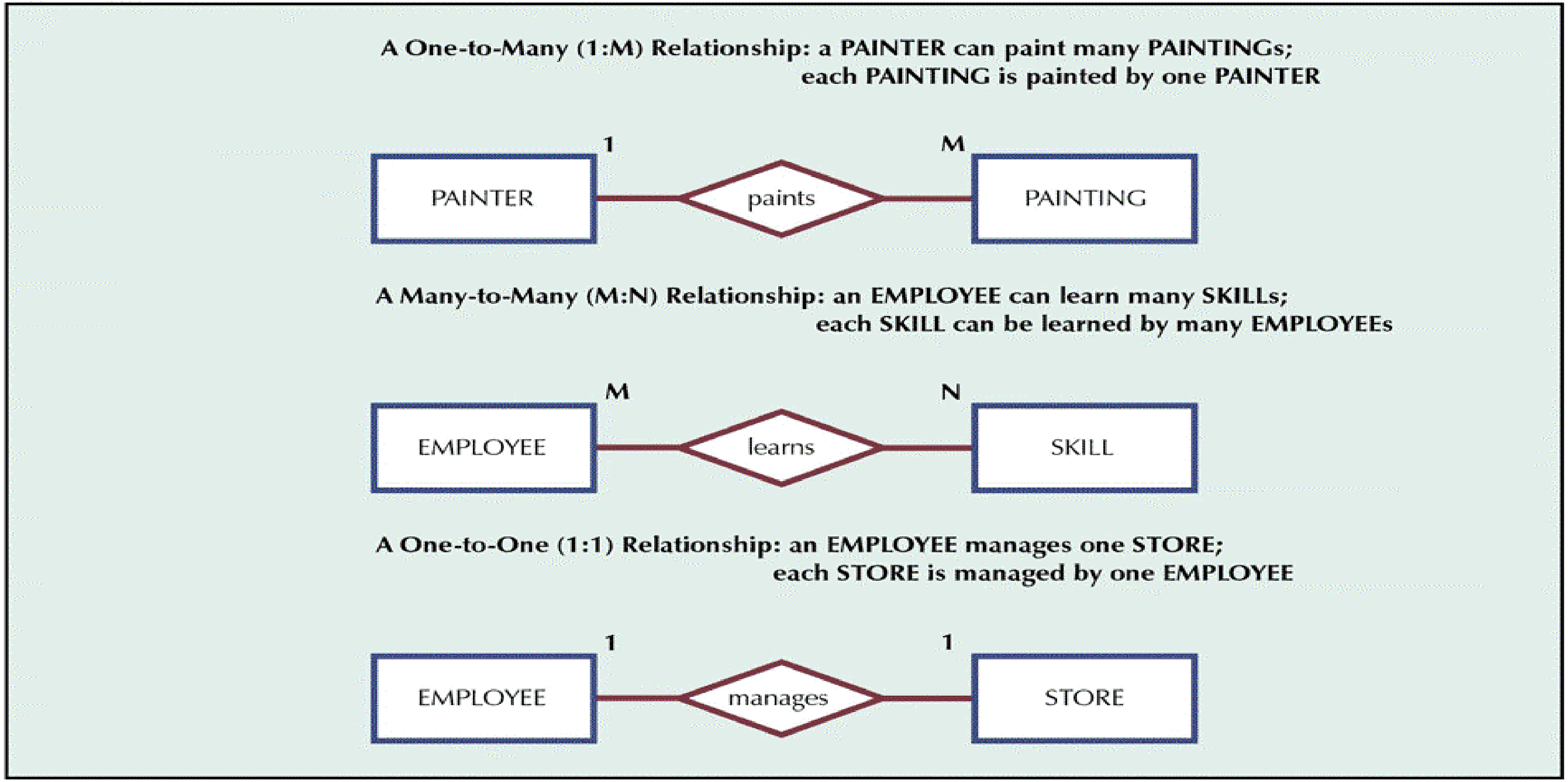
- Widely accepted and adapted graphical tool for data modeling
- Introduced by Chen in 1976
- Graphical representation of entities and their relationships in a database structure

The Entity Relationship Model—Basic Structure

- Entity relationship diagram (ERD)
  - Uses graphic representations to model database components
  - Entity is mapped to a relational table
- Entity instance (or occurrence) is row in table
- Entity set is collection of like entities
- Connectivity labels types of relationships
  - Diamond connected to related entities through a relationship line

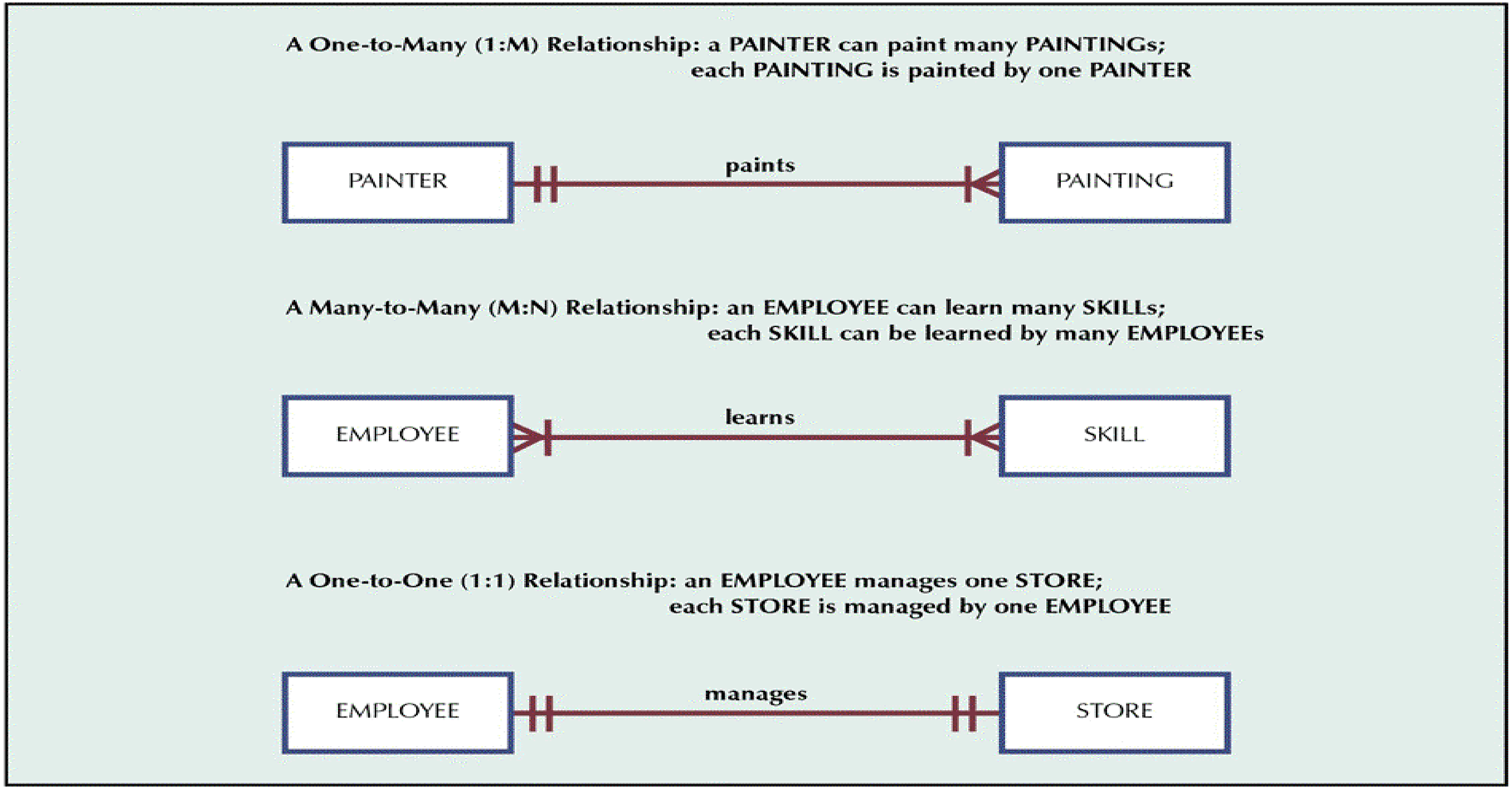
Relationships: The Basic Chen ERD

FIGURE 2.6 RELATIONSHIPS: THE BASIC CHEN ERD



Relationships:
The Basic Crow's Foot ERD

FIGURE 2.7 Relationships: The Basic Crow's Foot ERD



The Entity Relationship Model

- Advantages
  - Exceptional conceptual simplicity
  - Visual representation
  - Effective communication tool
  - Integrated with the relational data model

The Entity Relationship Model (continued)

- Disadvantages
  - Limited constraint representation
  - Limited relationship representation
  - No data manipulation language
  - Loss of information content

The Object Oriented Model

- Semantic data model (SDM) developed by Hammer and McLeod in 1981
- Modeled both data and their relationships in a single structure known as an object
- Basis of object oriented data model (OODM)
- OODM becomes the basis for the object oriented database management system (OODBMS)

The Object Oriented Model (continued)

- Object is described by its factual content
  - Like relational model's entity
- Includes information about relationships between facts within object and relationships with other objects
  - Unlike relational model's entity
- Subsequent OODM development allowed an object to also contain operations
- Object becomes basic building block for autonomous structures

Developments that Boosted OODM's Popularity

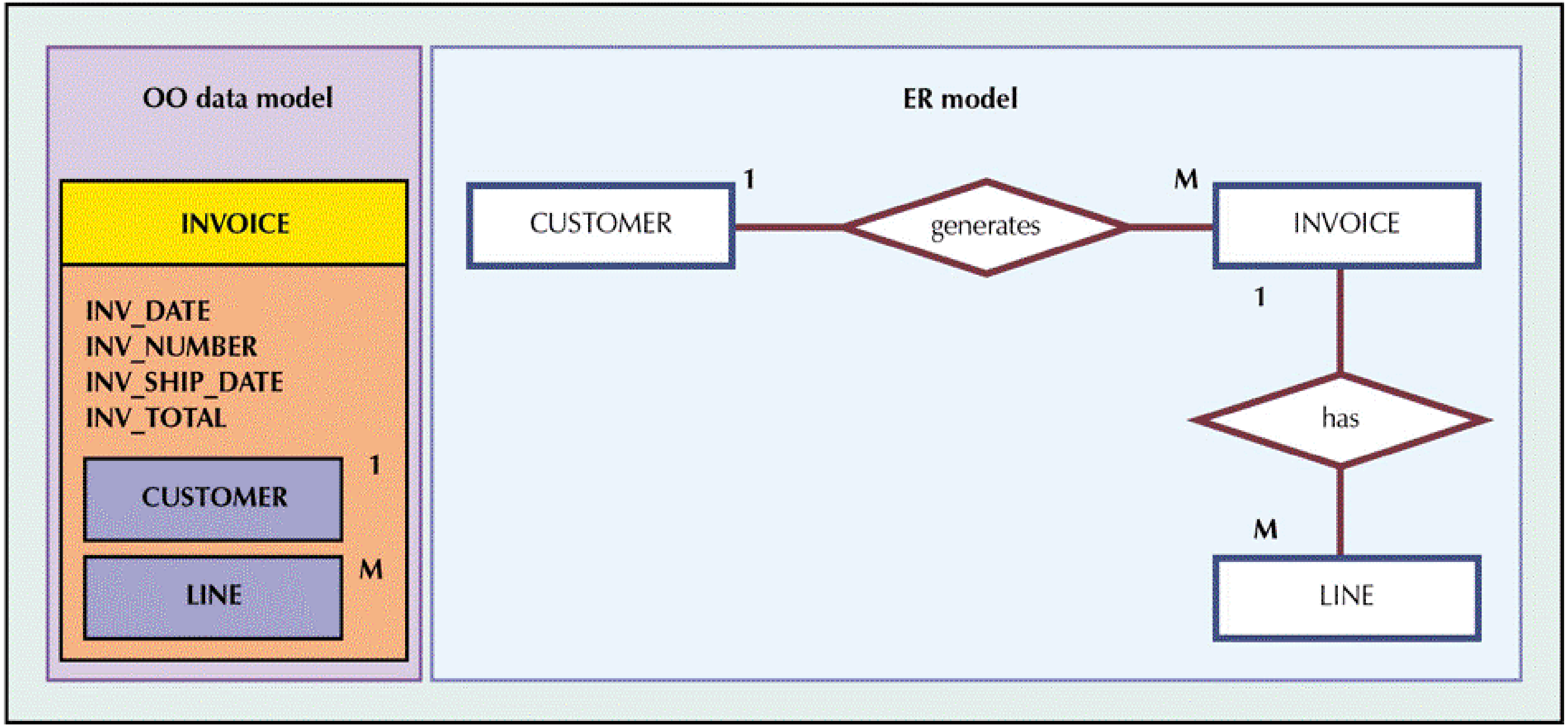
- Growing costs put a premium on code reusability
- Complex data types and system requirements became difficult to manage with a traditional RDBMS
- Became possible to support increasingly sophisticated transaction & information requirements
- Ever-increasing computing power made it possible to support the large computing overhead required

Object Oriented Data Model—Basic Structure

- Object: abstraction of a real-world entity
- Attributes describe the properties of an object
- Objects that share similar characteristics are grouped in classes
- Classes are organized in a class hierarchy
- Inheritance is the ability of an object within the class hierarchy to inherit the attributes and methods of classes above it

A Comparison of the OO Model and the ER Model

FIGURE 2.8 A Comparison of the OO Model and the ER Model



The Object Oriented Model

- Advantages
  - Adds semantic content
  - Visual presentation includes semantic content
  - Database integrity
  - Both structural and data independence

The Object Oriented Model (continued)

- Disadvantages
  - Slow pace of OODM standards development
  - Complex navigational data access
  - Steep learning curve
  - High system overhead slows transactions
  - Lack of market penetration

Other Models

- Extended Relational Data Model (ERDM)
  - Semantic data model developed in response to increasing complexity of applications
  - DBMS based on the ERDM often described as an object/relational database management system (O/RDBMS)
  - Primarily geared to business applications

Other Models (continued)

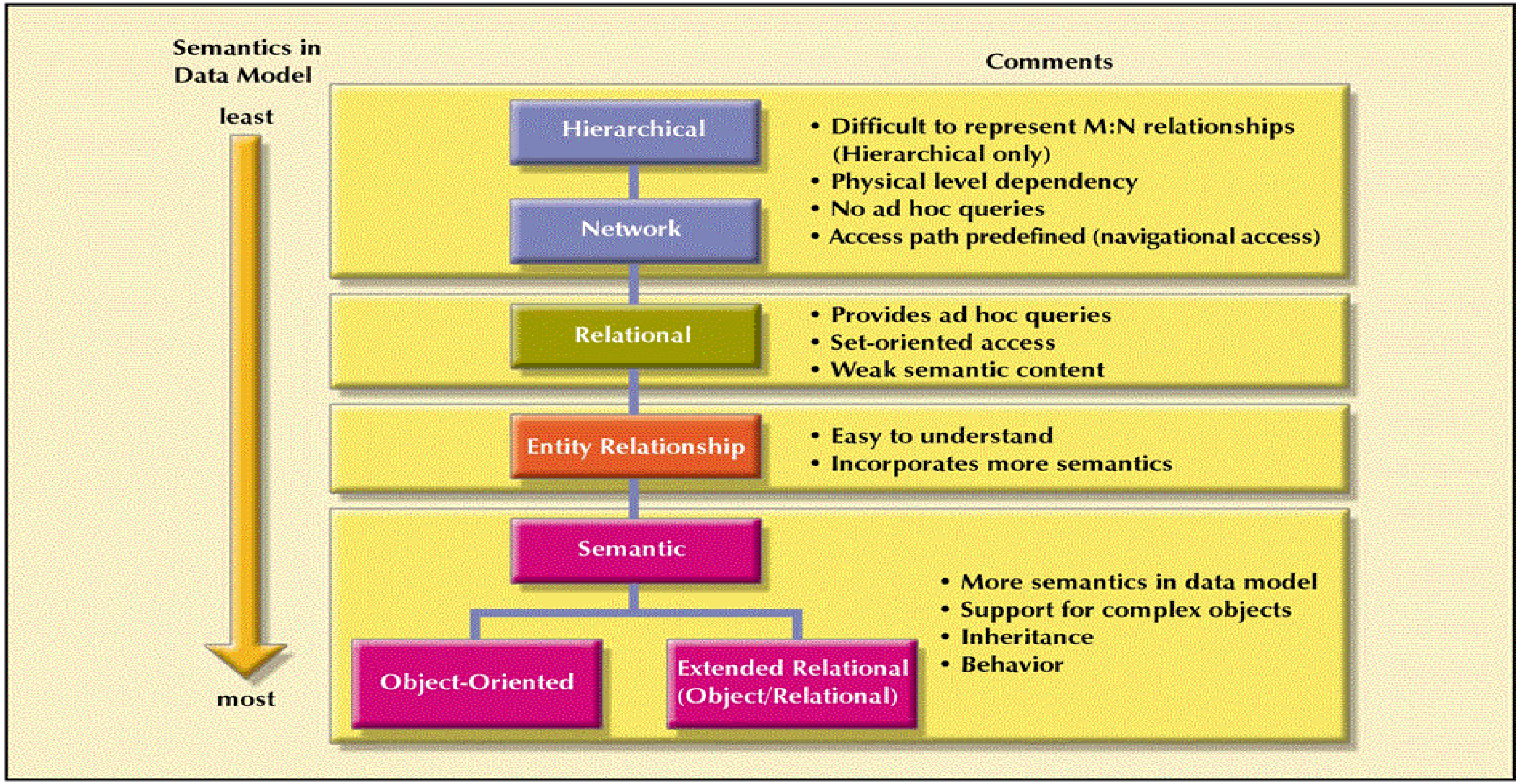
- Date's objections to ERDM label
  - Given proper support for domains, relational data models are quite capable of handling complex data
    - Therefore, capability that is supposedly being extended is already there
  - O/RDM label is not accurate because the relational data model's domain is not an object model structure

Data Models: A Summary

- Each new data model capitalized on the shortcomings of previous models
- Common characteristics:
  - Conceptual simplicity without compromising the semantic completeness of the database
  - Represent the real world as closely as possible
  - Representation of real-world transformations (behavior) must be in compliance with consistency and integrity characteristics of any data model

The Development of Data Models

FIGURE 2.9 The Development of Data Models



Database Models and the Internet

- Characteristics of successful “Internet age” databases
  - Flexible, efficient, and secure Internet access that is easily used, developed, and supported
  - Support for complex data types and relationships
  - Seamless interfacing with multiple data sources and structures

Database Models and the Internet (continued)

– Relative conceptual simplicity to make database design and implementation less cumbersome

– An abundance of available database design, implementation, and application development tools

– A powerful DBMS graphical user interface (GUI) to help make the DBA's job easier

Degrees of Data Abstraction

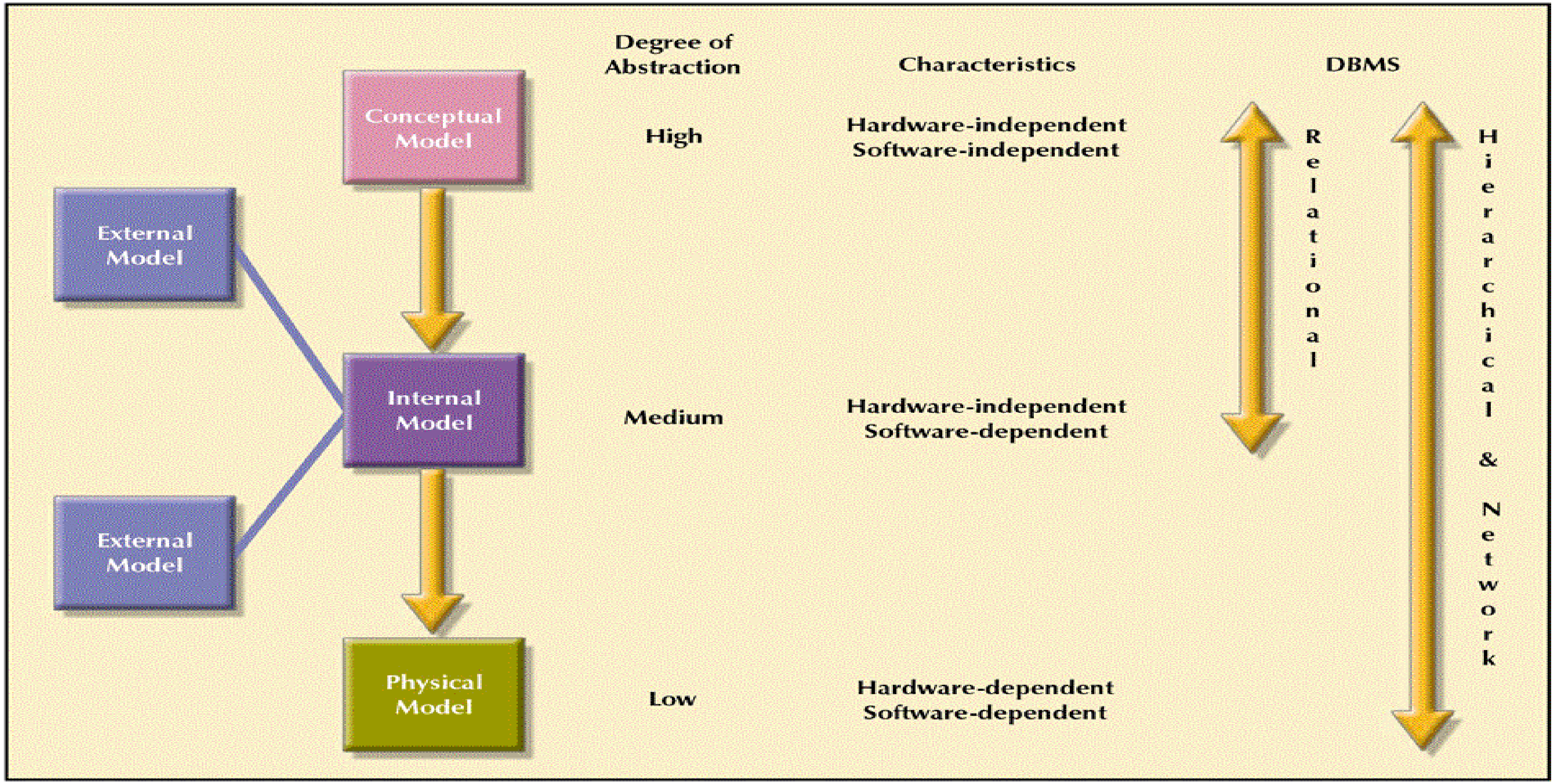
- Way of classifying data models
- Many processes begin at high level of abstraction and proceed to an ever-increasing level of detail
- Designing a usable database follows the same basic process

Degrees of Data Abstraction (continued)

- American National Standards Institute/Standards Planning and Requirements Committee (ANSI/SPARC)
  - Classified data models according to their degree of abstraction (1970s):
    - Conceptual
    - External
    - Internal

Data Abstraction Levels

FIGURE 2.10 DATA ABSTRACTION LEVELS

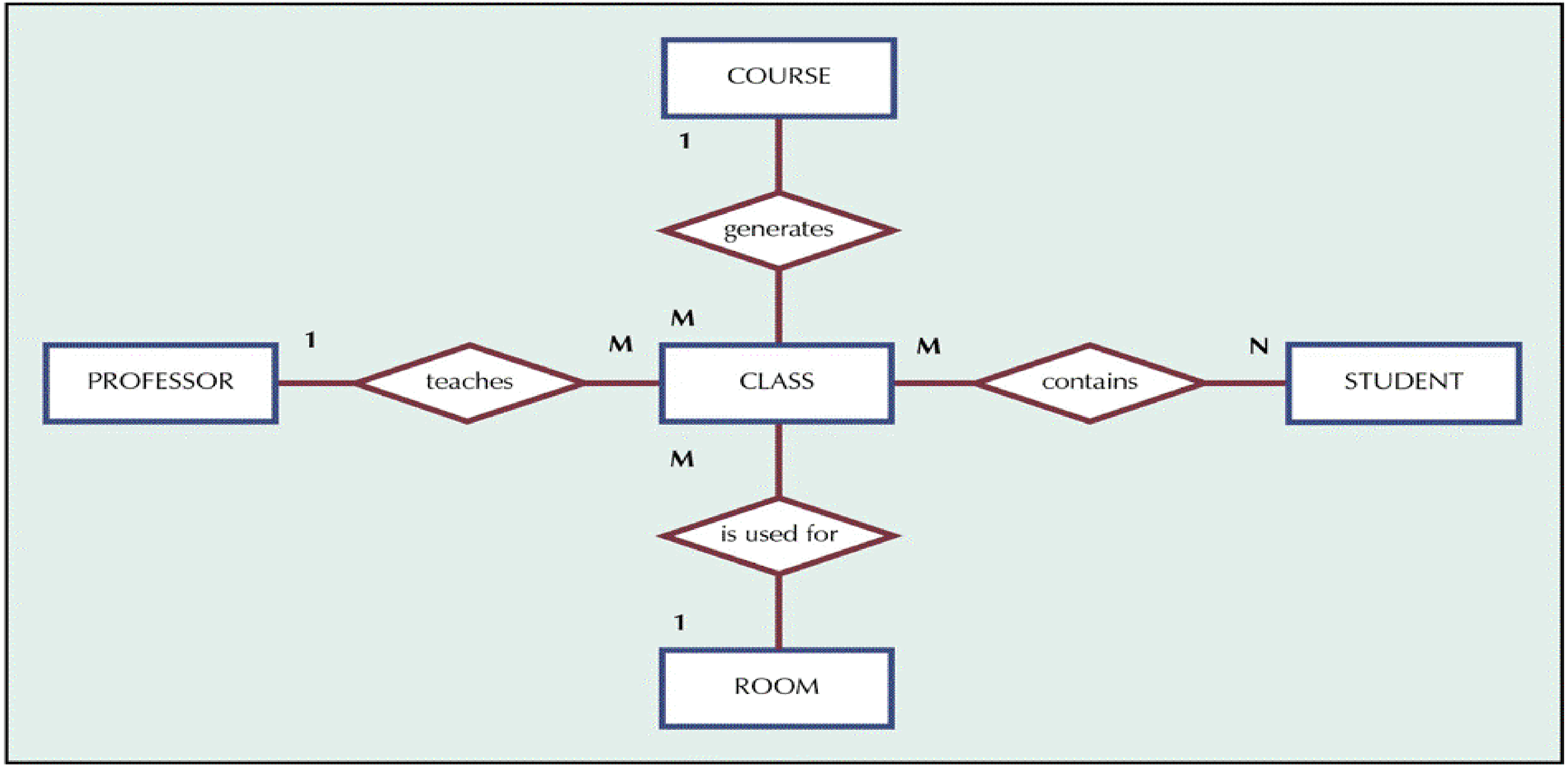


The Conceptual Model

- Represents global view of the database
- Enterprise-wide representation of data as viewed by high-level managers
- Basis for identification and description of main data objects, avoiding details
- Most widely used conceptual model is the entity relationship (ER) model

A Conceptual Model for Tiny College

FIGURE 2.12 A Conceptual Model for Tiny College



Advantages of Conceptual Model

- Provides a relatively easily understood macro level view of data environment
- Independent of both software and hardware
  - Does not depend on the DBMS software used to implement the model
  - Does not depend on the hardware used in the implementation of the model
  - Changes in either the hardware or the DBMS software have no effect on the database design at the conceptual level

The Internal Model

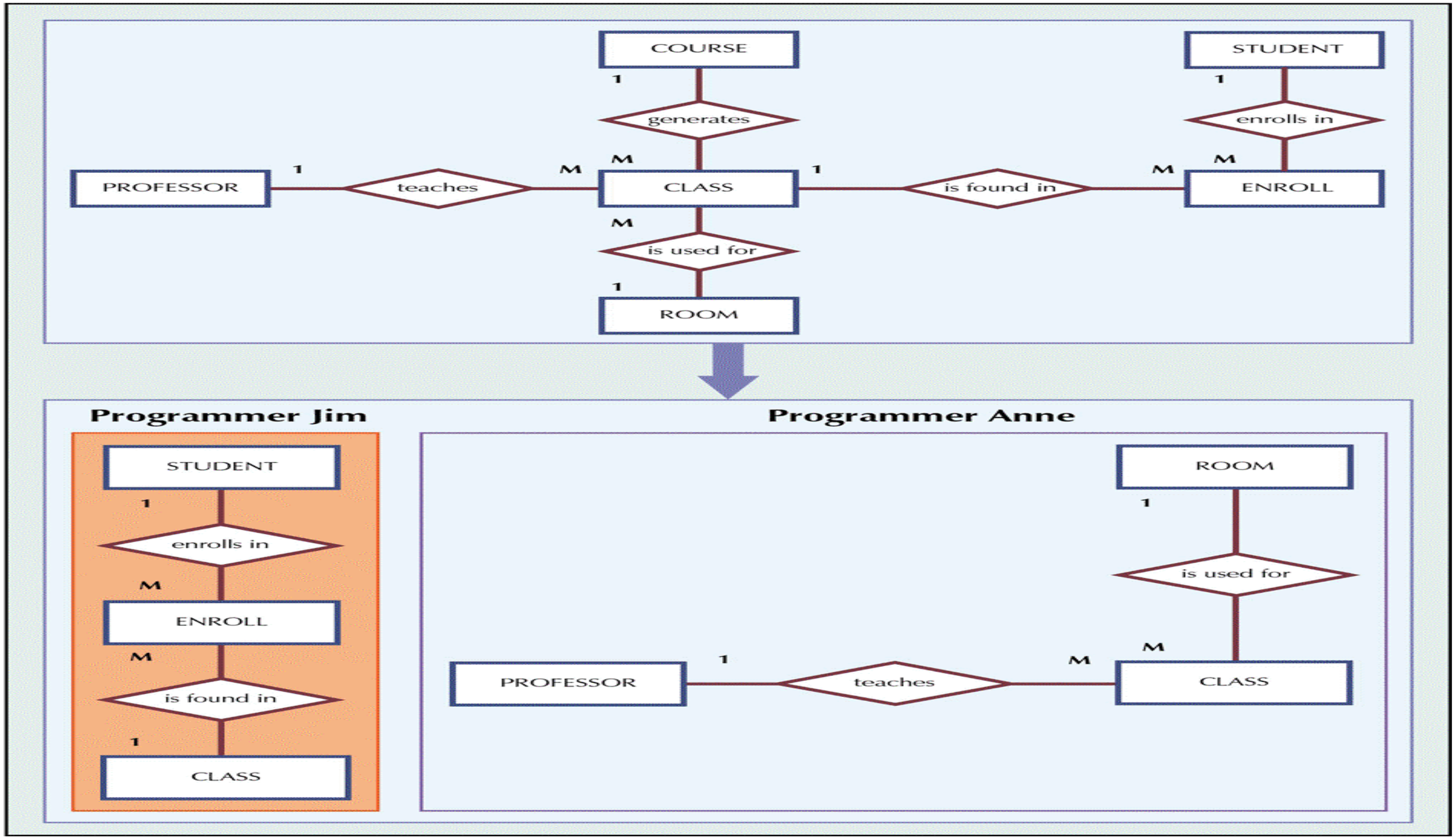
- Representation of the database as “seen” by the DBMS
- Adapts the conceptual model to the DBMS
- Software dependent
- Hardware independent

The External Model

- End users' view of the data environment
- Requires that the modeler subdivide set of requirements and constraints into functional modules that can be examined within the framework of their external models
- Good design should:
  - Consider such relationships between views
  - Provide programmers with a set of restrictions that govern common entities

A Division of an Internal Model into External Models

FIGURE 2.13 A Division of an Internal Model into External Models



Advantages of External Models

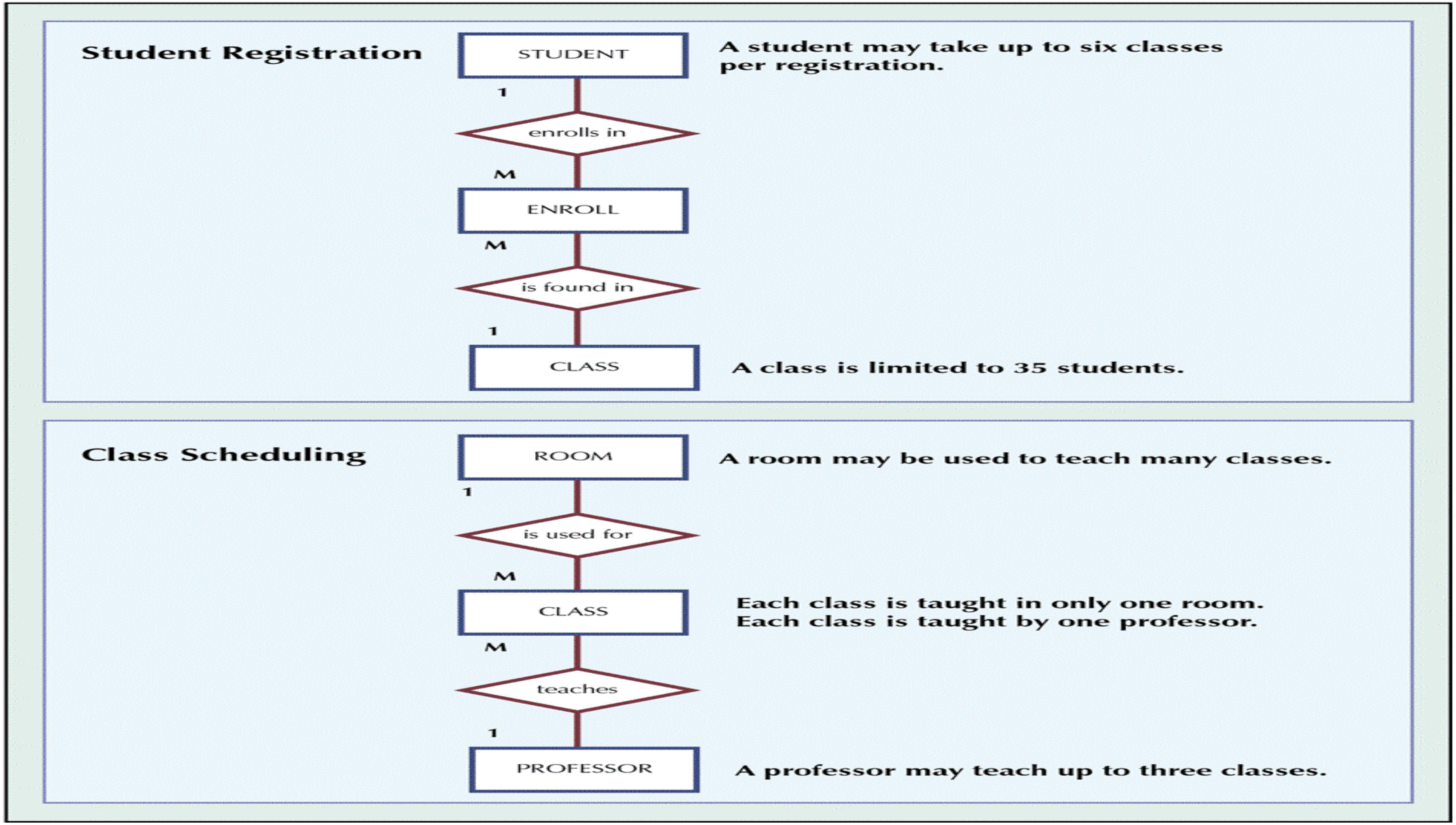
- Use of database subsets makes application program development much simpler
  - Facilitates designer's task by making it easier to identify specific data required to support each business unit's operations
  - Provides feedback about the conceptual model's adequacy
- Creation of external models helps to ensure security constraints in the database design

The External Model

- DBMS dependent
- Hardware independent

The External Models for Tiny College

FIGURE 2.14 The External Models for Tiny College



The Physical Model

- Operates at lowest level of abstraction, describing the way data are saved on storage media such as disks or tapes
- Software and hardware dependent
- Requires that database designers have a detailed knowledge of the hardware and software used to implement database design

Levels of Data Abstraction

TABLE 2.2 LEVELS OF DATA ABSTRACTION

MODEL	DEGREE OF ABSTRACTION	DATA MODEL	FOCUS	INDEPENDENT OF
Conceptual	High	Entity	Global	Hardware and software
External		ER components	Subset	Hardware
Internal		Relational and others	Global	Hardware
Physical	Low	Physical storage methods	N/A	Neither hardware nor software

Summary

- A good DBMS will perform poorly with a poorly designed database
- A data model is a (relatively) simple abstraction of a complex real-world data-gathering environment
- Basic data modeling components are:
  - Entities
  - Attributes
  - Relationships

Summary (continued)

- Hierarchical model
  - Based on a tree structure composed of a root segment, parent segments, and child segments
  - Depicts a set of one-to-many (l:M) relationships between a parent and its children
  - Does not include ad hoc querying capability

Summary (continued)

- Network model attempts to deal with many of the hierarchical model's limitations
- Relational model:
  - Current database implementation standard
  - Much simpler than hierarchical or network design
- Object is basic modeling structure of object oriented model
- Data modeling requirements are a function of different data views (global vs. local) and level of data abstraction