(ADVANCED) DATABASE SYSTEMS (DATABASE MANAGEMENTS)

PROF. DR. HASAN HÜSEYİN BALIK

(1ST WEEK)

OUTLINE

- Course Information and Policies
- Course Syllabus
- 1. The Database Environment and Development Process

COURSE INFORMATION

- Instructor: Prof. Dr. Hasan Huseyin BALIK, hasanbalik@gmail.com and www.hasanbalik.com
- Class Homepage:

  http://www.hasanbalik.com/LectureNotes/DataBase/

  Book: Modern Database Management, 12th Edition, 2016
- Grading: Assignment (%40), Attendence (%20) and Final (%40)
- http://www.hasanbalik.com/LectureNotes/DataBase/syllabus.pdf

COURSE SYLLABUS-1

- The Context of Database Management
  - The Database Environment and Development Process
- Database Analysis
  - Modeling Data in the Organization
  - The Enhanced E-R (Entity-Relationship) Model
- Database Design
  - Logical Database Design and the Relational Model
  - Physical Database Design and Performance

COURSE SYLLABUS-2

- Implementation
  - Introduction to SQL
  - Advanced SQL
  - Database Application Development
  - Data Warehousing
- Advanced Database Topics
  - Data Quality and Integration
  - Data and Database Administration

1. OUTLINE

1. The Context of Database Management

1.1 The Database Environment and Development Process

1.1 THE DATABASE ENVIRONMENT AND DEVELOPMENT PROCESS

OBJECTIVES

- Define terms
- Name limitations of conventional file processing
- Explain advantages of databases
- Identify costs and risks of databases
- List components of database environment
- Identify categories of database applications
- Describe database system development life cycle
- Explain prototyping and agile development approaches
- Explain roles of individuals
- Explain the three-schema architecture for databases

DEFINITIONS

- Database: organized collection of logically related data
- Data: stored representations of meaningful objects and events
  - Structured: numbers, text, dates
  - Unstructured: images, video, documents
- Information: data processed to increase knowledge in the person using the data
- Metadata: data that describes the properties and context of user data

Data in context

Class Roster

Course: MGT 500 Business Policy

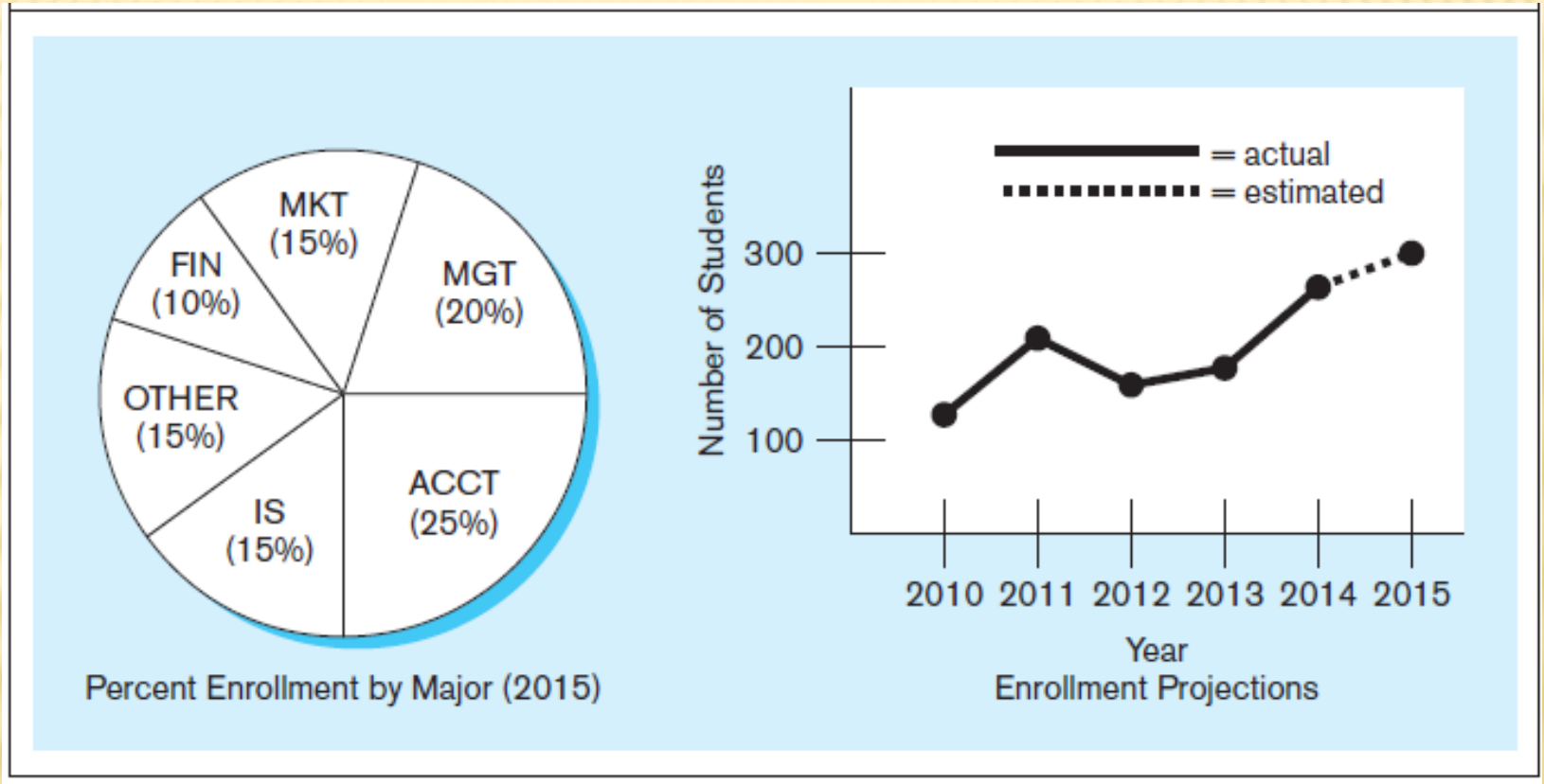
Semester: Spring 2015

Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Context helps users understand data

Summarized data



Graphical displays turn data into useful information that managers can use for decision making and interpretation

TABLE 1-1 Example Metadata for Class Roster

Data Item			Metadata			
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	Integer	9			Student ID (SSN)	Student IS
Major	Alphanumeric	4			Student major	Student IS
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

Descriptions of the properties or characteristics of the data, including data types, field sizes, allowable values, and data context

DISADVANTAGES OF FILE PROCESSING

- **Program-Data Dependence**
  - All programs maintain metadata for each file they use
- **Duplication of Data**
  - Different systems/programs have separate copies of the same data
- **Limited Data Sharing**
  - No centralized control of data
- **Lengthy Development Times**
  - Programmers must design their own file formats
- **Excessive Program Maintenance**
  - 80% of information systems budget

PROBLEMS WITH DATA DEPENDENCY

- Each application programmer must maintain his/her own data
- Each application program needs to include code for the metadata of each file
- Each application program must have its own processing routines for reading, inserting, updating, and deleting data
- Lack of coordination and central control
- Non-standard file formats

Duplicate Data

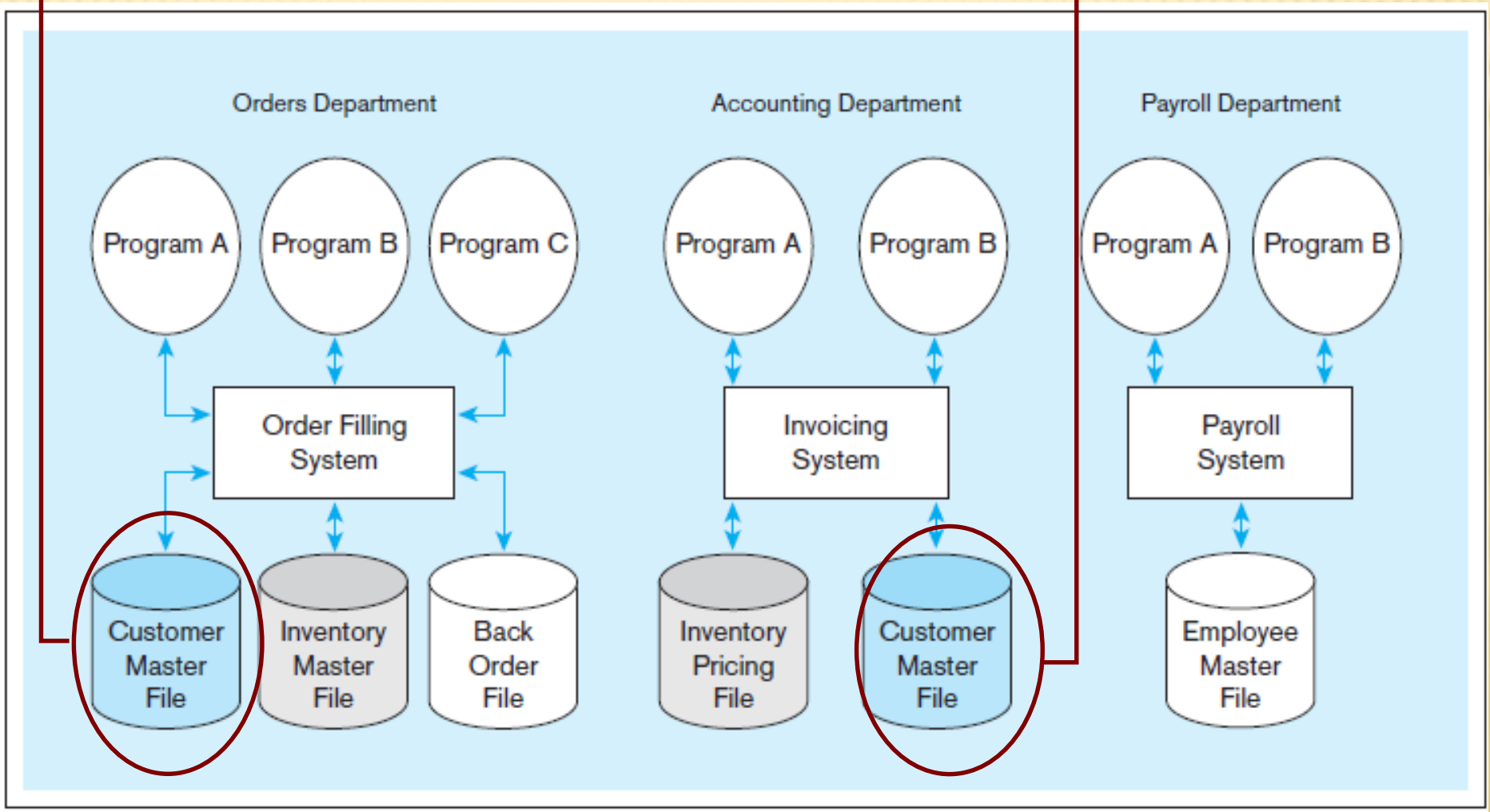


FIGURE 1-2 **Old file processing systems at Pine Valley Furniture Company**

PROBLEMS WITH DATA REDUNDANCY

- Waste of space to have duplicate data
- Causes more maintenance headaches
- The biggest problem:
  - **Data changes in one file could cause inconsistencies**
  - Compromises in ***data integrity***
- "data integrity" refers to ensuring the validity, security, and availability of a company's data

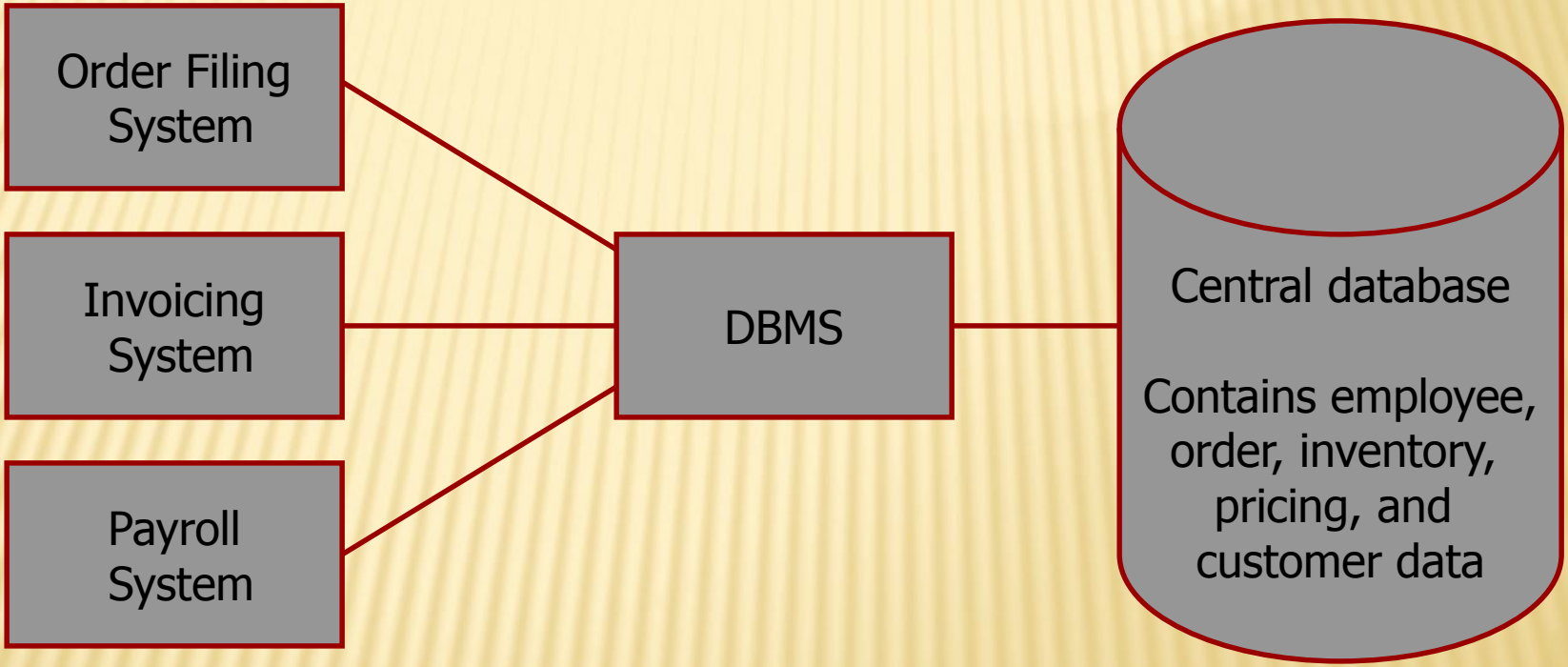
SOLUTION: THE DATABASE APPROACH

- Central repository of shared data
- Data is managed by a controlling agent
- Stored in a standardized, convenient form

Requires a Database Management System (DBMS)

DATABASE MANAGEMENT SYSTEM

- A software system that is used to create, maintain, and provide controlled access to user databases



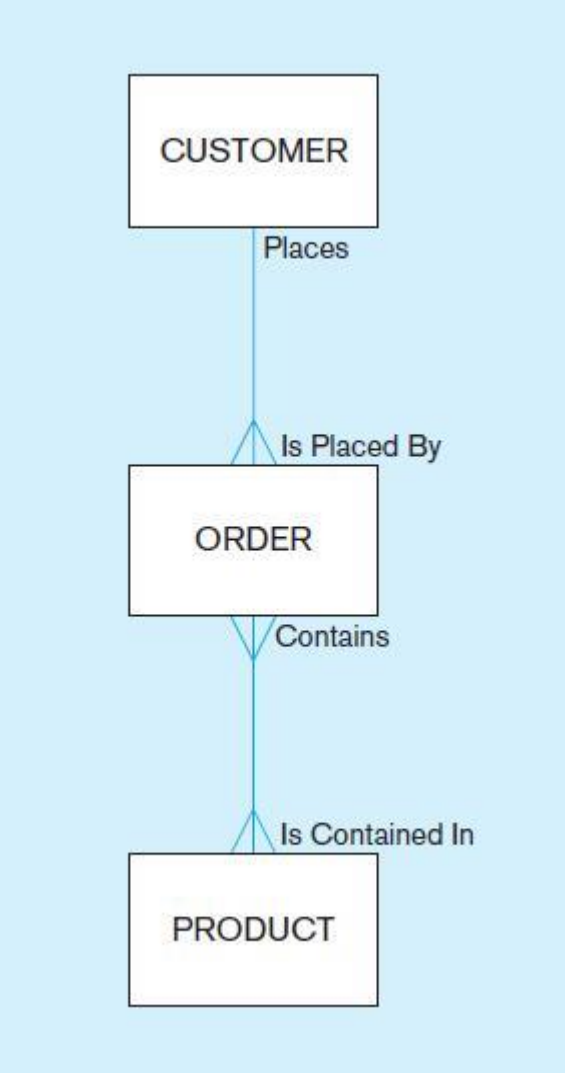
DBMS manages data resources like an operating system manages hardware resources

ELEMENTS OF THE DATABASE APPROACH

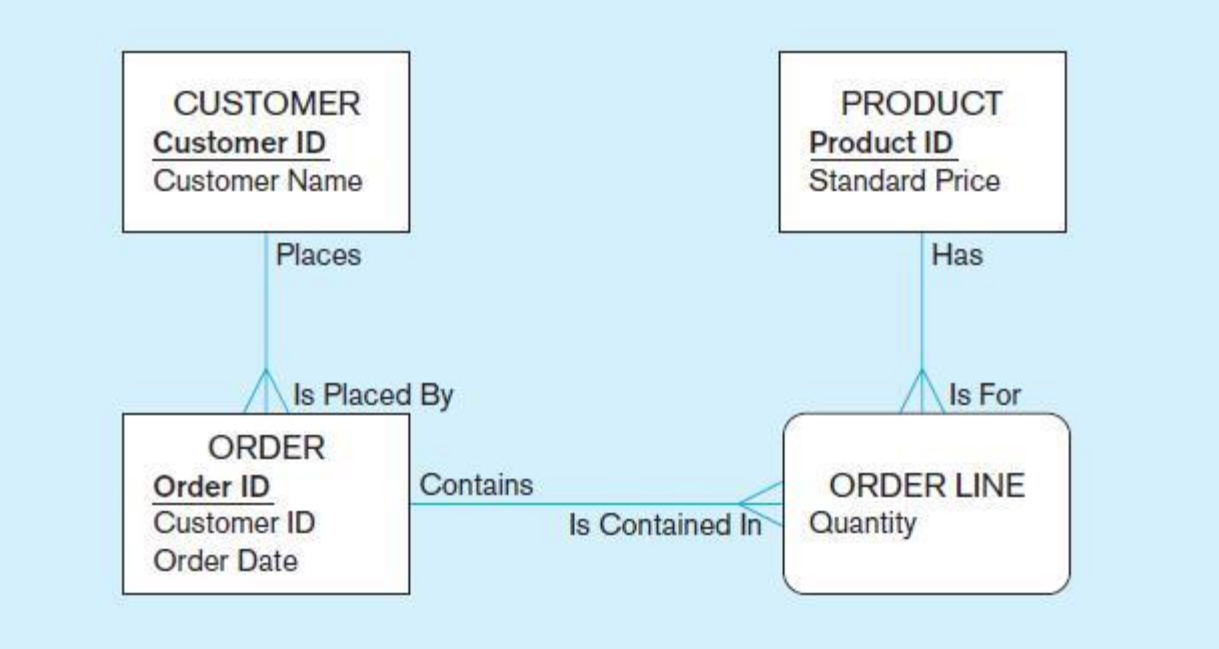
- Data models
  - Graphical diagram capturing nature and relationship of data
  - Enterprise Data Model–high-level entities and relationships for the organization
  - Project Data Model–more detailed view, matching data structure in database or data warehouse
- Entities
  - Noun form describing a person, place, object, event, or concept
  - Composed of attributes
- Relationships
  - Between entities
  - Usually one-to-many (1:M) or many-to-many (M:N), but could also be one-to-one (1:1)
- Relational Databases
  - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships

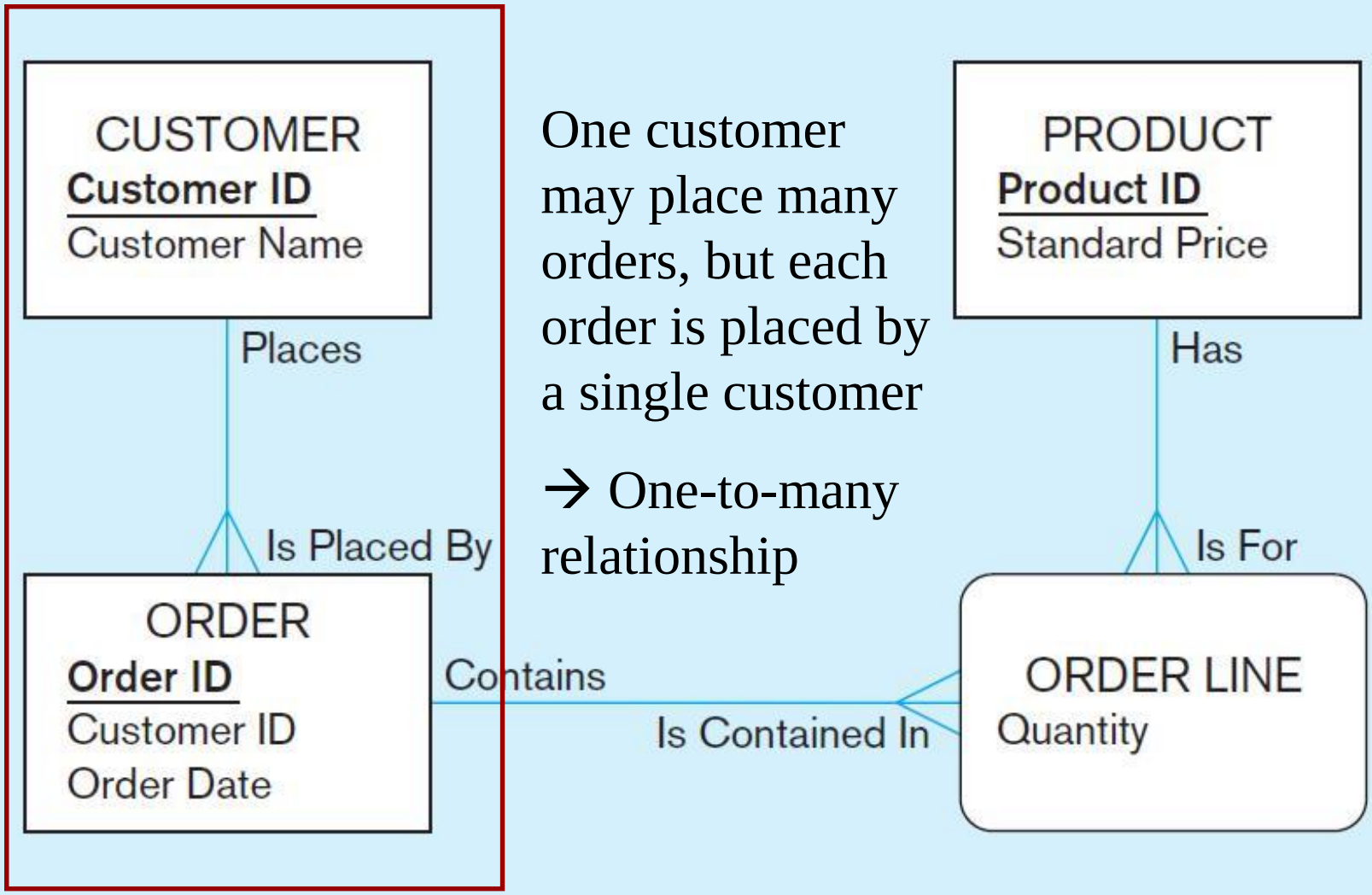
Comparison of enterprise and project level data models

Segment of an enterprise data model



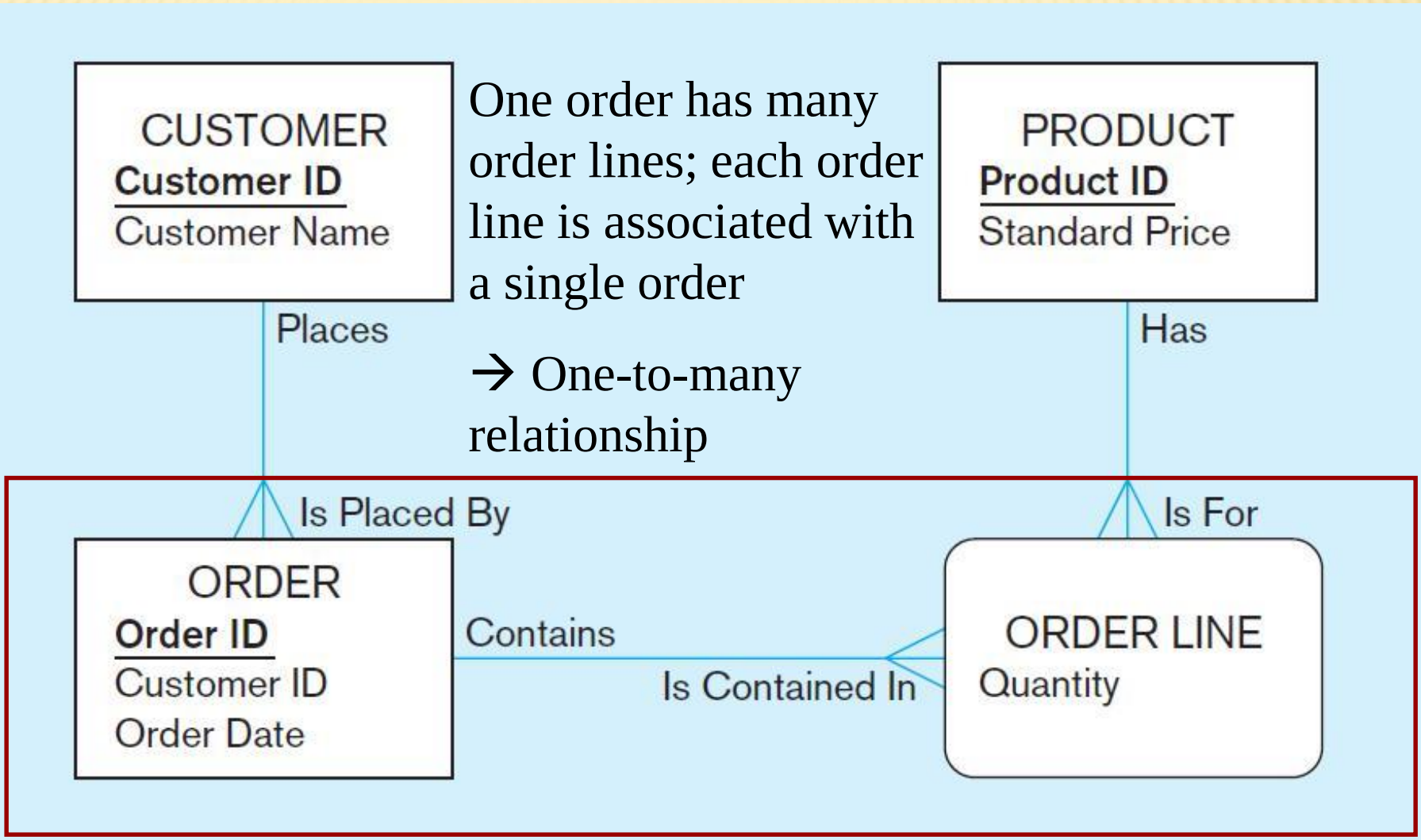
Segment of a project-level data model





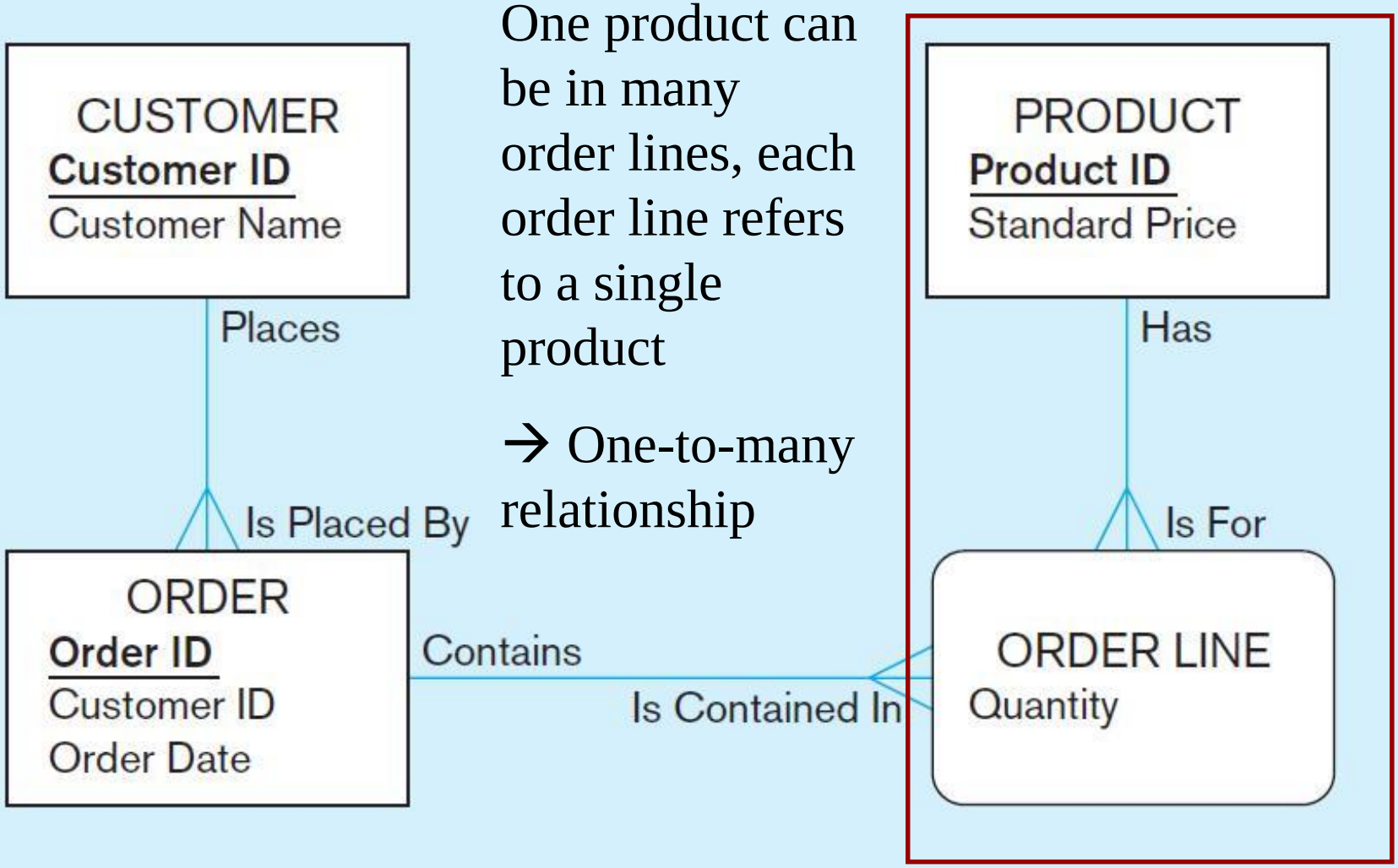
One customer may place many orders, but each order is placed by a single customer

→ One-to-many relationship



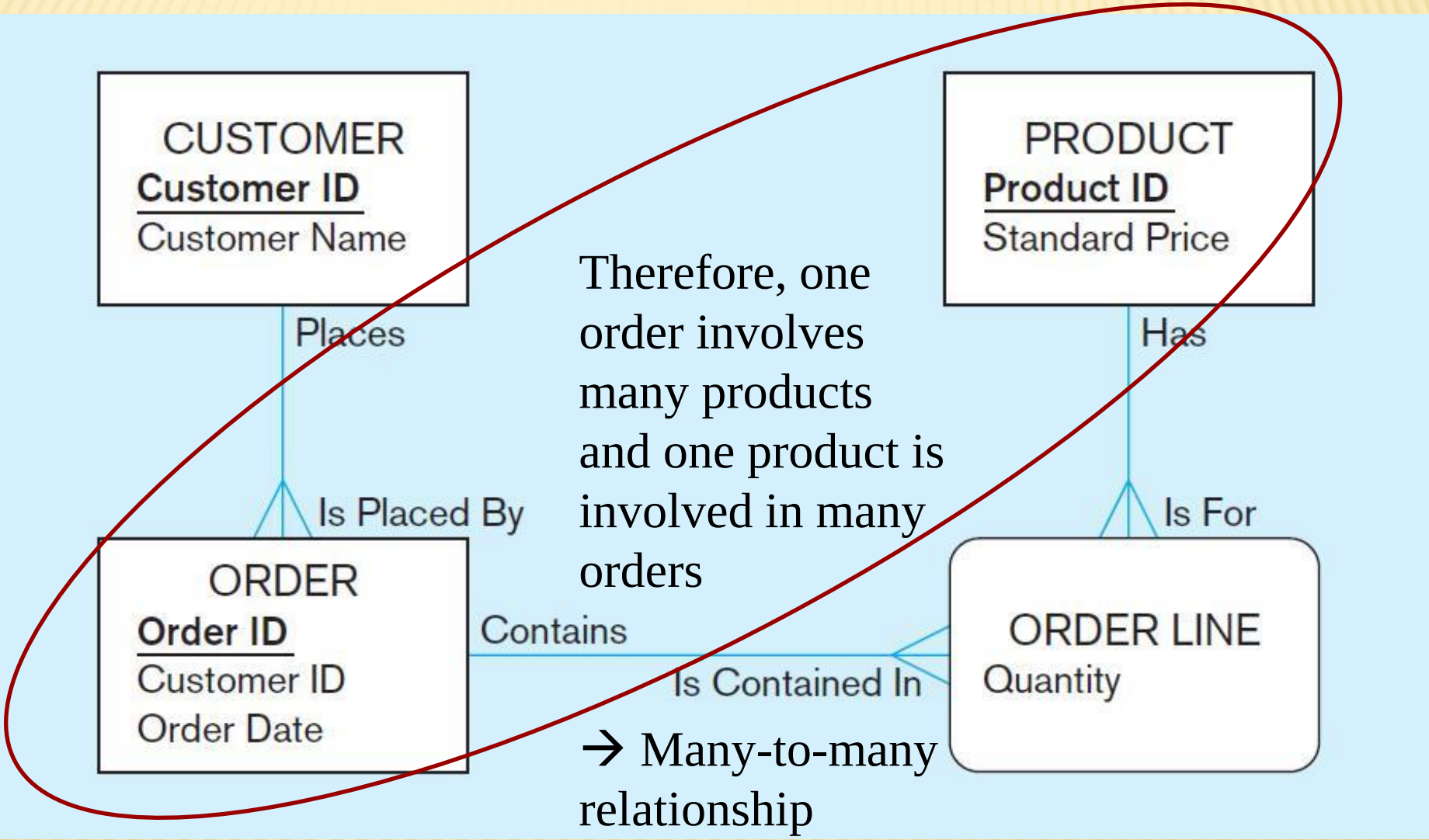
One order has many order lines; each order line is associated with a single order

→ One-to-many relationship



One product can be in many order lines, each order line refers to a single product

→ One-to-many relationship


CUSTOMER
Customer ID
Customer Name
PRODUCT
Product ID
Standard Price
Places
Has
Is Placed By
Is For
ORDER
Order ID
Customer ID
Order Date
Contains
Is Contained In
ORDER LINE
Quantity
Therefore, one order involves many products and one product is involved in many orders
→ Many-to-many relationship

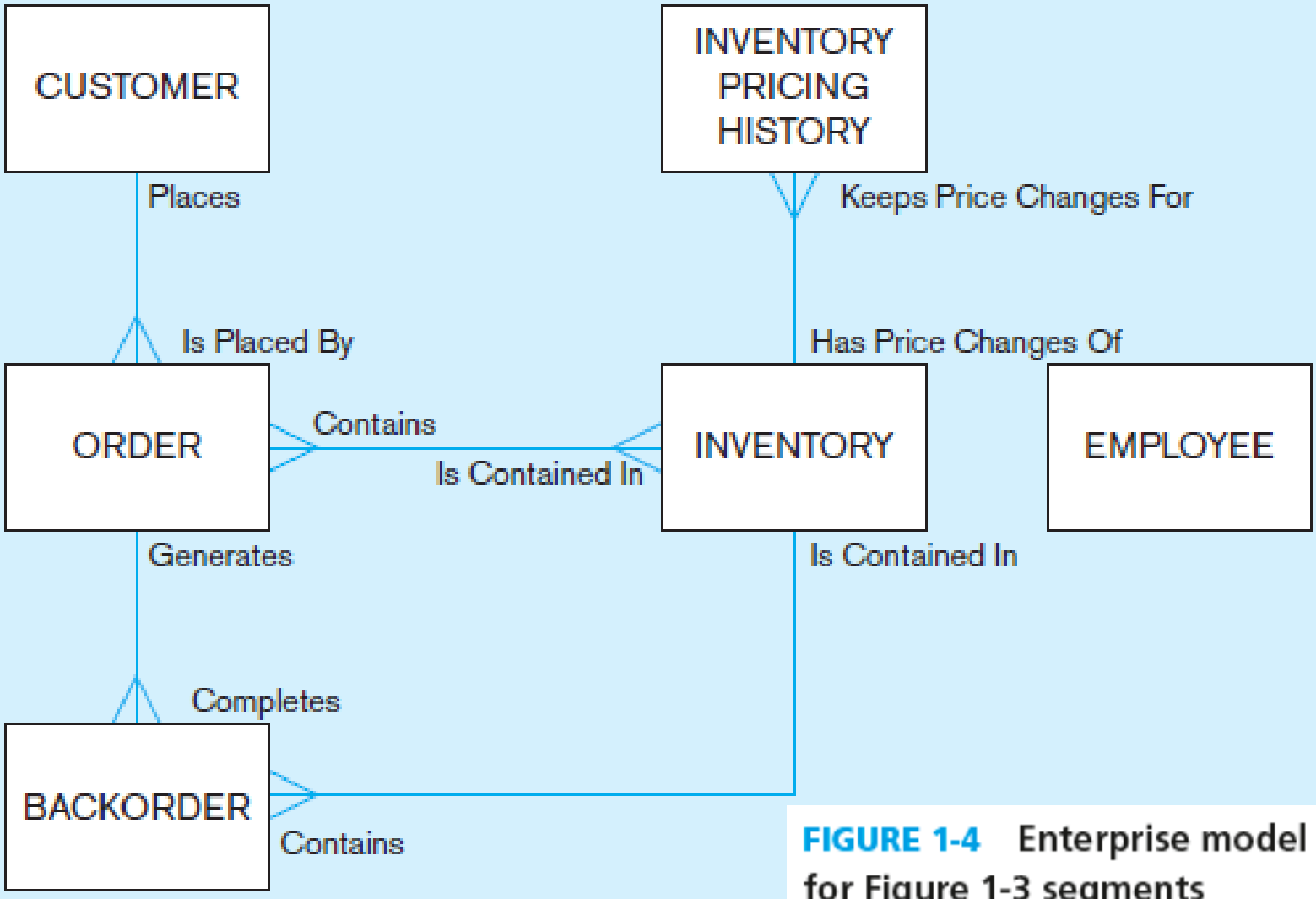


FIGURE 1-4 Enterprise model for Figure 1-3 segments

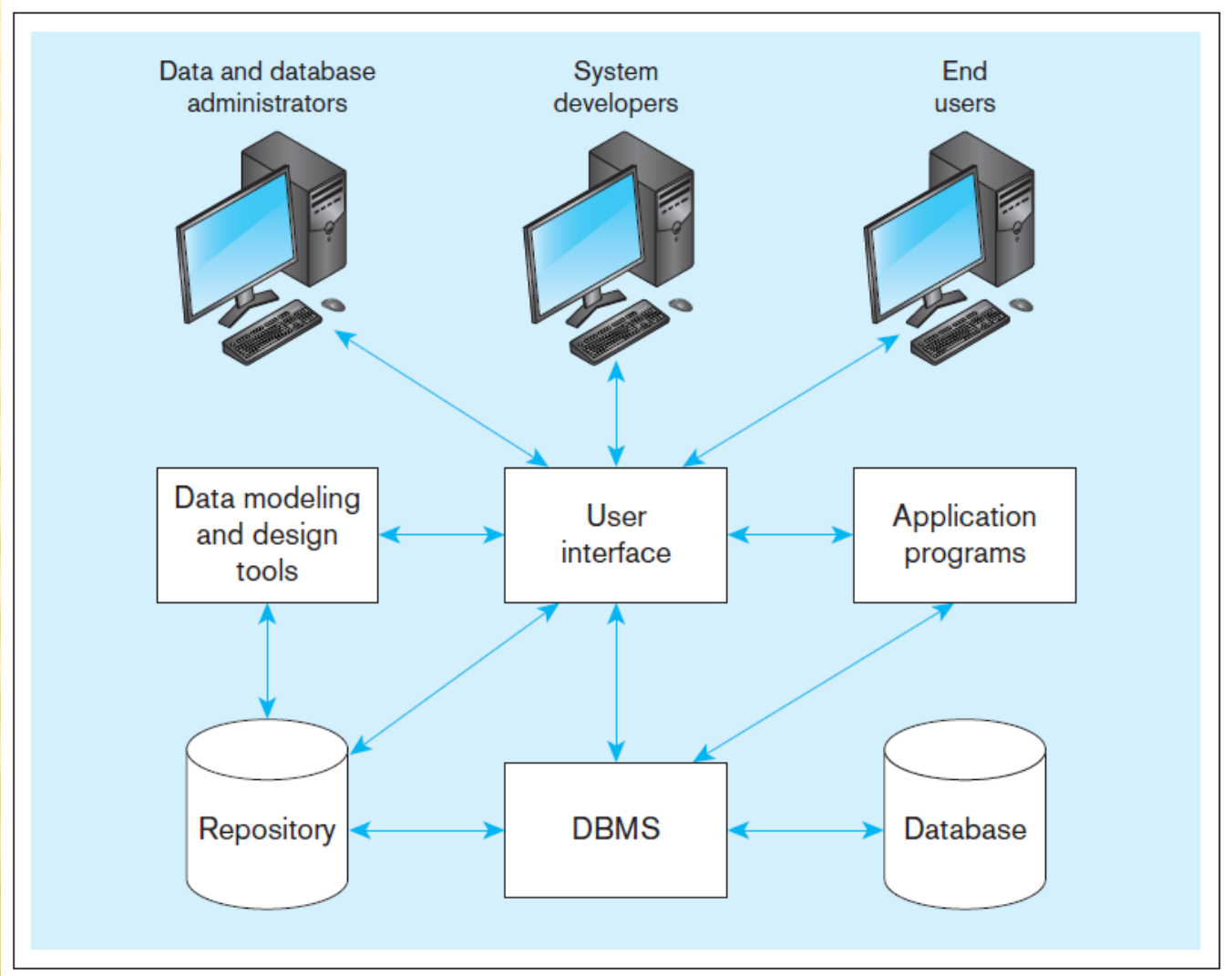
ADVANTAGES OF THE DATABASE APPROACH

- Program-data independence
- Planned data redundancy
- Improved data consistency
- Improved data sharing
- Increased application development productivity
- Enforcement of standards
- Improved data quality
- Improved data accessibility and responsiveness
- Reduced program maintenance
- Improved decision support

COSTS AND RISKS OF THE DATABASE APPROACH

- New, specialized personnel
- Installation and management cost and complexity
- Conversion costs
- Need for explicit backup and recovery
- Organizational conflict

Components of the database environment



COMPONENTS OF THE DATABASE ENVIRONMENT

- **Data modeling and design tools** -- automated tools used to design databases and application programs
- **Repository**–centralized storehouse of metadata
- **Database Management System (DBMS)** –software for managing the database
- **Database**–storehouse of the data
- **Application Programs**–software using the data
- **User Interface**–text, graphical displays, menus, etc. for user
- **Data/Database Administrators**–personnel responsible for maintaining the database
- **System Developers**–personnel responsible for designing databases and software
- **End Users**–people who use the applications and databases

ENTERPRISE DATA MODEL

- First step in the database development process
- Specifies scope and general content
- Overall picture of organizational data at high level of abstraction
- Entity-relationship diagram
- Descriptions of entity types
- Relationships between entities
- Business rules

Example business function-to-data entity matrix

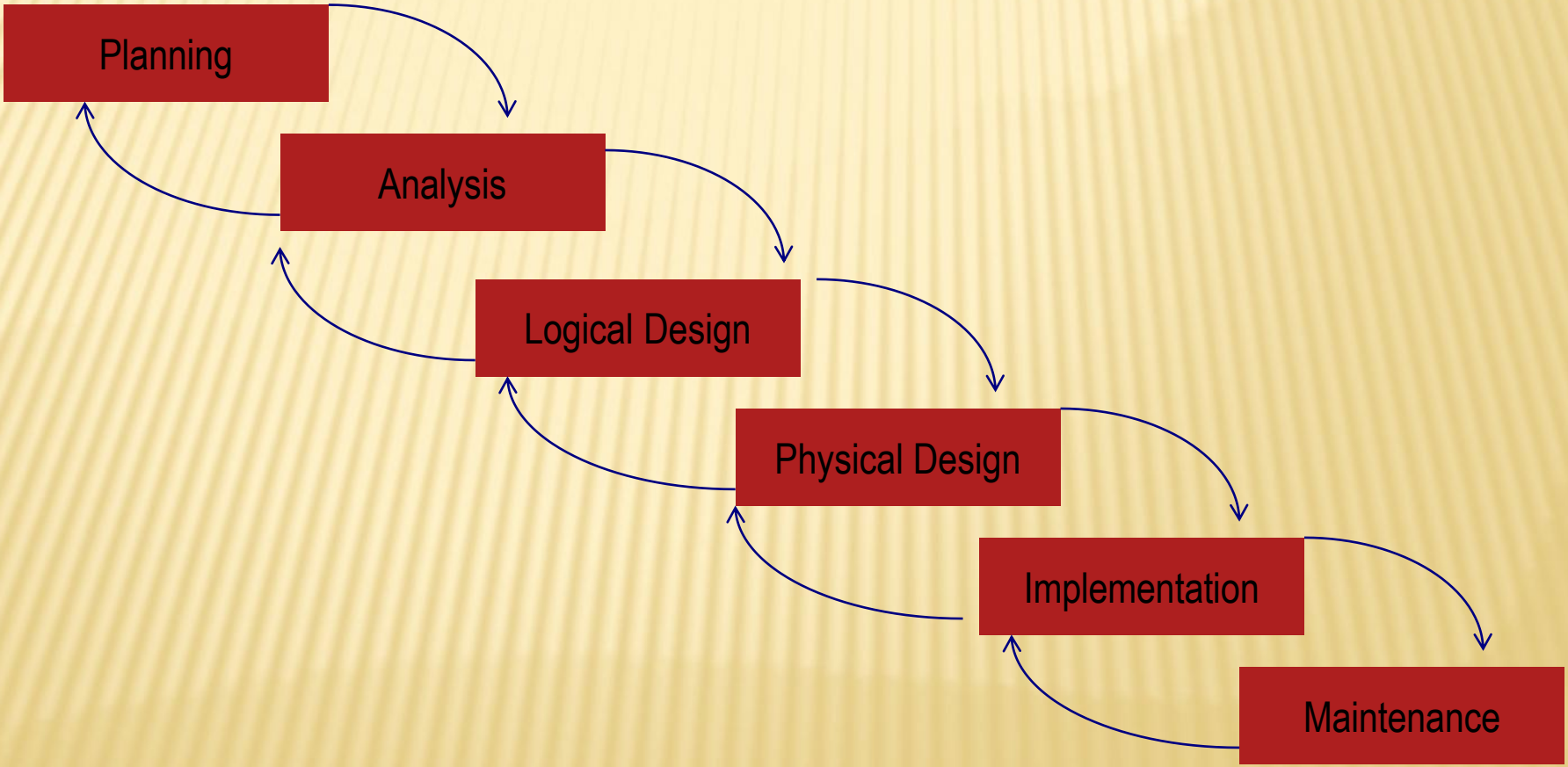
Business Functions \ Data Entity Types	Customer	Product	Raw Material	Order	Work Center	Work Order	Invoice	Equipment	Employee
Business Planning	X	X						X	X
Product Development		X	X		X			X	
Materials Management		X	X	X	X	X		X	
Order Fulfillment	X	X	X	X	X	X	X	X	X
Order Shipment	X	X		X	X		X		X
Sales Summarization	X	X		X			X		X
Production Operations		X	X	X	X	X		X	X
Finance and Accounting	X	X	X	X	X		X	X	X

X = data entity is used within business function

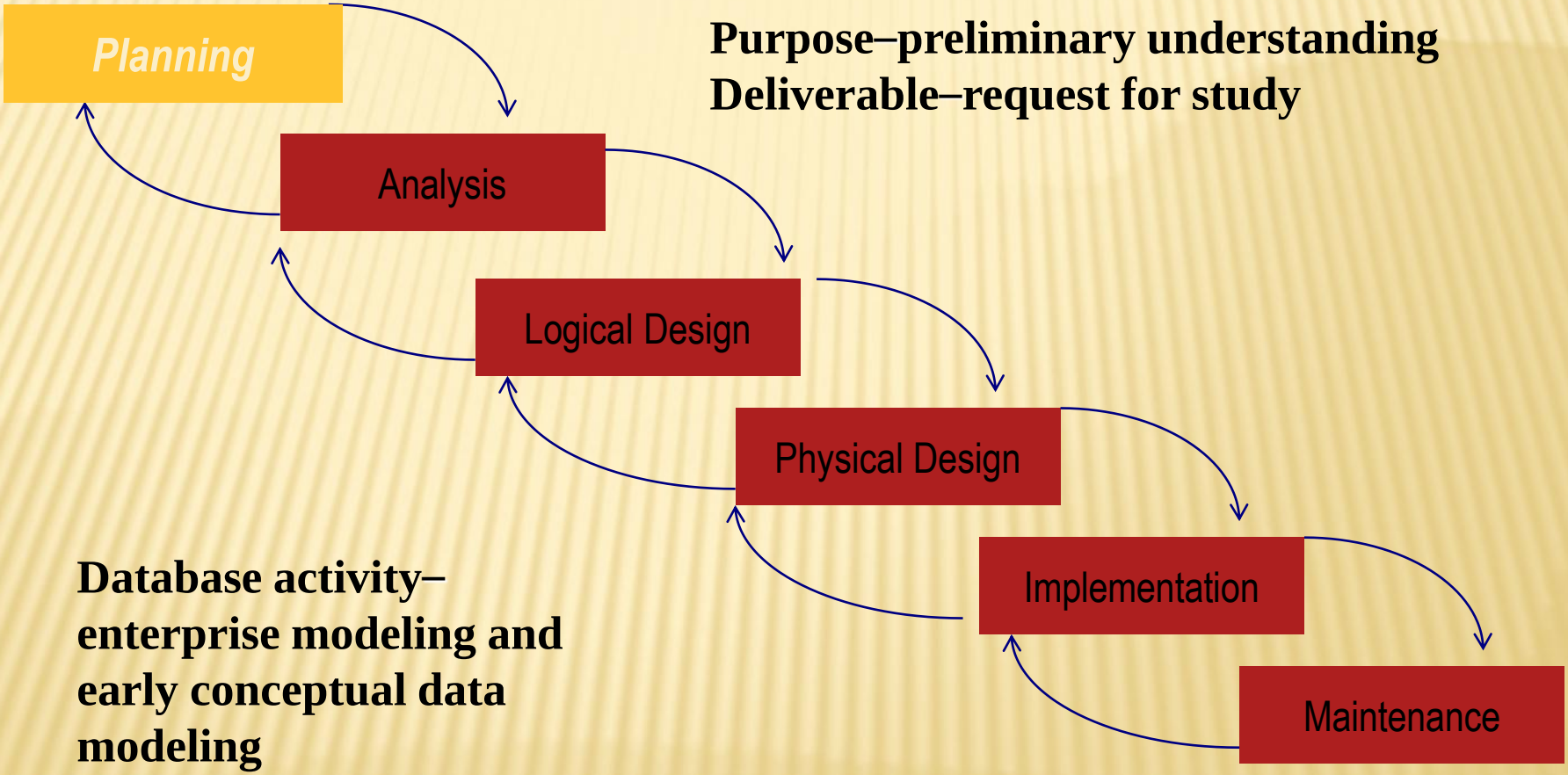
TWO APPROACHES TO DATABASE AND IS DEVELOPMENT

- SDLC
  - System Development Life Cycle
  - Detailed, well-planned development process
  - Time-consuming, but comprehensive
  - Long development cycle
- Prototyping
  - Rapid application development (RAD)
  - Cursory attempt at conceptual data modeling
  - Define database during development of initial prototype
  - Repeat implementation and maintenance activities with new prototype versions

SYSTEMS DEVELOPMENT LIFE CYCLE ("WATERFALL" APPROACH)



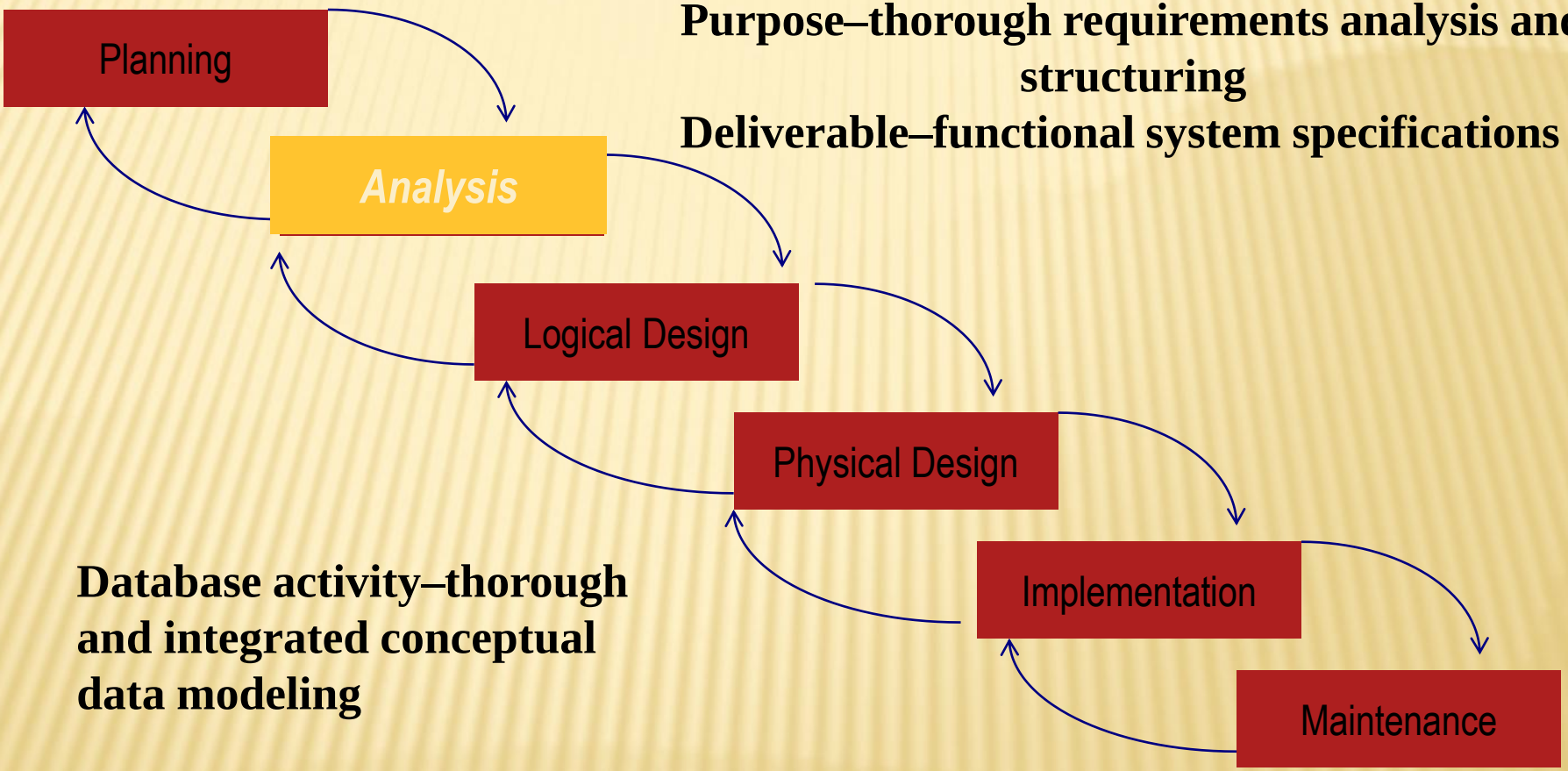
SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)



Purpose–preliminary understanding
Deliverable–request for study

Database activity–enterprise modeling and early conceptual data modeling

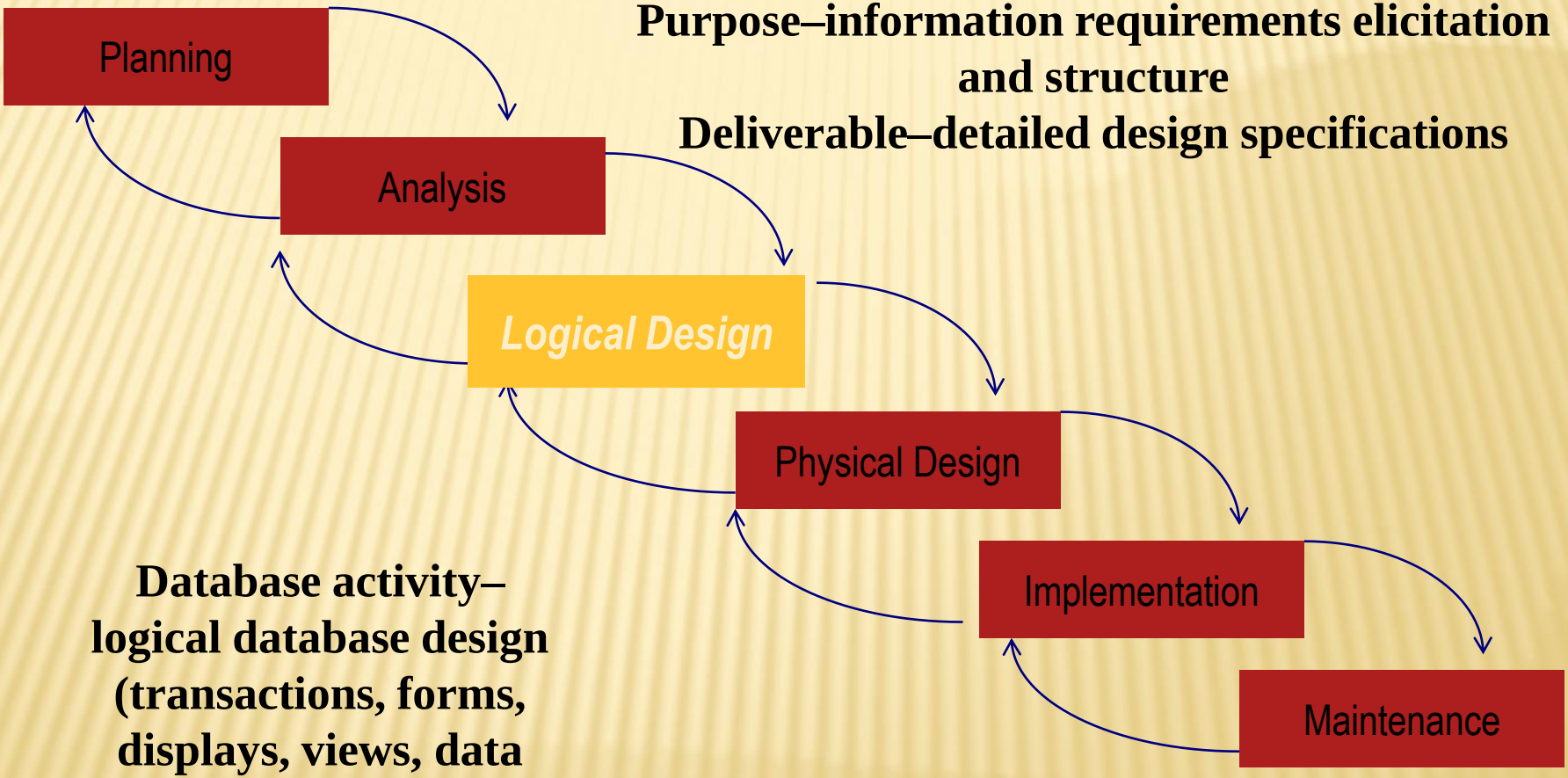
SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)



Purpose–thorough requirements analysis and structuring
Deliverable–functional system specifications

Database activity–thorough and integrated conceptual data modeling

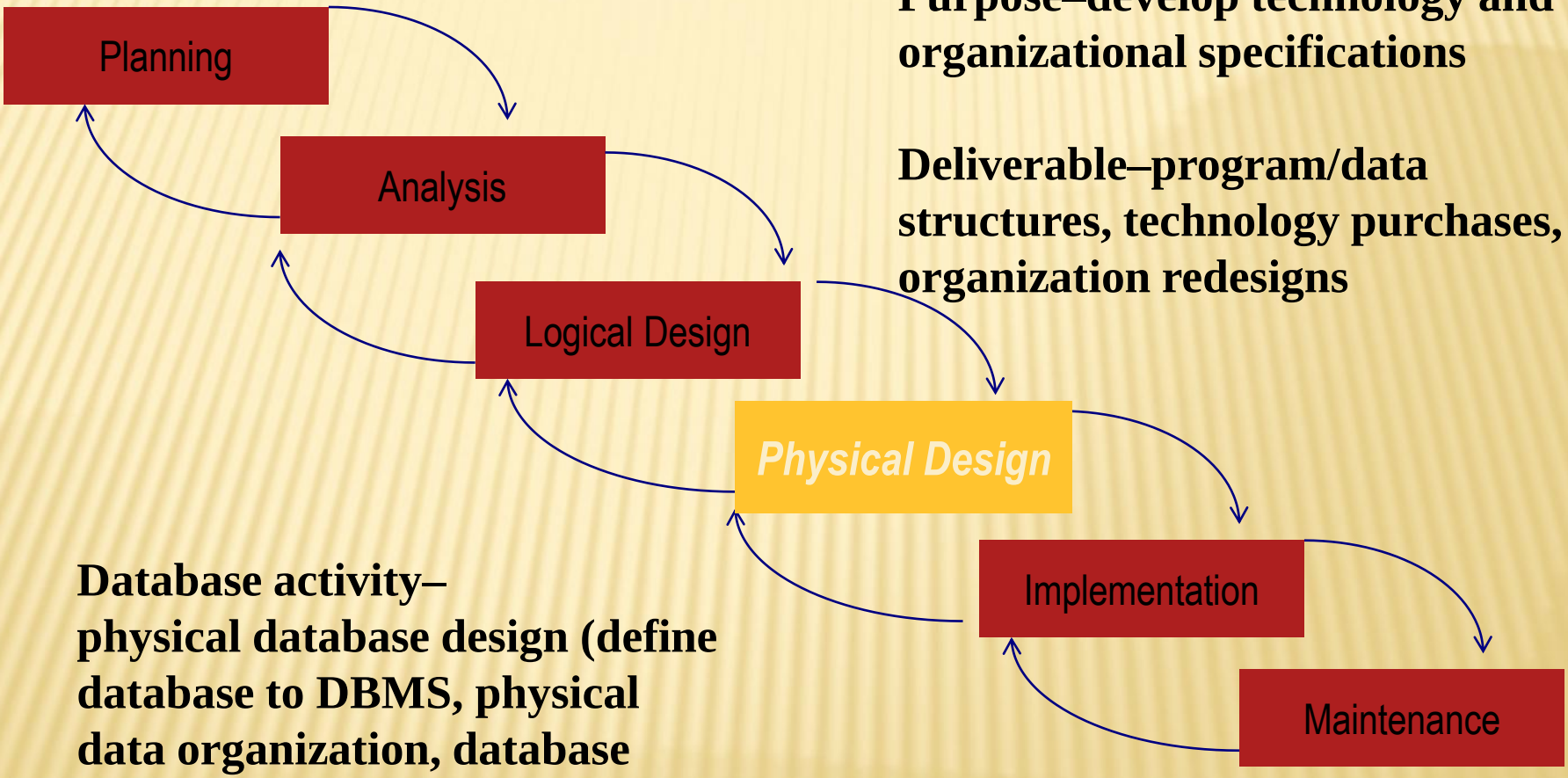
SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)



Purpose–information requirements elicitation and structure
Deliverable–detailed design specifications

Database activity–logical database design (transactions, forms, displays, views, data integrity and security)

SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)

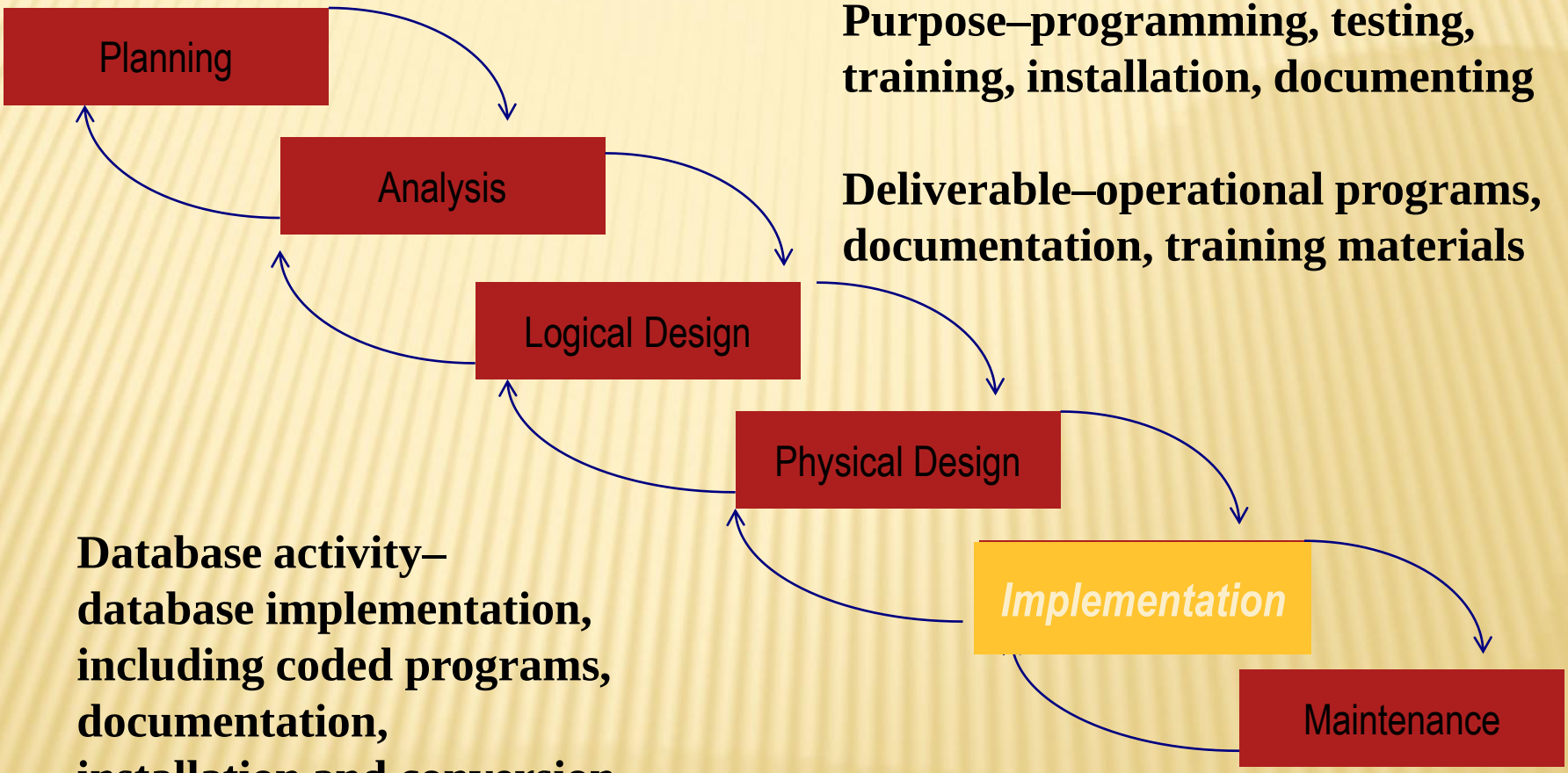


Purpose–develop technology and organizational specifications

Deliverable–program/data structures, technology purchases, organization redesigns

Database activity–physical database design (define database to DBMS, physical data organization, database processing programs)

SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)

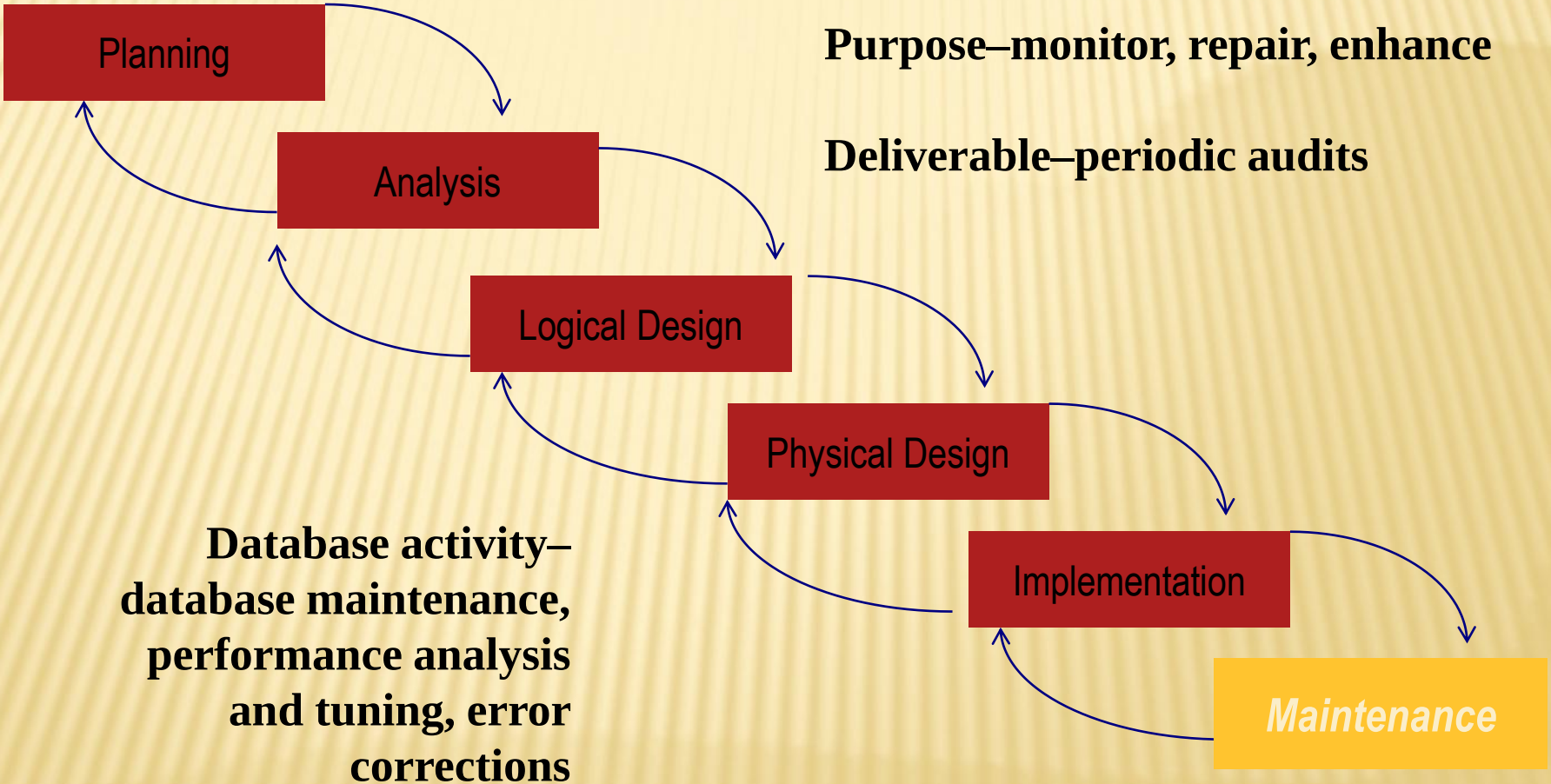


Purpose–programming, testing, training, installation, documenting

Deliverable–operational programs, documentation, training materials

Database activity–database implementation, including coded programs, documentation, installation and conversion

SYSTEMS DEVELOPMENT LIFE CYCLE (CONT.)

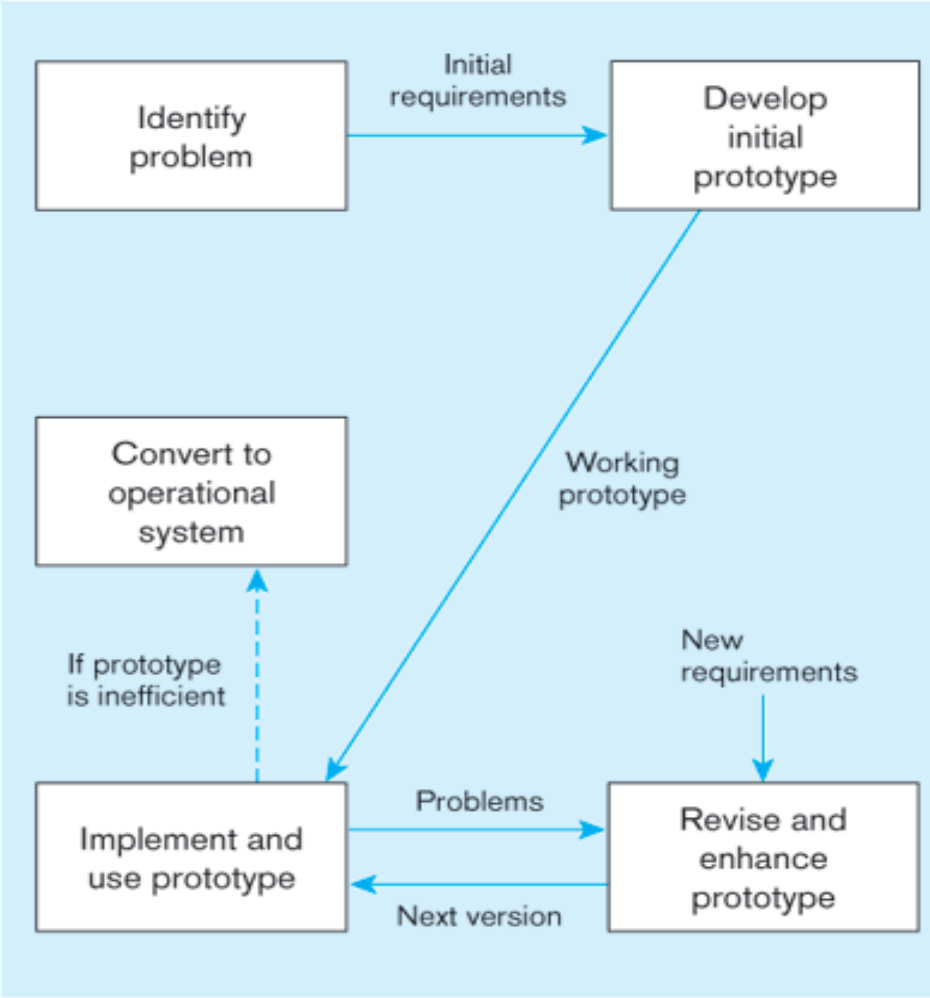


Purpose–monitor, repair, enhance

Deliverable–periodic audits

Database activity–database maintenance, performance analysis and tuning, error corrections

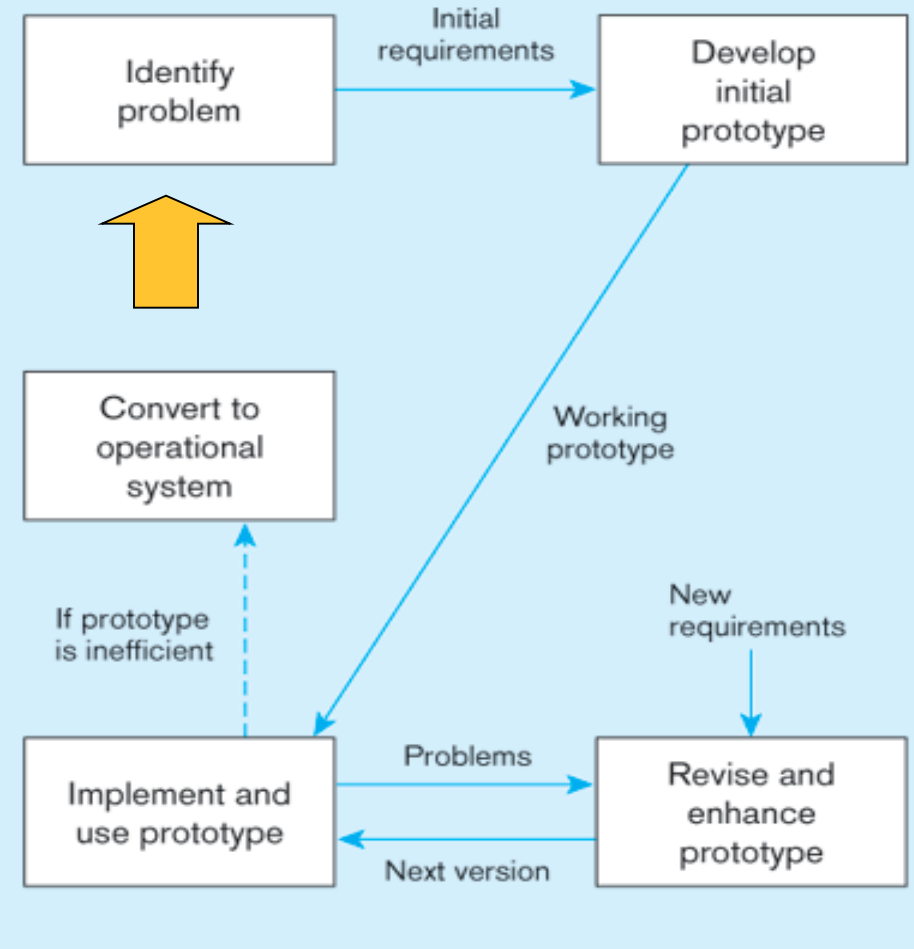
PROTOTYPING DATABASE METHODOLOGY



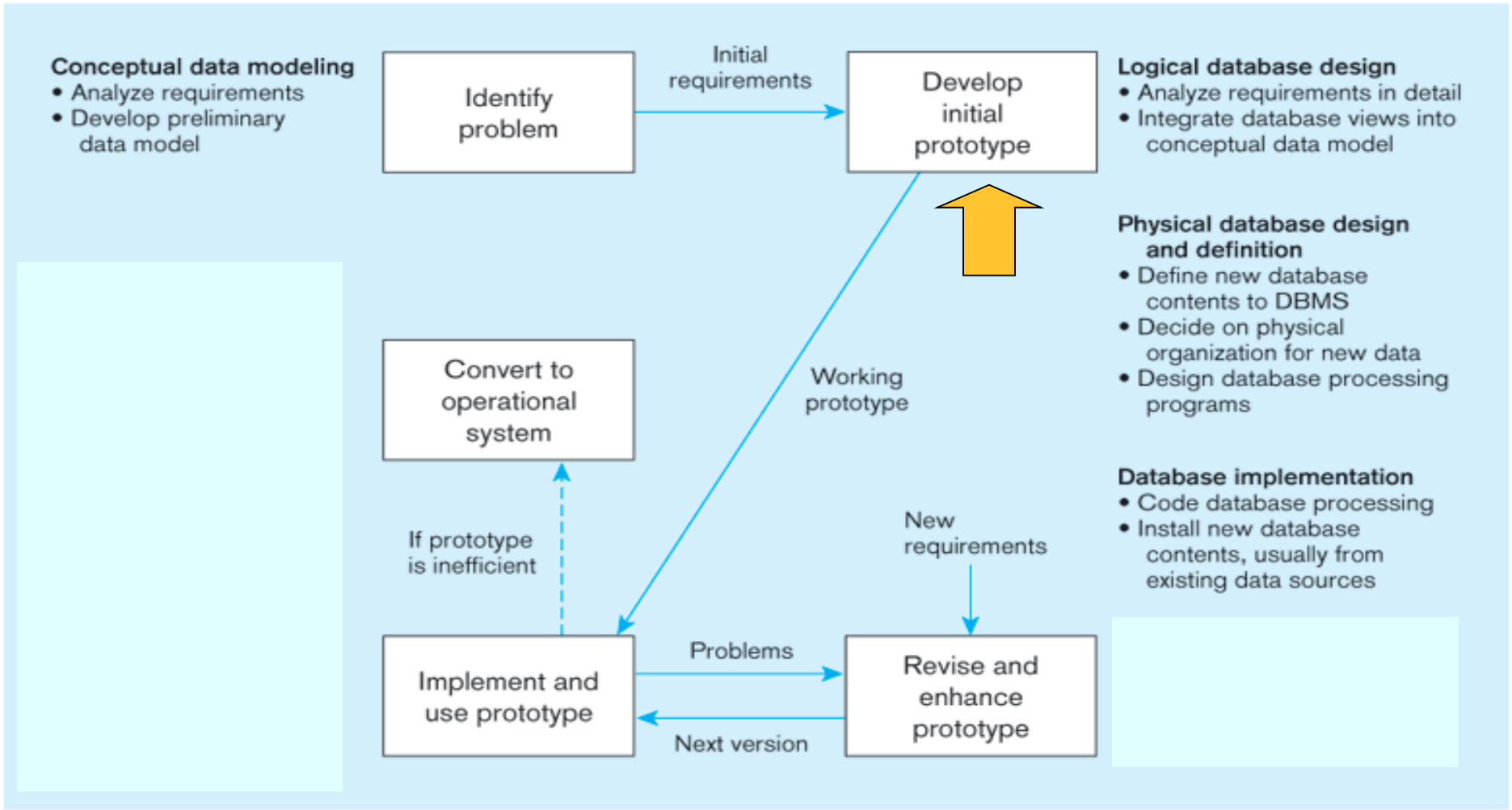
Prototyping is a classical Rapid Application Development (RAD) approach

PROTOTYPING DATABASE METHODOLOGY

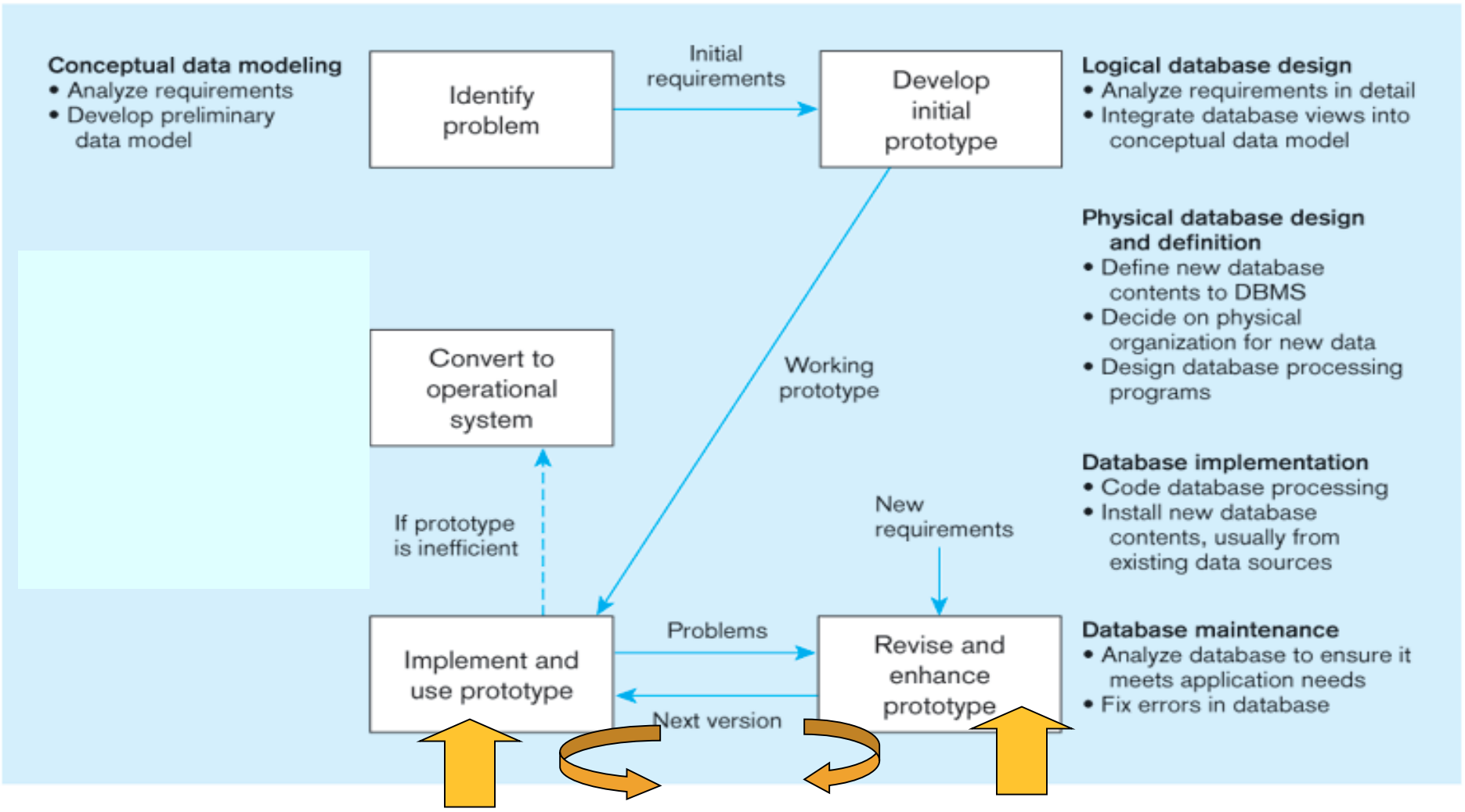
Conceptual data modeling
- Analyze requirements
- Develop preliminary data model



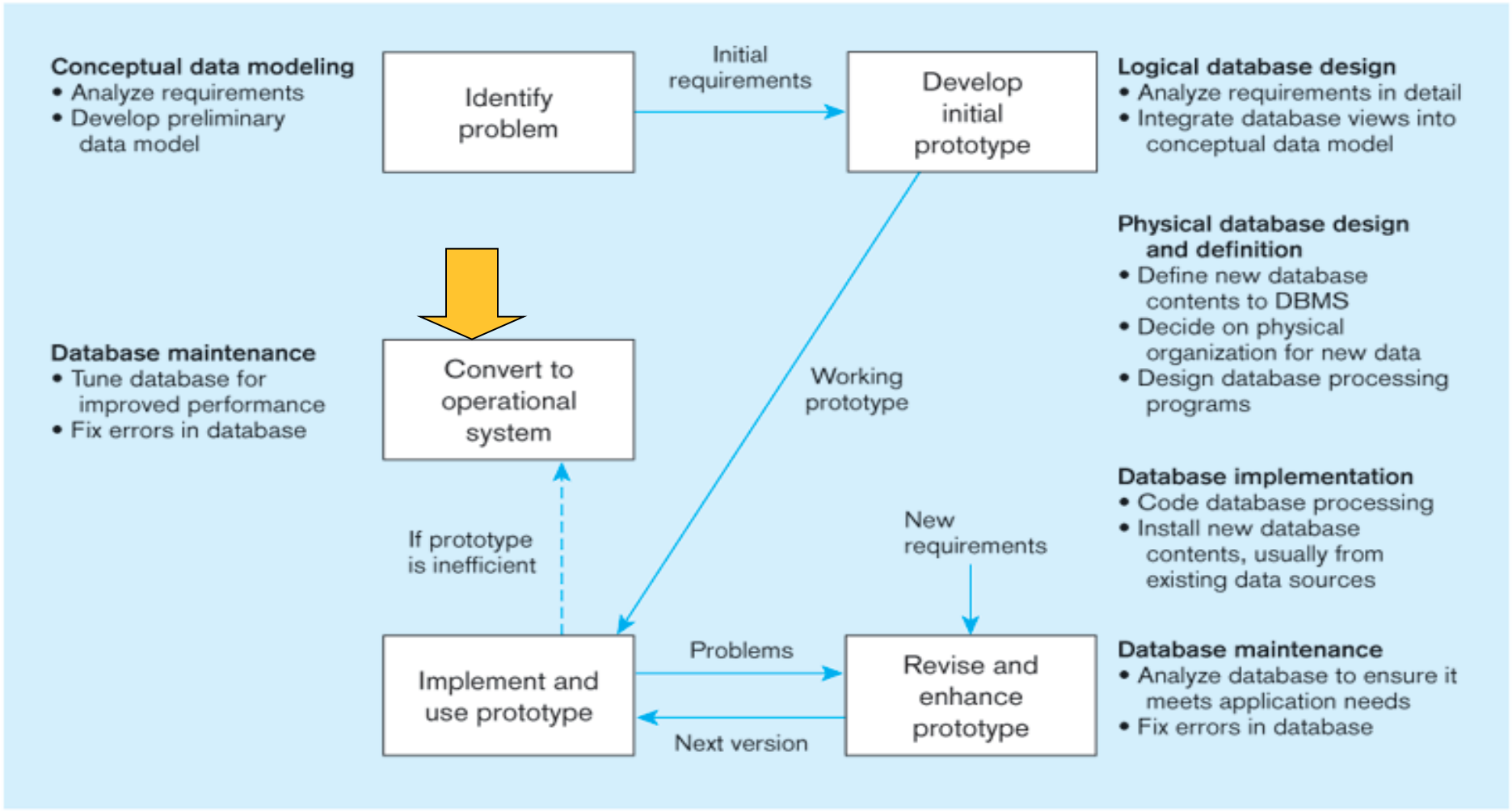
PROTOTYPING DATABASE METHODOLOGY



PROTOTYPING DATABASE METHODOLOGY



PROTOTYPING DATABASE METHODOLOGY



OTHER RAPID APPLICATION (RAD) APPROACHES

- Agile – emphasizes “individuals and interactions over processes and tools, working software over comprehensive documentation, customer collaboration over contract negotiation, and response to change over following a plan.” (The Agile Manifesto)

- Examples of agile programming methodologies
  - eXtreme programming
  - Scrum
  - DSDM (**Dynamic Systems Development Method**) Consortium
  - Feature-driven development

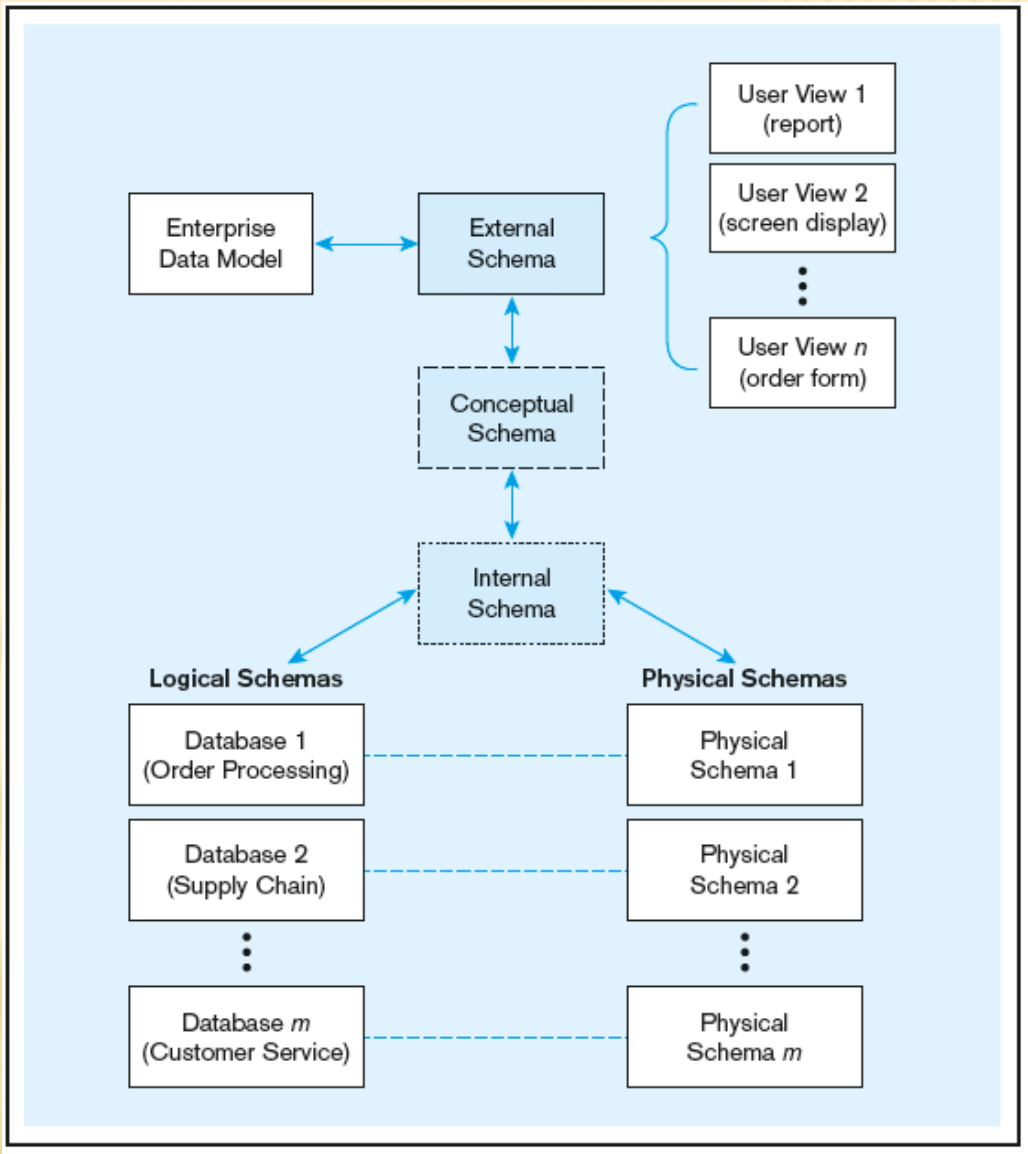
DATABASE SCHEMA

- External Schema
  - User Views
  - Subsets of Conceptual Schema
  - Can be determined from business-function/data entity matrices
  - DBA determines schema for different users
- Conceptual Schema
  - E-R models–
- Internal Schema
  - Logical structures
  - Physical structures

Three-schema architecture

Different people have different views of the database…these are the external schema

The internal schema is the underlying design and implementation



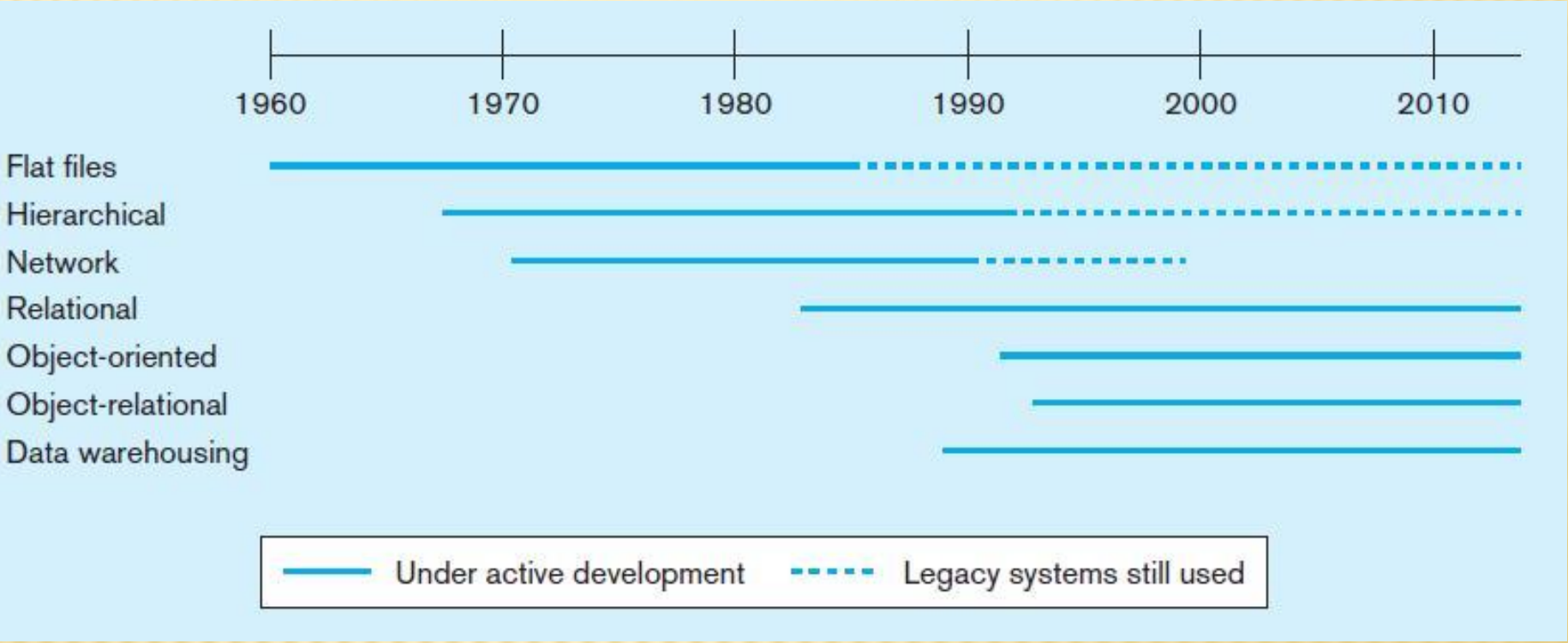
MANAGING PEOPLE AND PROJECTS

- Project–a planned undertaking of related activities to reach an objective that has a beginning and an end
- Initiated and planned in planning stage of SDLC
- Executed during analysis, design, and implementation
- Closed at the end of implementation

MANAGING PEOPLE AND PROJECTS

- Business analysts
- Systems analysts
- Database analysts and data modelers
- Users
- Programmers
- Database architects
- Data administrators
- Project managers
- Other technical experts

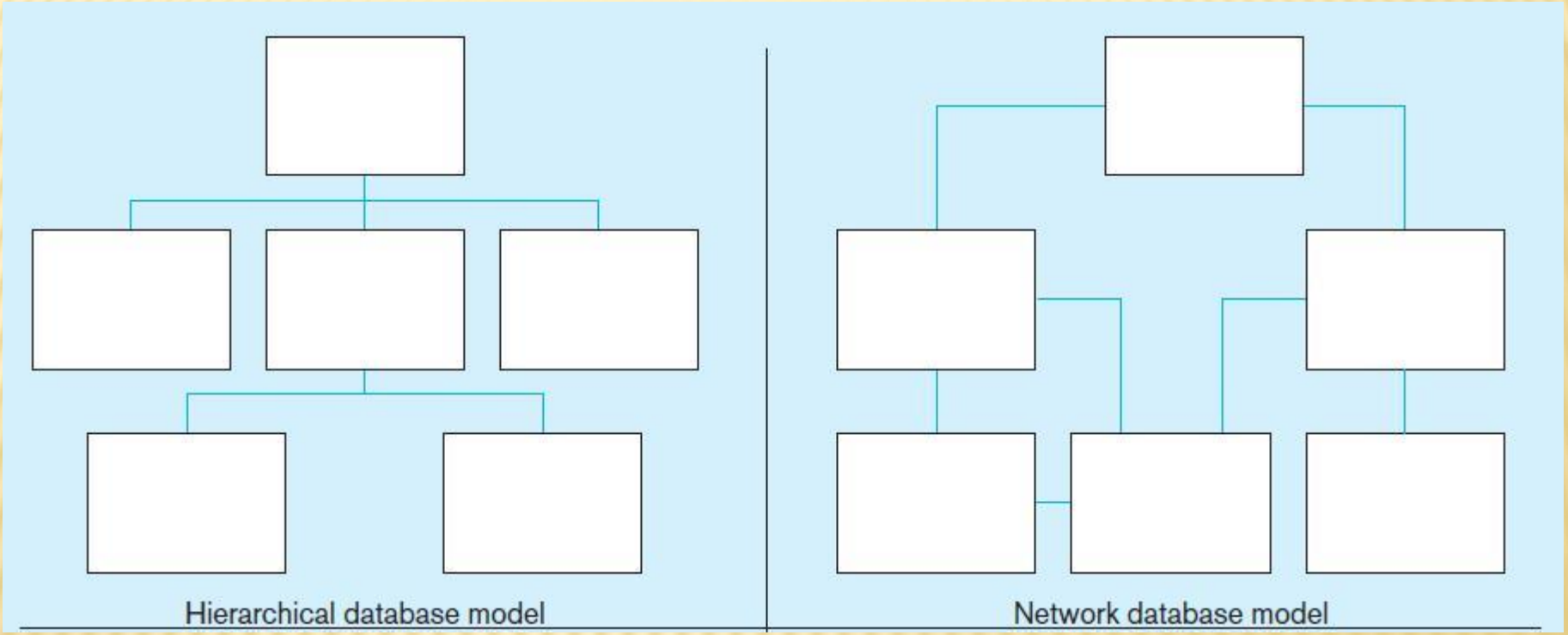
Evolution of database technologies



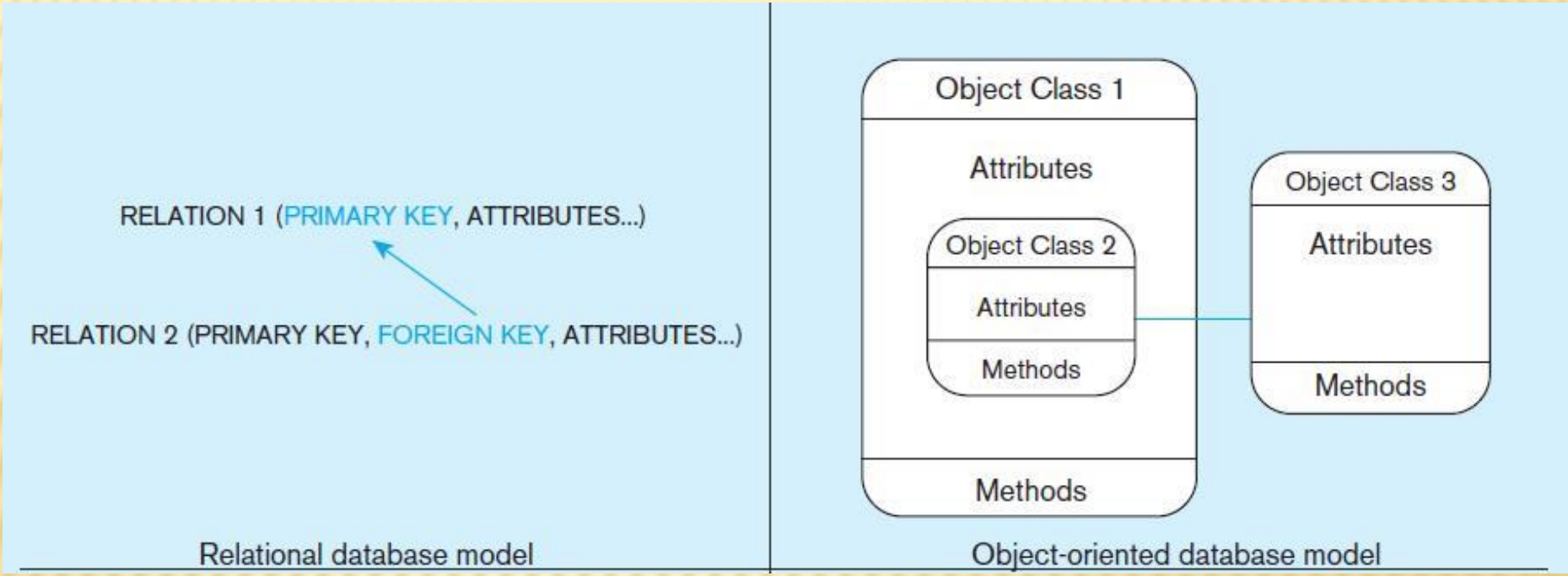
EVOLUTION OF DATABASE SYSTEMS

- Driven by four main objectives:
  - Need for program-data independence → reduced maintenance
  - Desire to manage more complex data types and structures
  - Ease of data access for less technical personnel
  - Need for more powerful decision support platforms

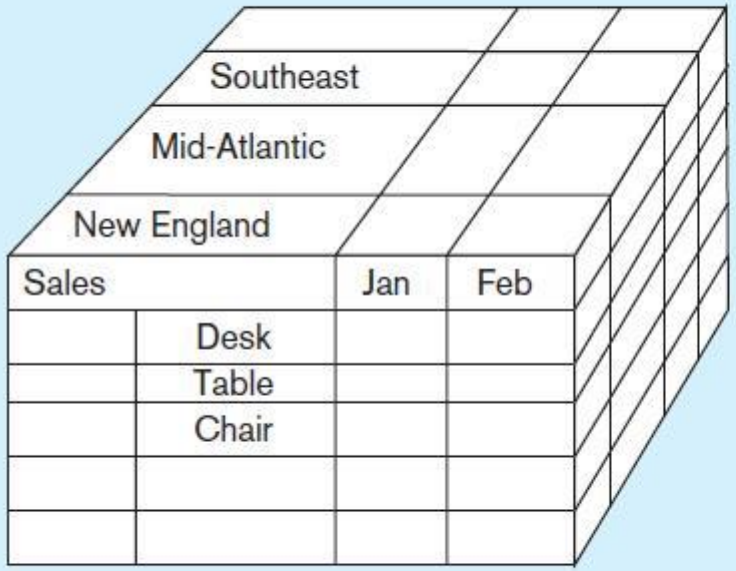
Database architectures



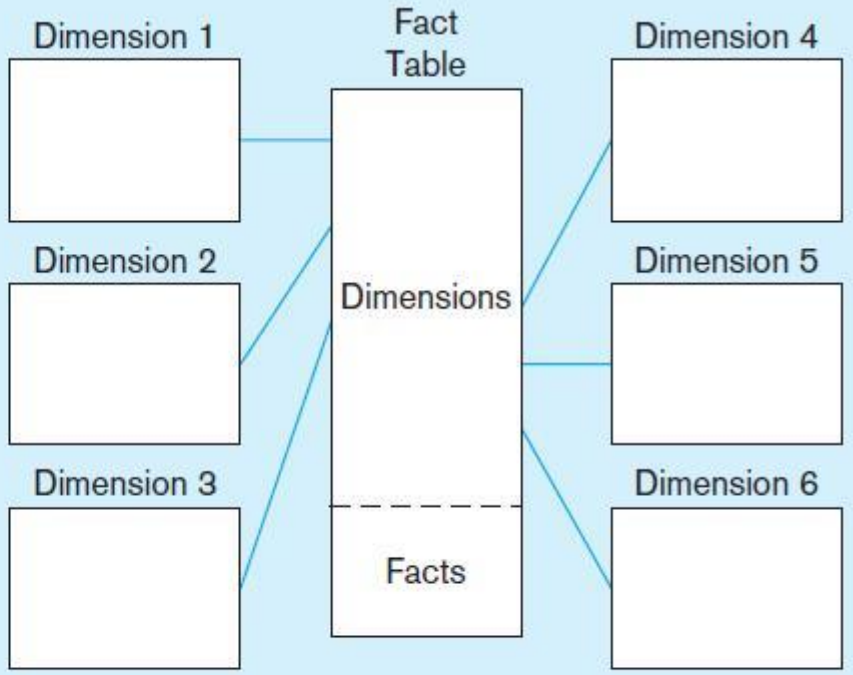
Database architectures (cont.)



Database architectures (cont.)



Multidimensional database model – multidimensional cube view



Multidimensional database model – star-schema view

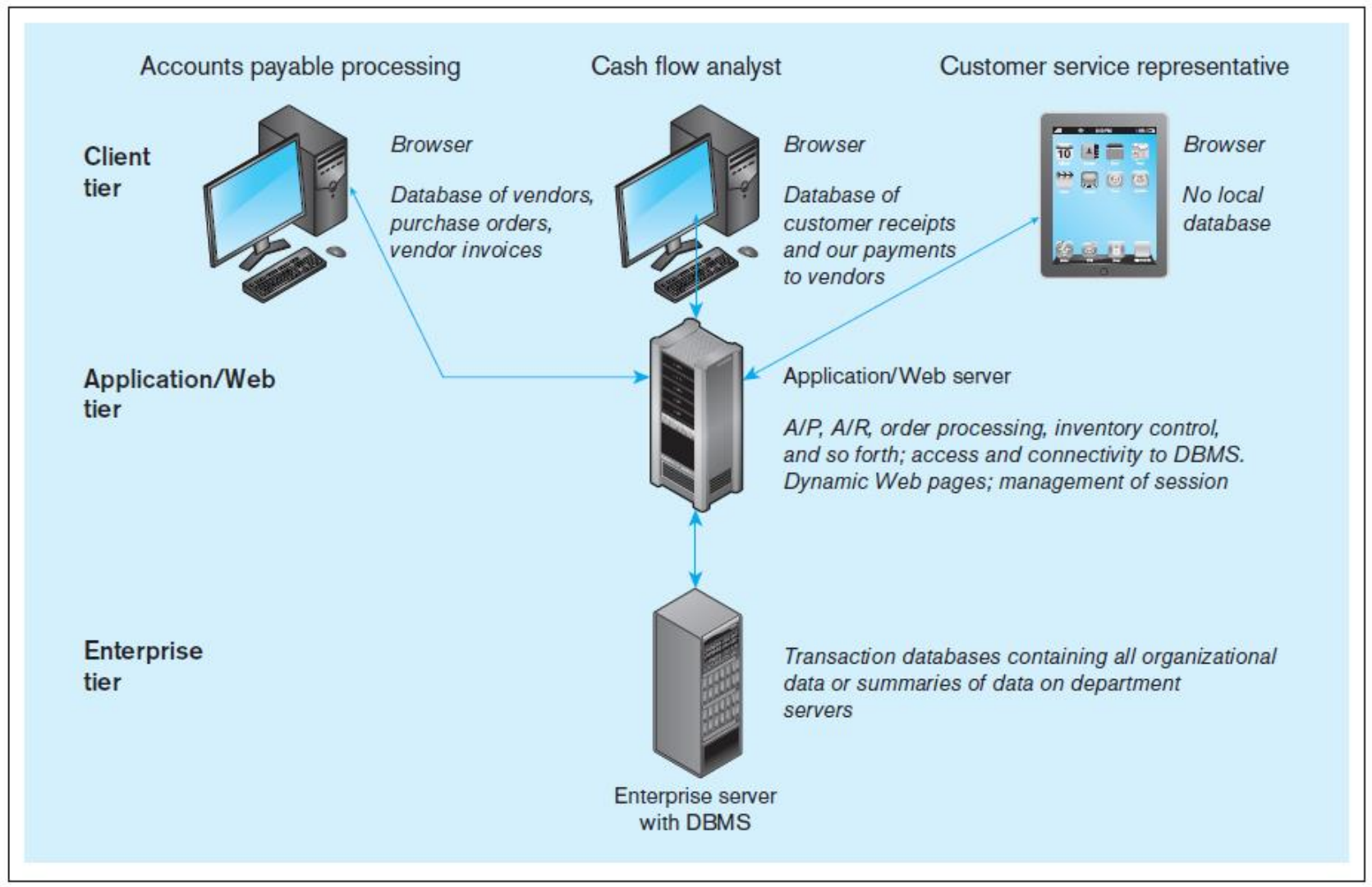
THE RANGE OF DATABASE APPLICATIONS

- Personal databases
- Two-tier and N-tier Client/Server databases
- Enterprise applications
  - Enterprise resource planning (ERP) systems
  - Data warehousing implementations

TABLE 1-5 Summary of Database Applications

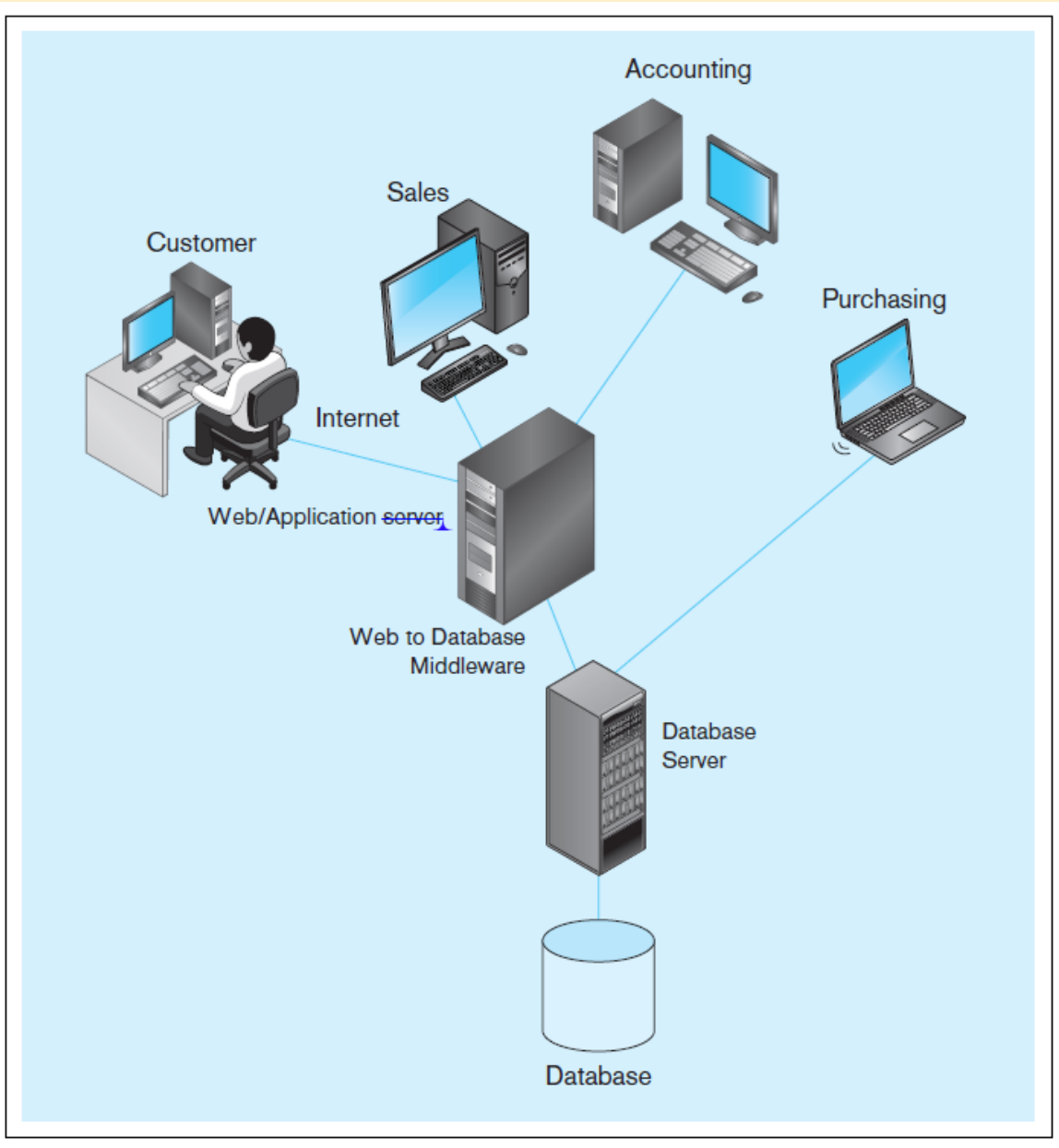
Type of Database / Application	Typical Number of Users	Typical Size of Database
Personal	1	Megabytes
Multitier Client/Server	100–1000	Gigabytes
Enterprise resource planning	>100	Gigabytes–terabytes
Data warehousing	>100	Terabytes–petabytes

Architecture



ENTERPRISE DATABASE APPLICATIONS

- Enterprise Resource Planning (ERP)
  - Integrate all enterprise functions (manufacturing, finance, sales, marketing, inventory, accounting, human resources)
- Data Warehouse
  - Integrated decision support system derived from various operational databases



Computer System for Pine Valley Furniture Company

Project data model for Home Office product line marketing support system