The Database Environment

MSBA 7024 / MACC 7020

Database Design and Management

Handout 1

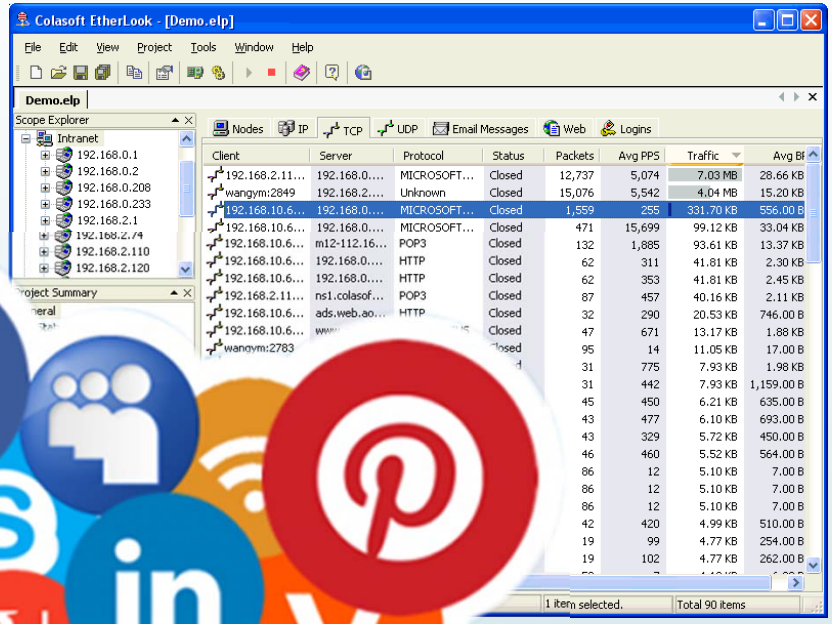
Objectives

- Definition of terms
- Explain growth and importance of databases
- Name limitations of conventional file processing
- Explain advantages and costs of databases
- List components of database environment
- Describe evolution of database systems
- Describe and compare the two major approaches to database and system development

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 be useful
- Metadata: data that describes the properties and context of user data

Data is everywhere



Accelerometer

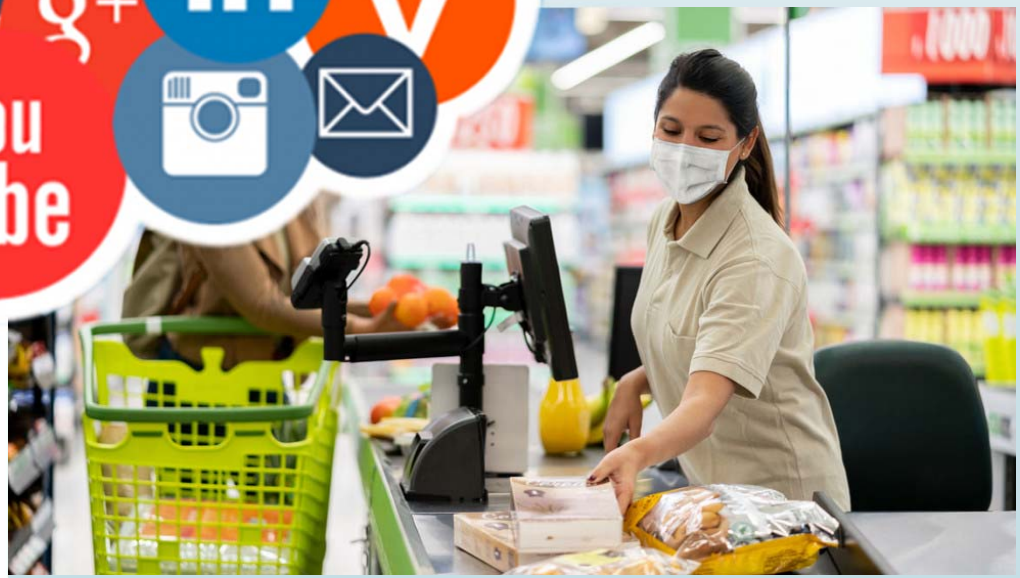
Digital Compass

Microphone

WiFi

Bluetooth

GPS



Data in context

Class Roster

Course: MGT 500 Business Policy

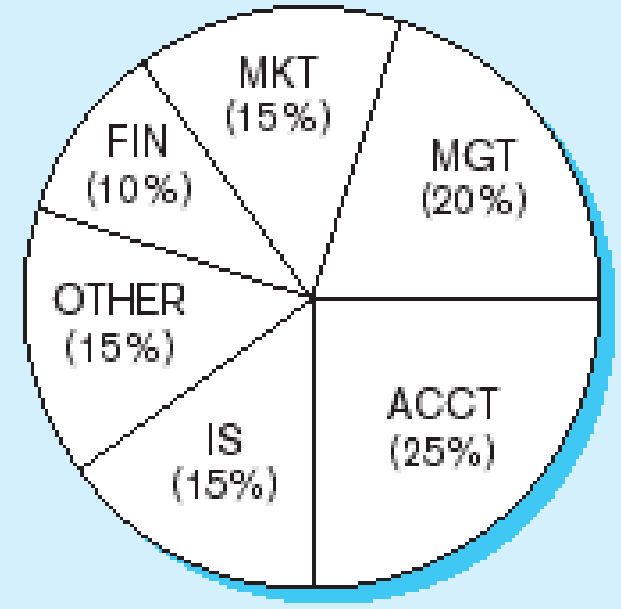
Semester: Spring 200X

Section: 2

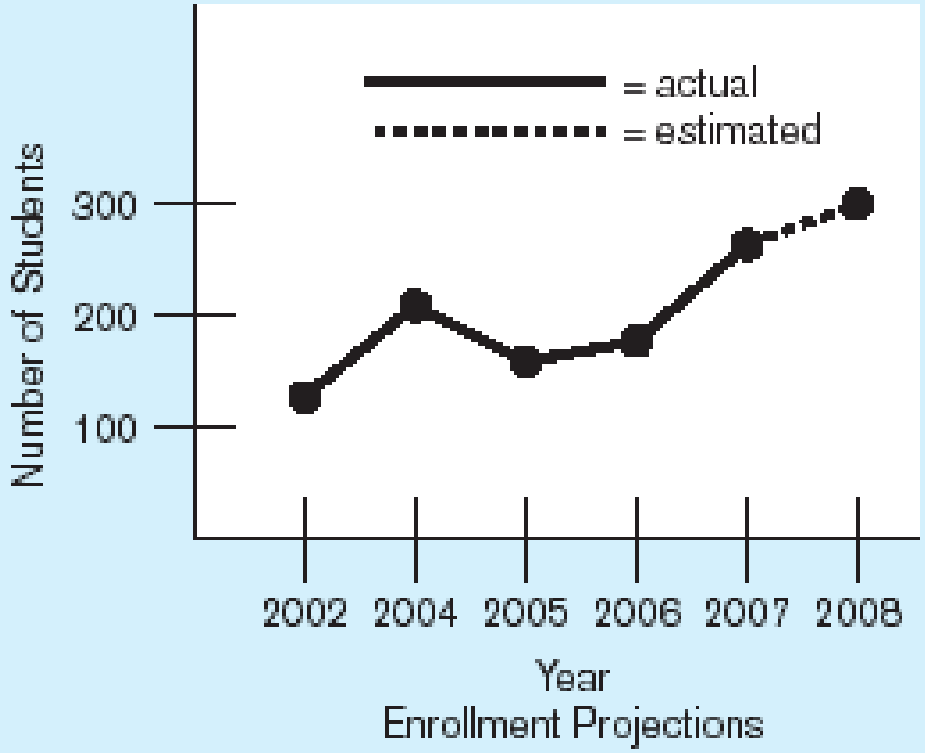
Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Context helps users understand data

Summarized data



Percent Enrollment by Major (200X)



Enrollment Projections

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

Table 1-1 Example Metadata for Class Roster

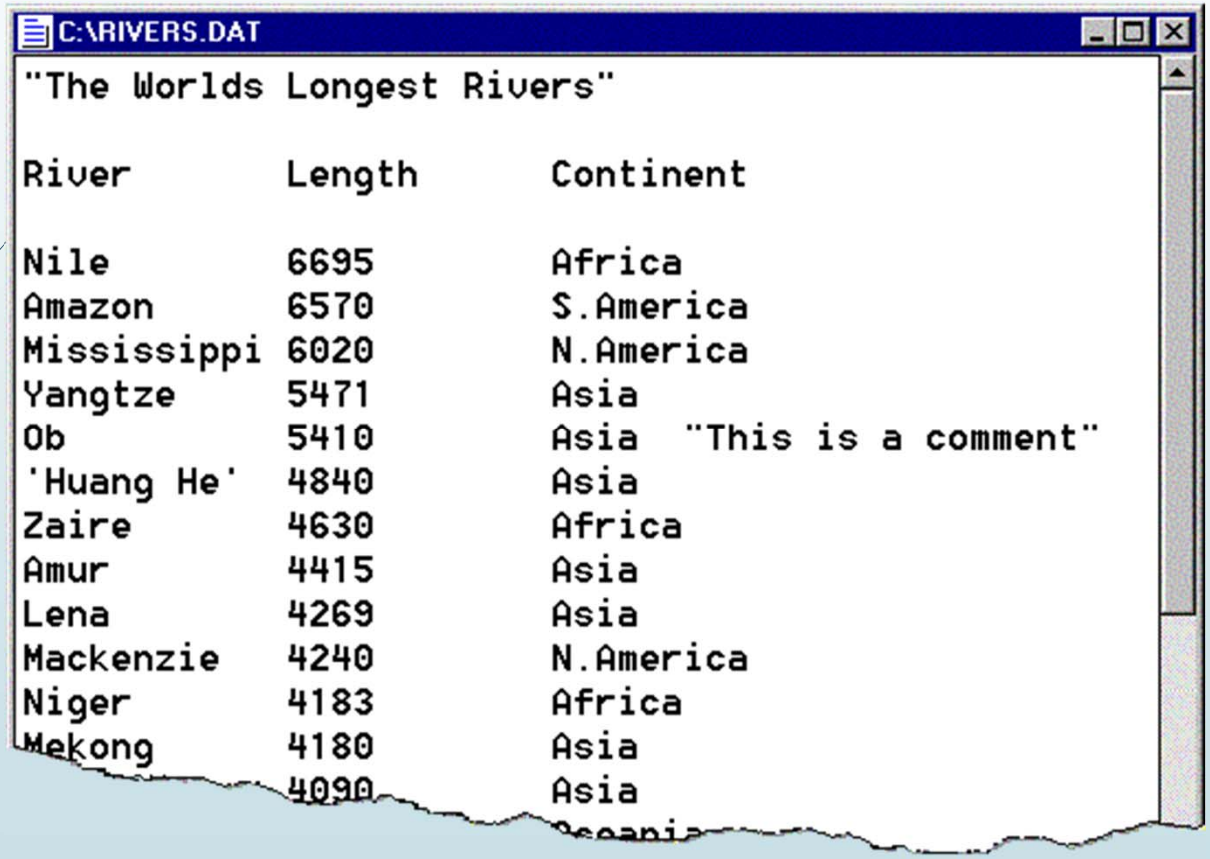
Data Item		*Value*				
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	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

How to store data in computers?

- In the beginning, data were stored as files.

Example data file



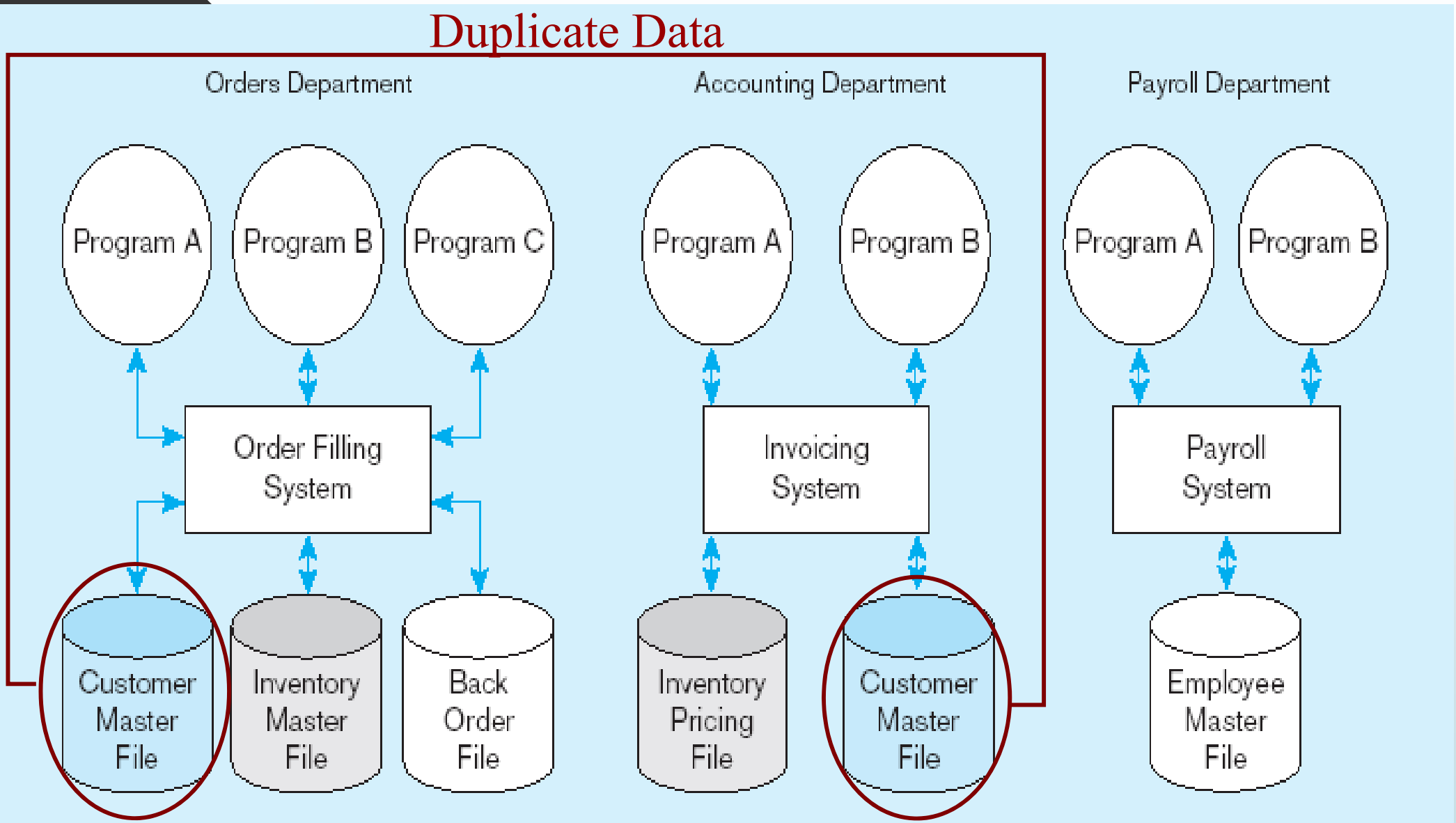
Disadvantages of File Processing

- **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

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***

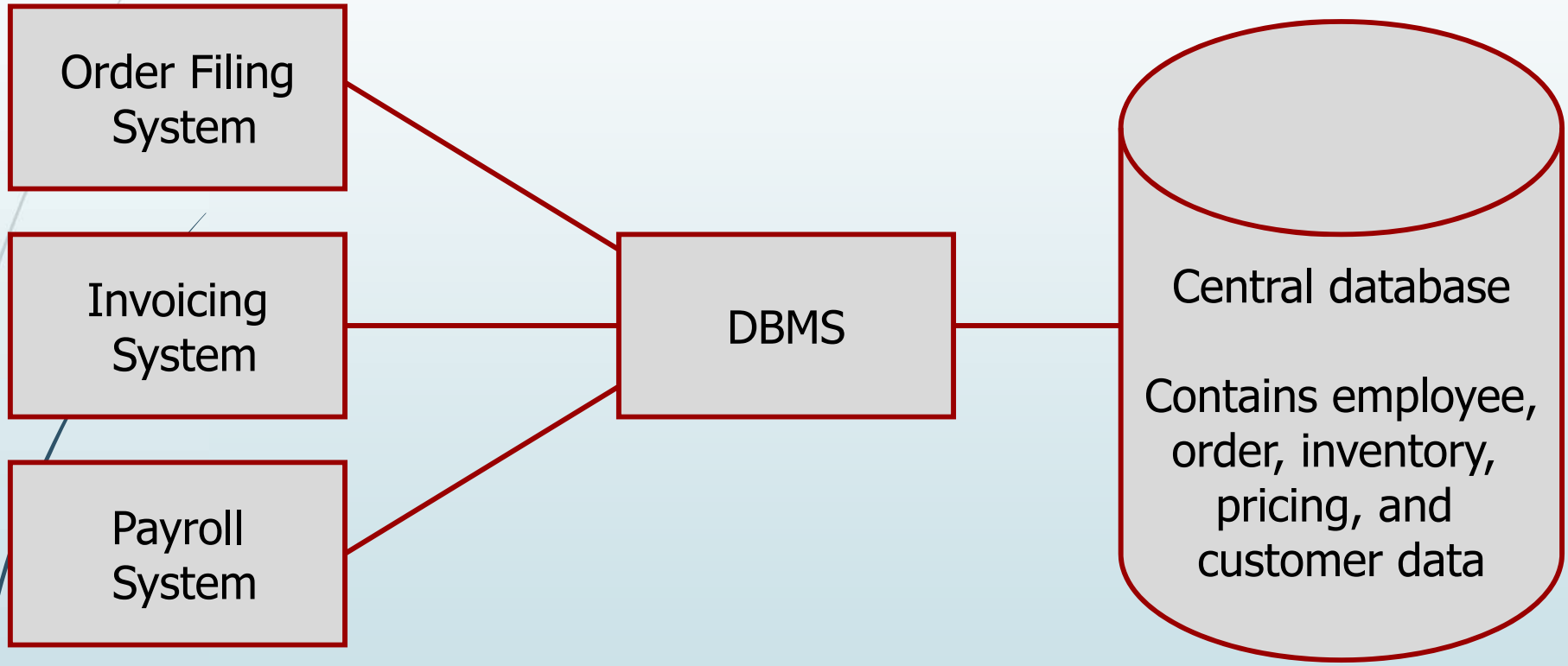
SOLUTION:
The DATABASE Approach

- 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

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

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

Elements of the Database Approach

- Data models
  - Graphical system 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
- Database management system
  - A software system that is used to create, maintain, and provide controlled access to user databases

Elements of the Database Approach

- Use of Internet Technology
  - Networks and telecommunications, distributed databases, client-server, and 3-tier architectures
- Database Applications
  - Application programs used to perform database activities (create, read, update, and delete) for database users

Denormalized vs Normalized Data

Denormalized Data

TransactionID	CustomerName	CustomerAddress	CustomerCity	ProductName	ProductType	Manufacturer	Price	Quantity
1	John Lee	123 Pokfulam Road	Hong Kong	Street Fighter 6	PS5 Game	Capcom	360	1
2	John Lee	123 Pokfulam Road	Hong Kong	FIFA	PS5 Game	EA Sports	320	1
3	John Lee	123 Pokfulam Road	Hong Kong	Baseball Cap	Apparel	HKU	130	2
4	Mary Chen	333 Nanjing Road	Shanghai	Blackpink Gift Box	Gift Box	Blackpink	150	3
5	Mary Chen	333 Nanjing Road	Shanghai	Baseball Cap	Apparel	HKU	130	1
6	Steven Brown	111 First Avenue	New York	ThinkPad Computer	Computer	Lenovo	7999	1
7	Steven Brown	111 First Avenue	New York	Baseball Cap	Apparel	HKU	130	2
8	Helen Wong	506 Nathan Road	Hong Kong	Blackpink Gift Box	Gift Box	Blackpink	150	1
9	Helen Wong	506 Nathan Road	Hong Kong	Harry Potter Book 1	Book	Bloomsbury	70	1

Normalized Data

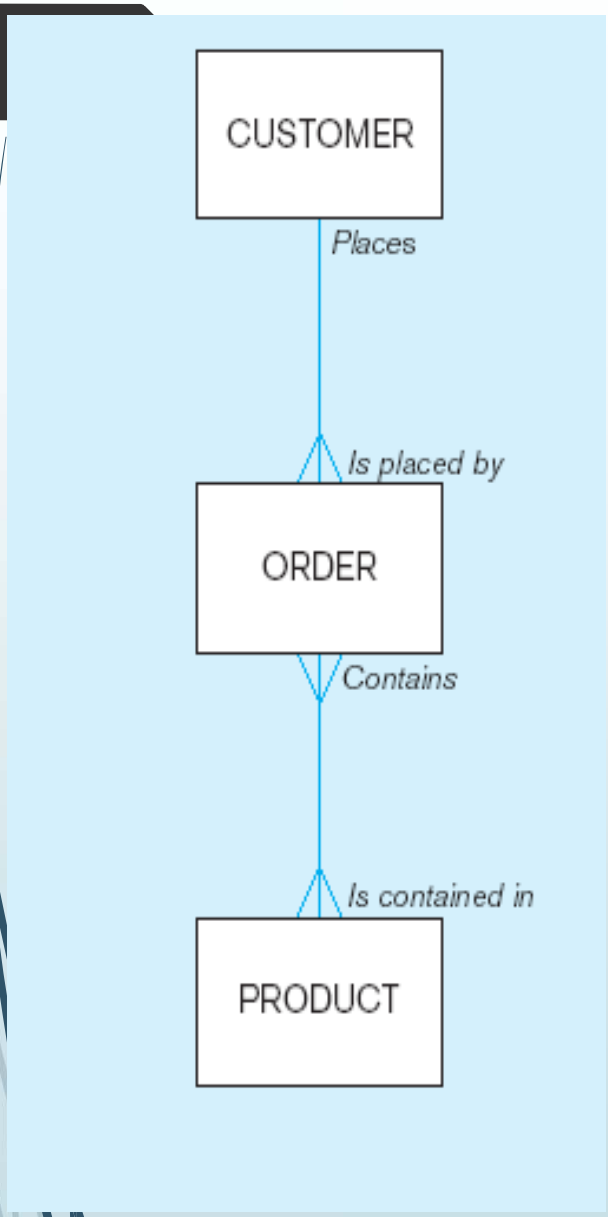
TRANSACTION			
TransactionID	CustomerID	ProductID	Quantity
1	1	1	1
2	1	2	1
3	1	3	2
4	2	4	3
5	2	3	1
6	3	5	1
7	3	3	2
8	4	4	1
9	4	6	1

CUSTOMER			
CustomerID	CustomerName	CustomerAddress	CustomerCity
1	John Lee	123 Pokfulam Road	Hong Kong
2	Mary Chen	333 Nanjing Road	Shanghai
3	Steven Brown	111 First Avenue	New York
4	Helen Wong	506 Nathan Road	Hong Kong

PRODUCT				
ProductID	ProductName	ProductType	Manufacturer	Price
1	Street Fighter 6	PS5 Game	Capcom	360
2	FIFA	PS5 Game	EA Sports	320
3	Baseball Cap	Apparel	HKU	130
4	Blackpink Gift Box	Gift Box	Blackpink	150
5	ThinkPad Computer	Computer	Lenovo	7999
6	Harry Potter Book 1	Book	Bloomsbury	70

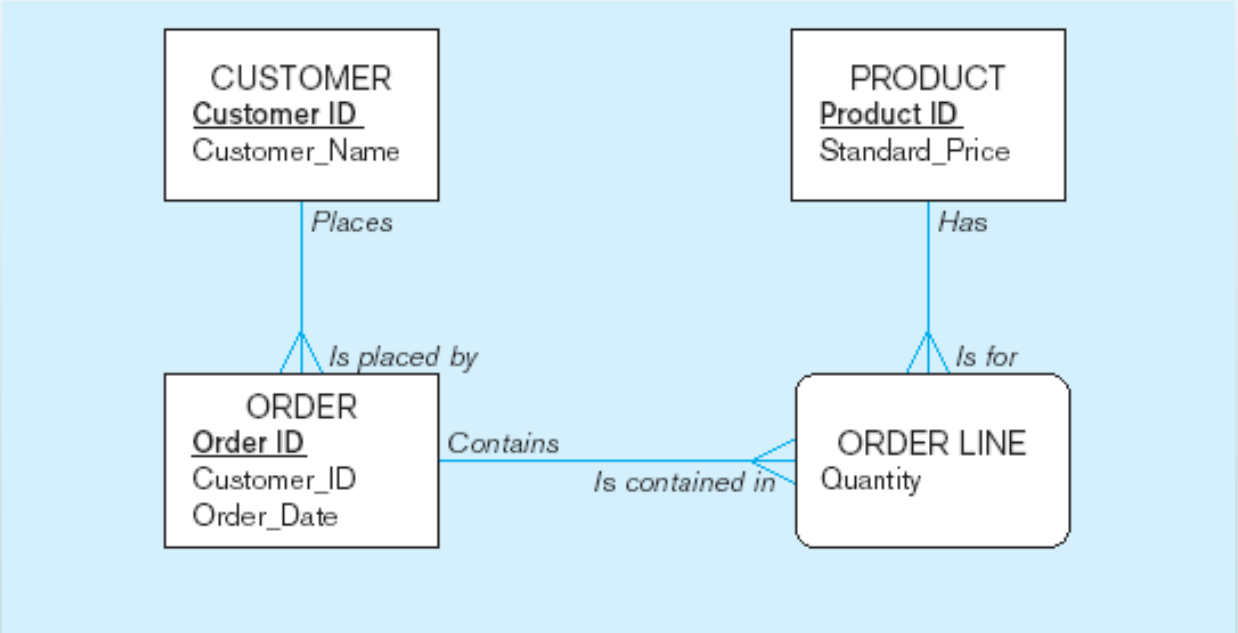
Example: Relational Model

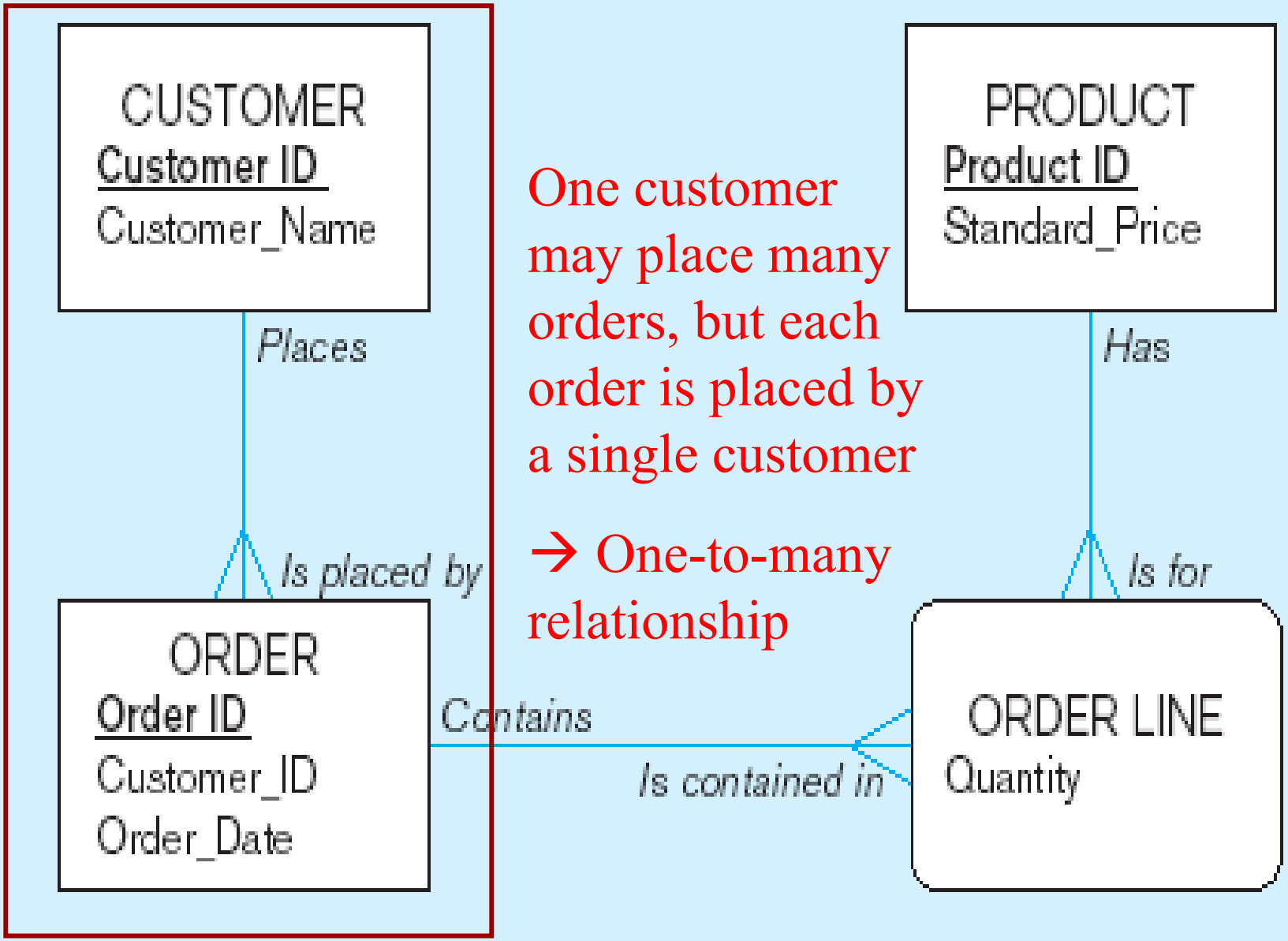
- Relational Databases
  - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships
- Store normalized data
- Most widely used in operational systems



Segment of an Enterprise Data Model

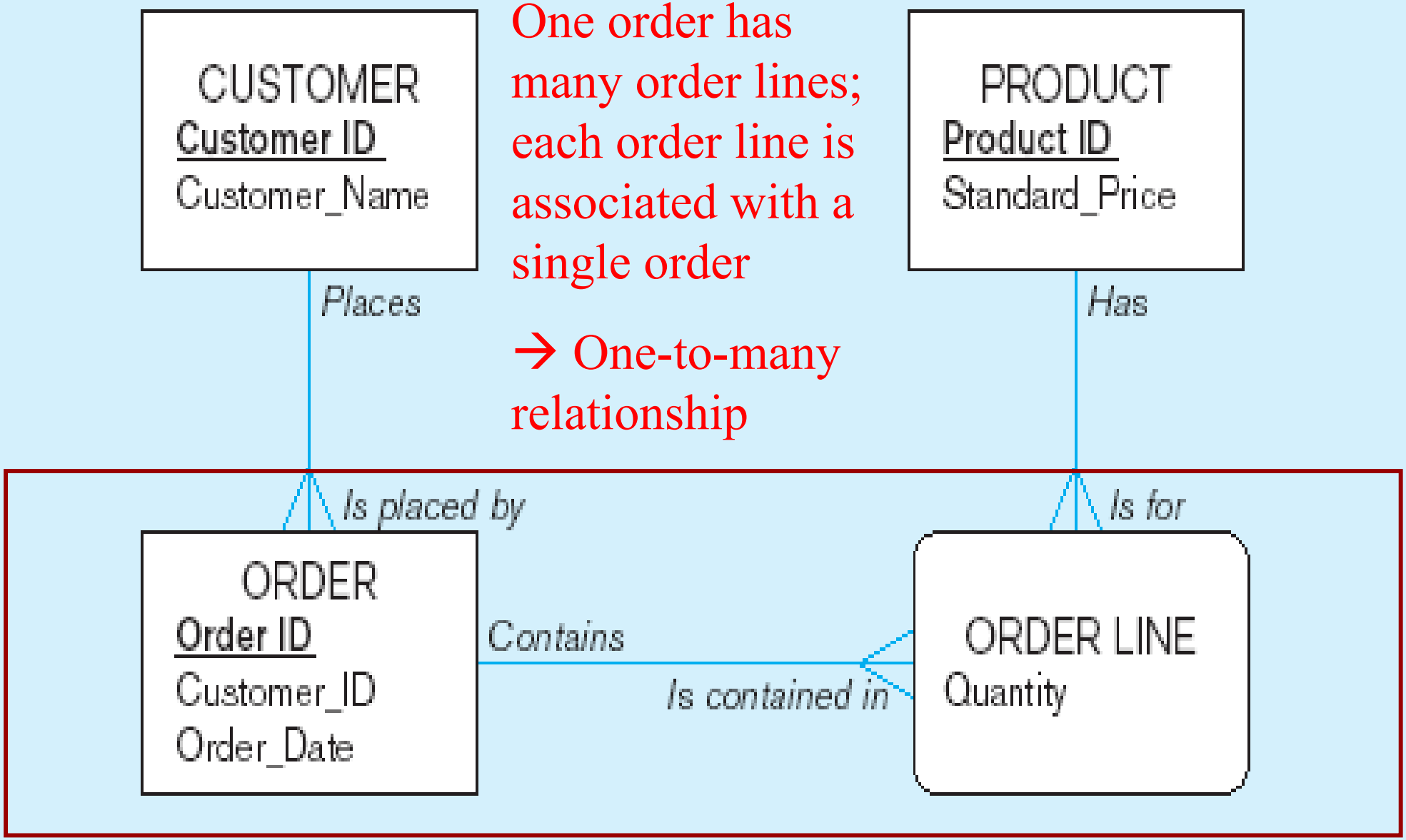
Segment of a Project-Level Data Model

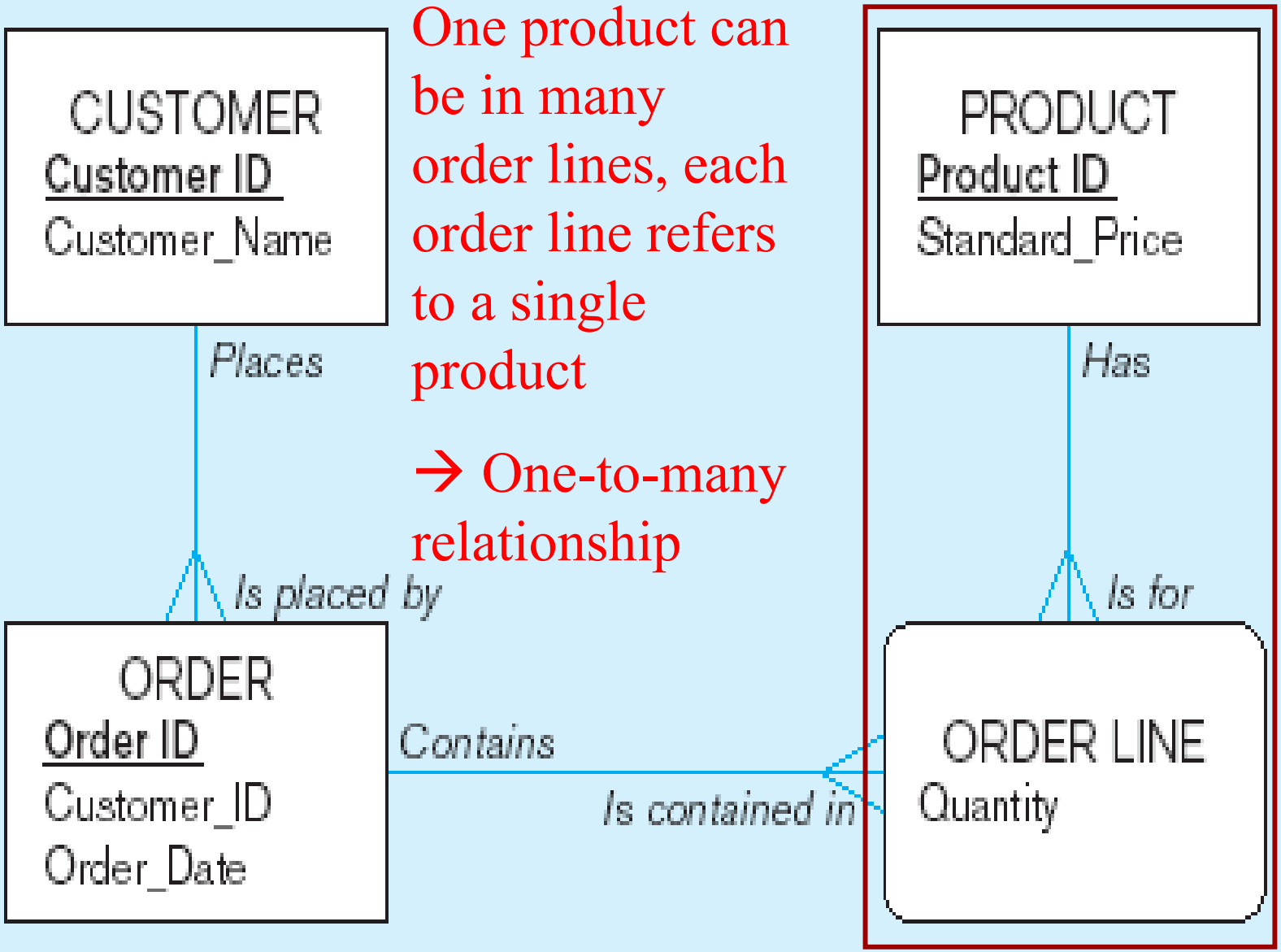




One customer may place many orders, but each order is placed by a single customer

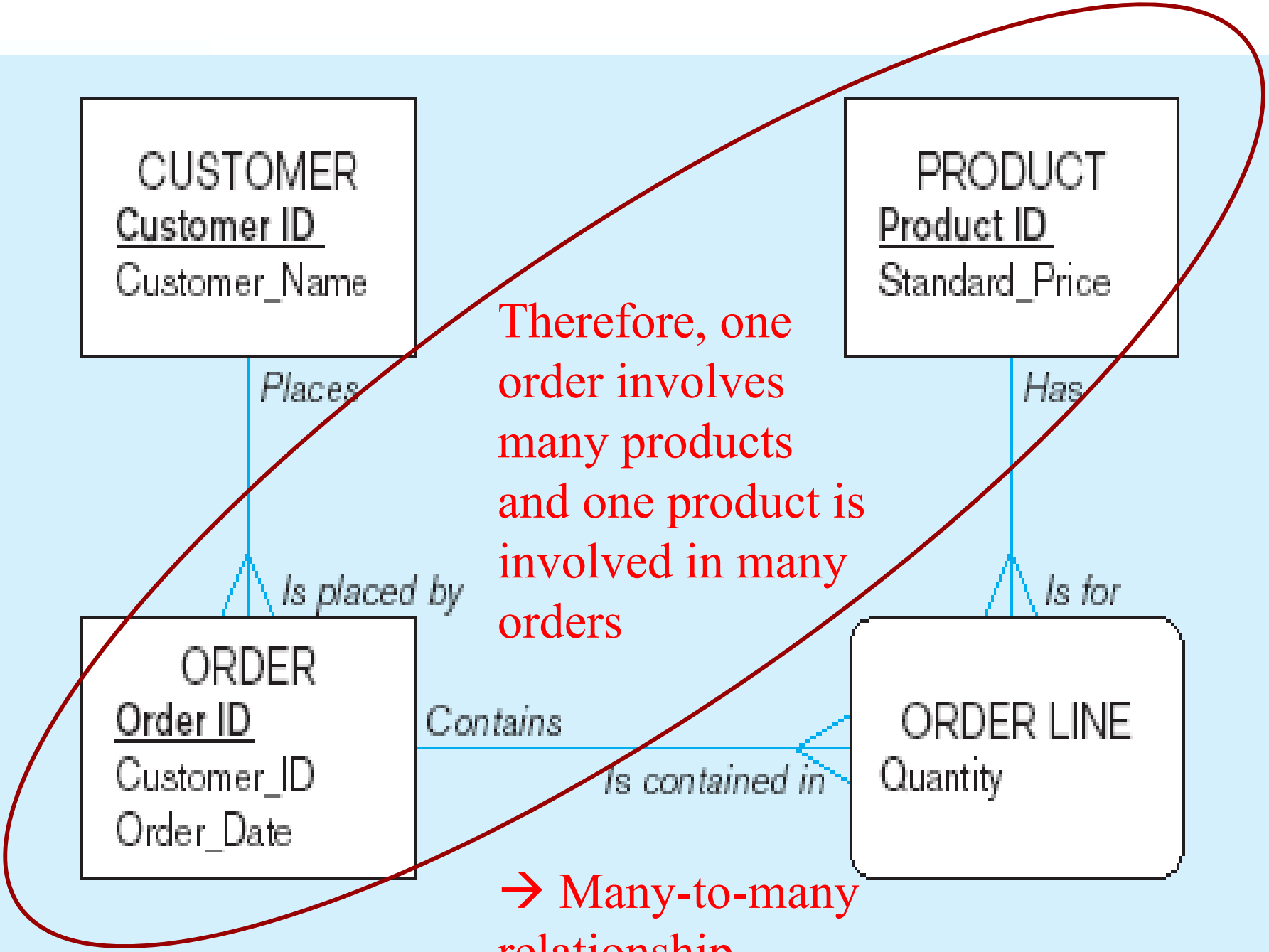
→ One-to-many relationship


CUSTOMER
Customer ID
Customer_Name
PRODUCT
Product ID
Standard_Price
One order has many order lines; each order line is associated with a single order
→ One-to-many relationship
Places
Has
Is placed by
Is for
ORDER
Order ID
Customer_ID
Order_Date
Contains
Is contained in
ORDER LINE
Quantity

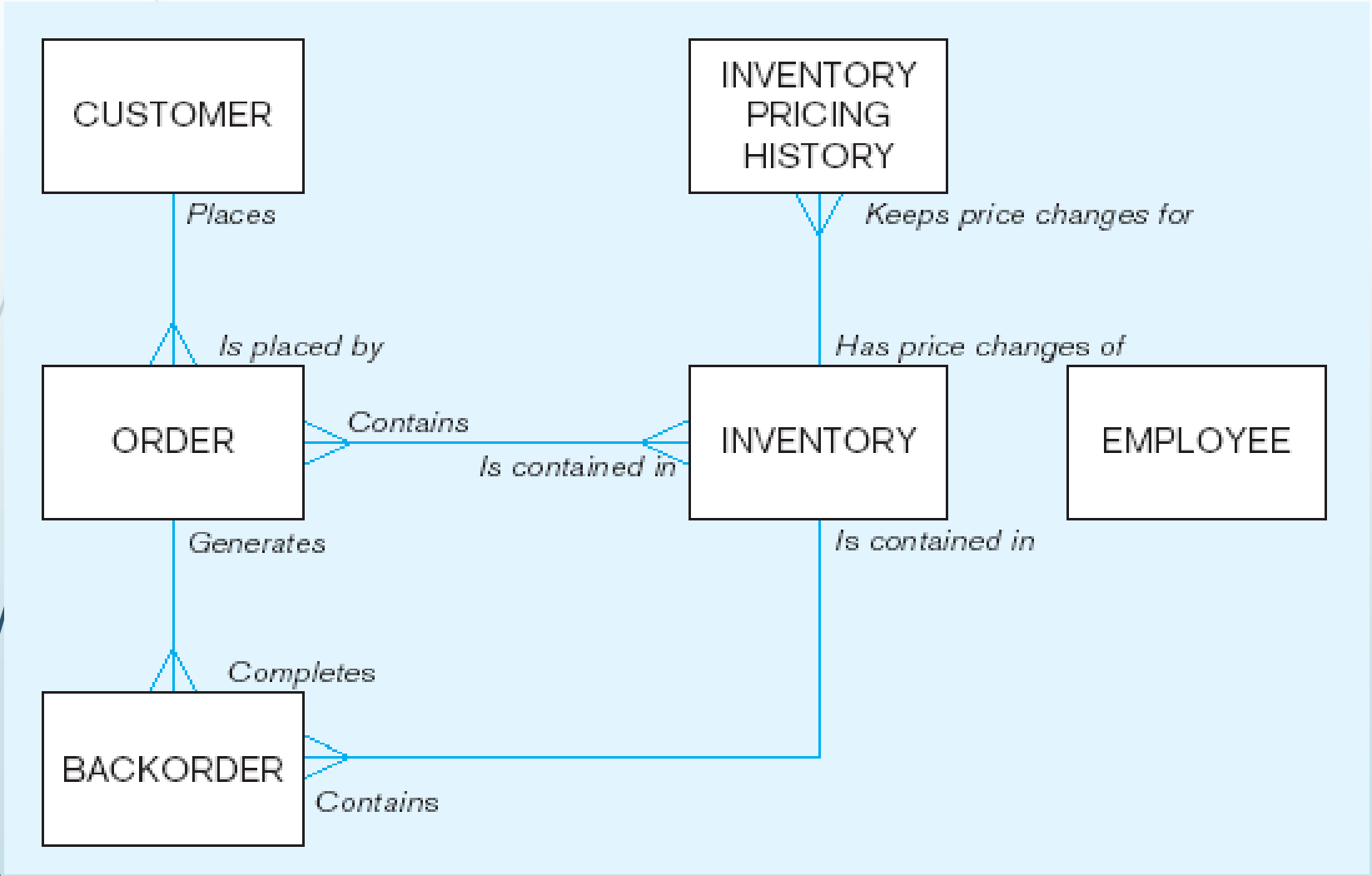


One product can be in many order lines, each order line refers to a single product

→ One-to-many relationship

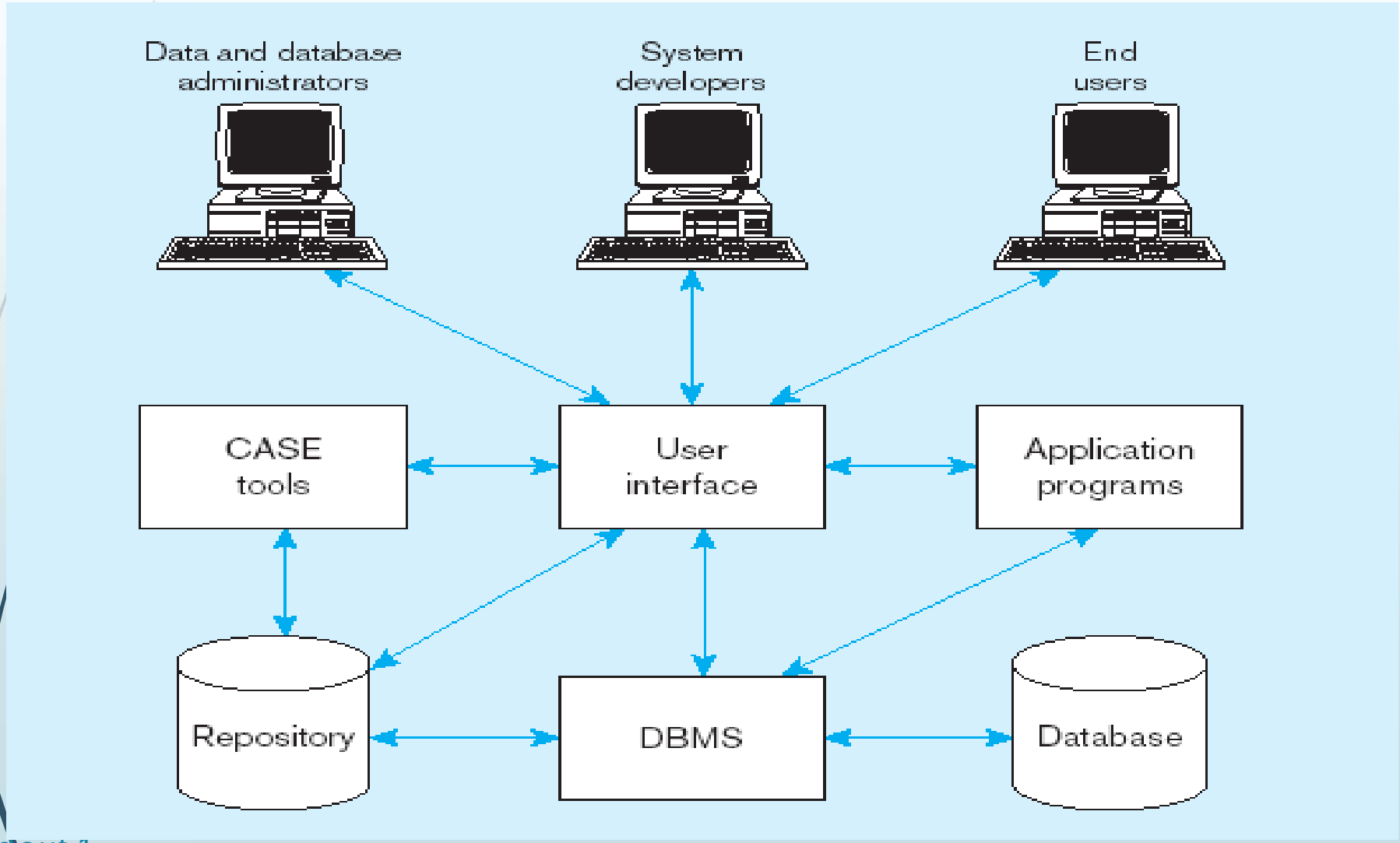

CUSTOMER
Customer ID
Customer_Name
PRODUCT
Product ID
Standard_Price
Places
Has
Is placed by
Is for
ORDER
Order ID
Customer_ID
Order_Date
Contains
Is contained in
ORDER LINE
Quantity
Therefore, one order involves many products and one product is involved in many orders
→ Many-to-many relationship

Enterprise data model



Handout 1

Components of the Database Environment



Components of the Database Environment

- **CASE Tools**–computer-aided software engineering
- **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 and graphical displays to users
- **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

The Range of Database Applications

- Personal databases
- Workgroup databases
- Departmental/divisional databases
- Enterprise database

TABLE 1-5 Summary of Database Applications

Type of Database / Application	Typical Number of Users	Typical Size of Database
Personal	1	Megabytes
Two-tier	5–100	Megabytes–gigabytes
Three-tier	100–1000	Gigabytes
Enterprise resource planning	>100	Gigabytes–terabytes
Data warehousing	>100	Terabytes–petabytes

Customer

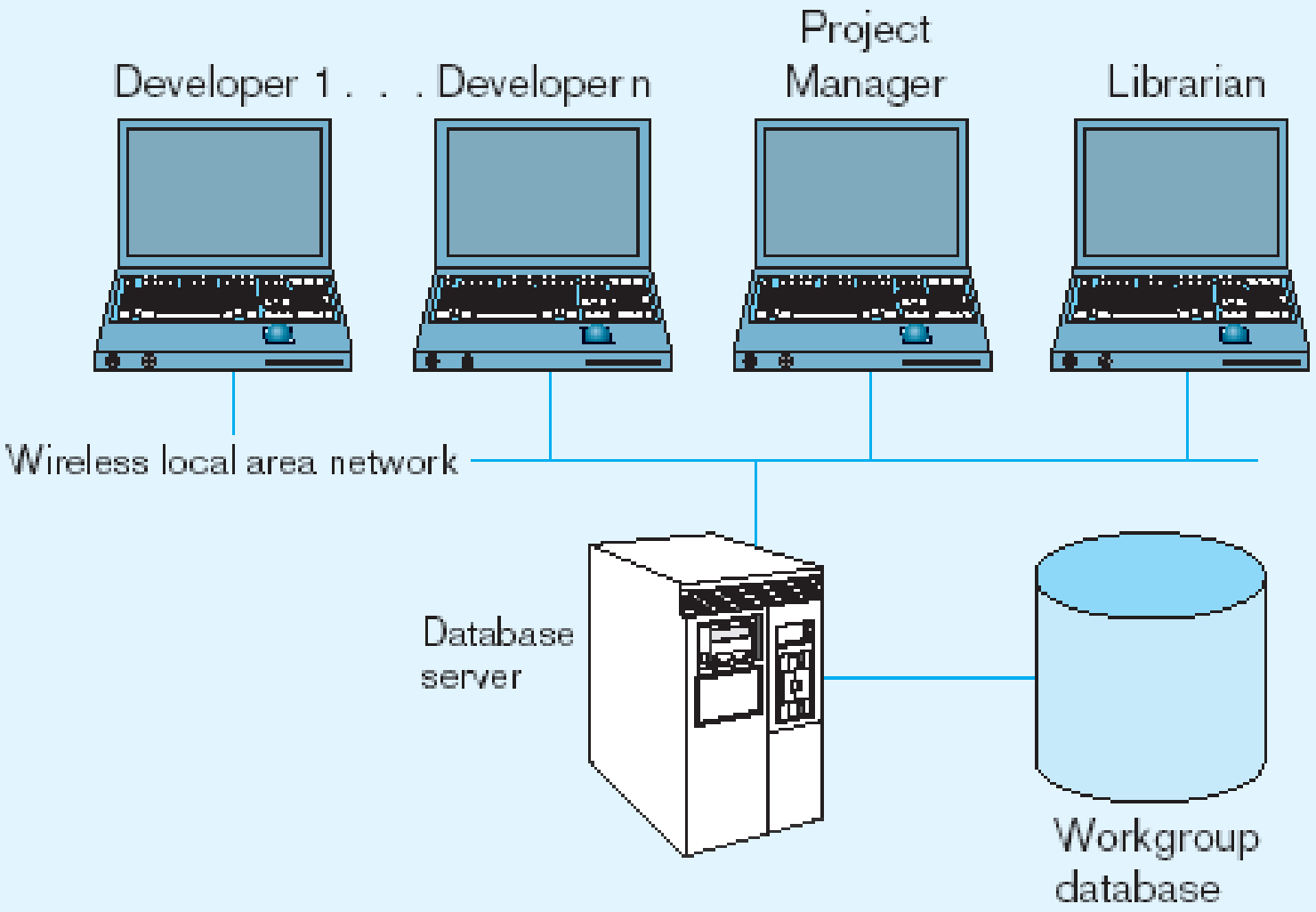
Customer Name: **Multi Media, Inc.**	
Address: **1000 River Road**	
City: **San Antonio**	
State: **TX**	
Zip: **76235**	
Phone: **(219) 864-2000**	
Next Contact Date: **10/17/2006**	Time: **10:30 AM**

Contact History for Customer

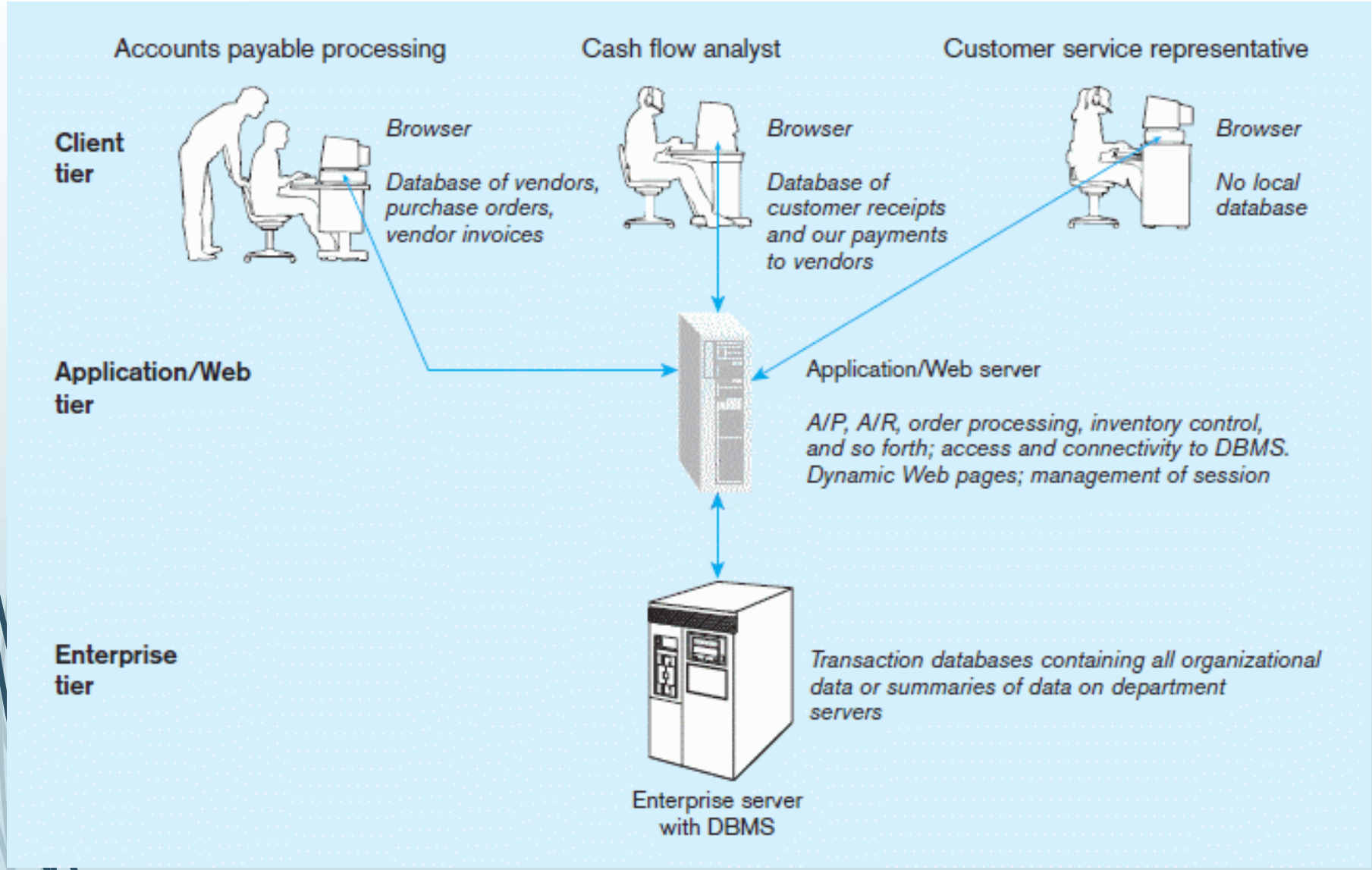
Date	Time	Contact	Comments
08/04/2006	10:00 AM	Roberts	Review proposal
08/19/2006	08:00 AM	Roberts	Revise schedule
09/10/2006	09:00 AM	Pearson	Sign contract
09/21/2006	02:00 PM	Roberts	Follow up

Typical data from a personal database

Workgroup database with wireless local area network



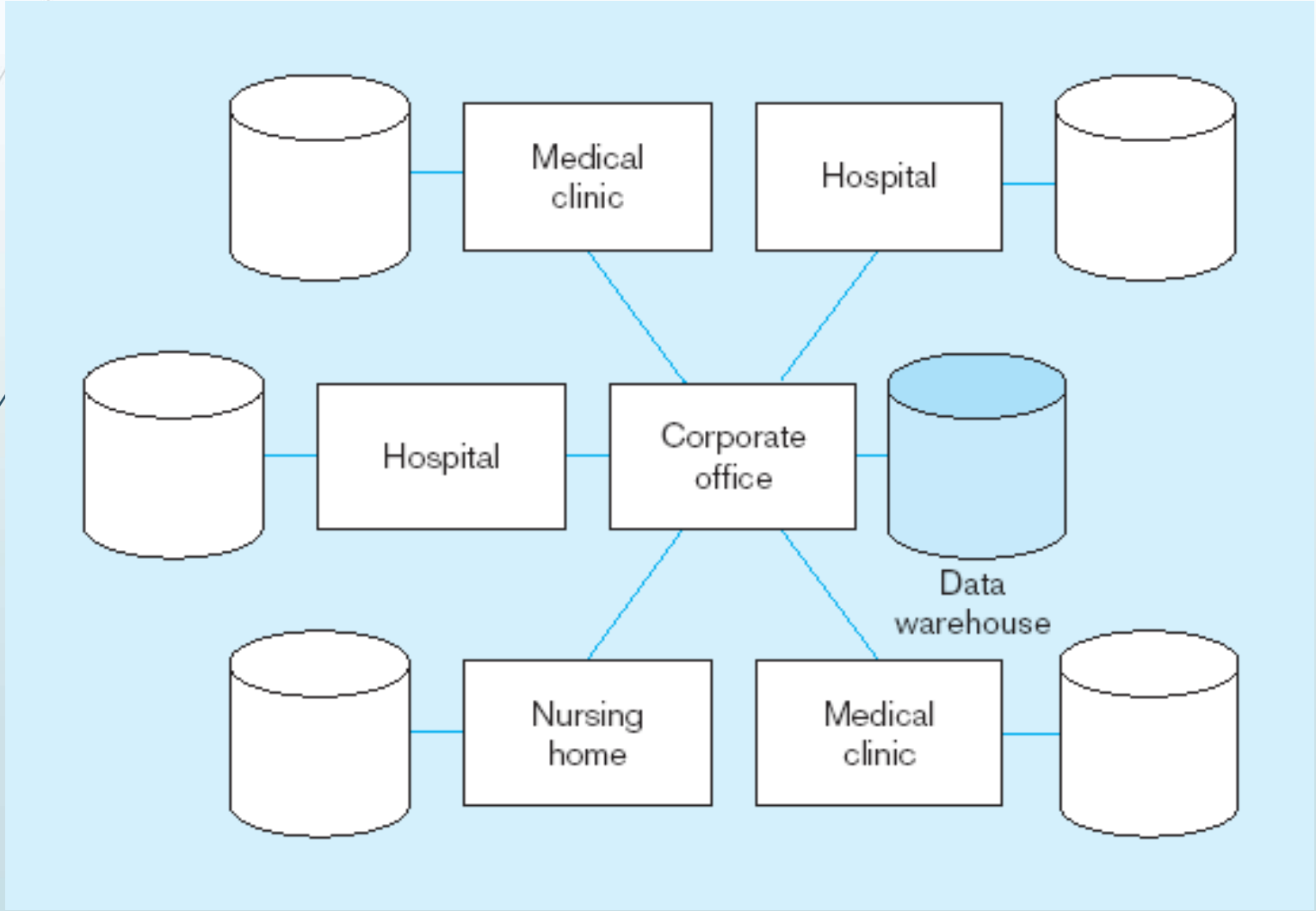
Three-tiered client/server database architecture



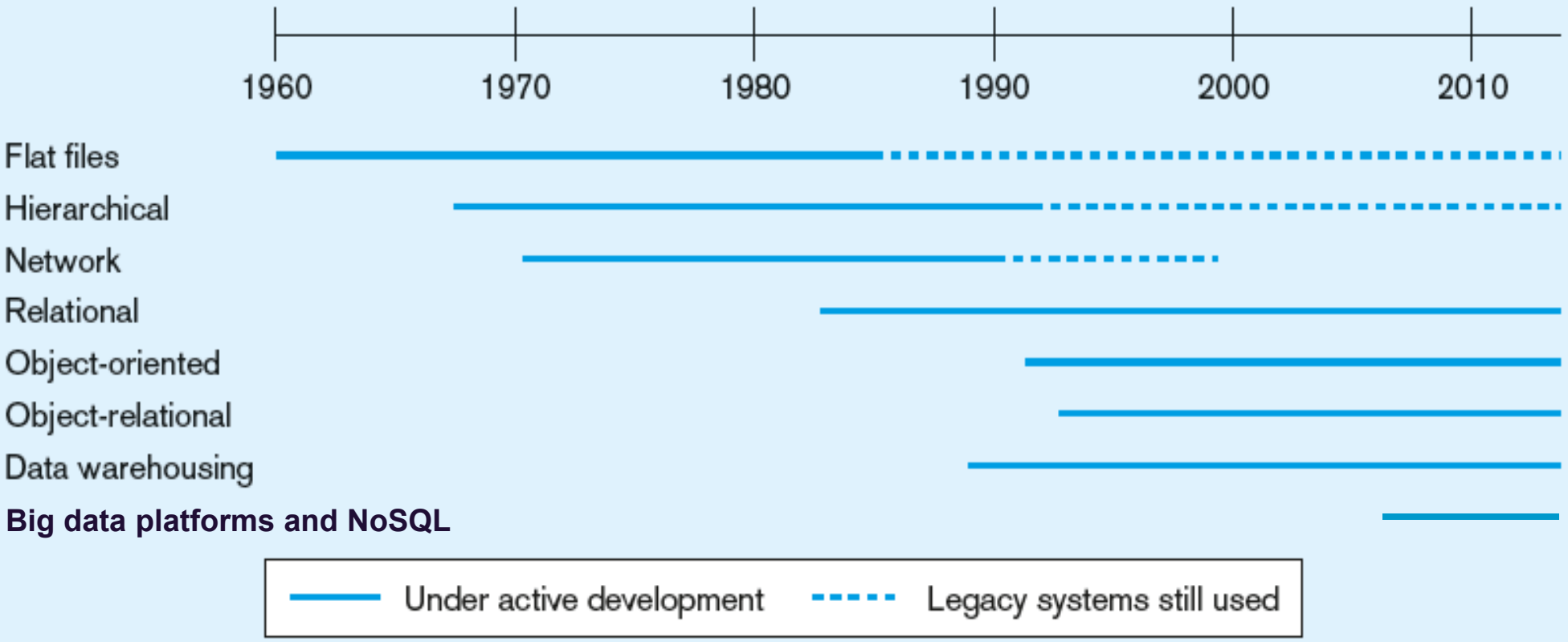
Enterprise Database Applications

- 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
- Big Data and Business Analytics
  - Massive amounts of real-time and multimedia data processed by computer clusters in data center for decision support and business forecasting

An enterprise data warehouse



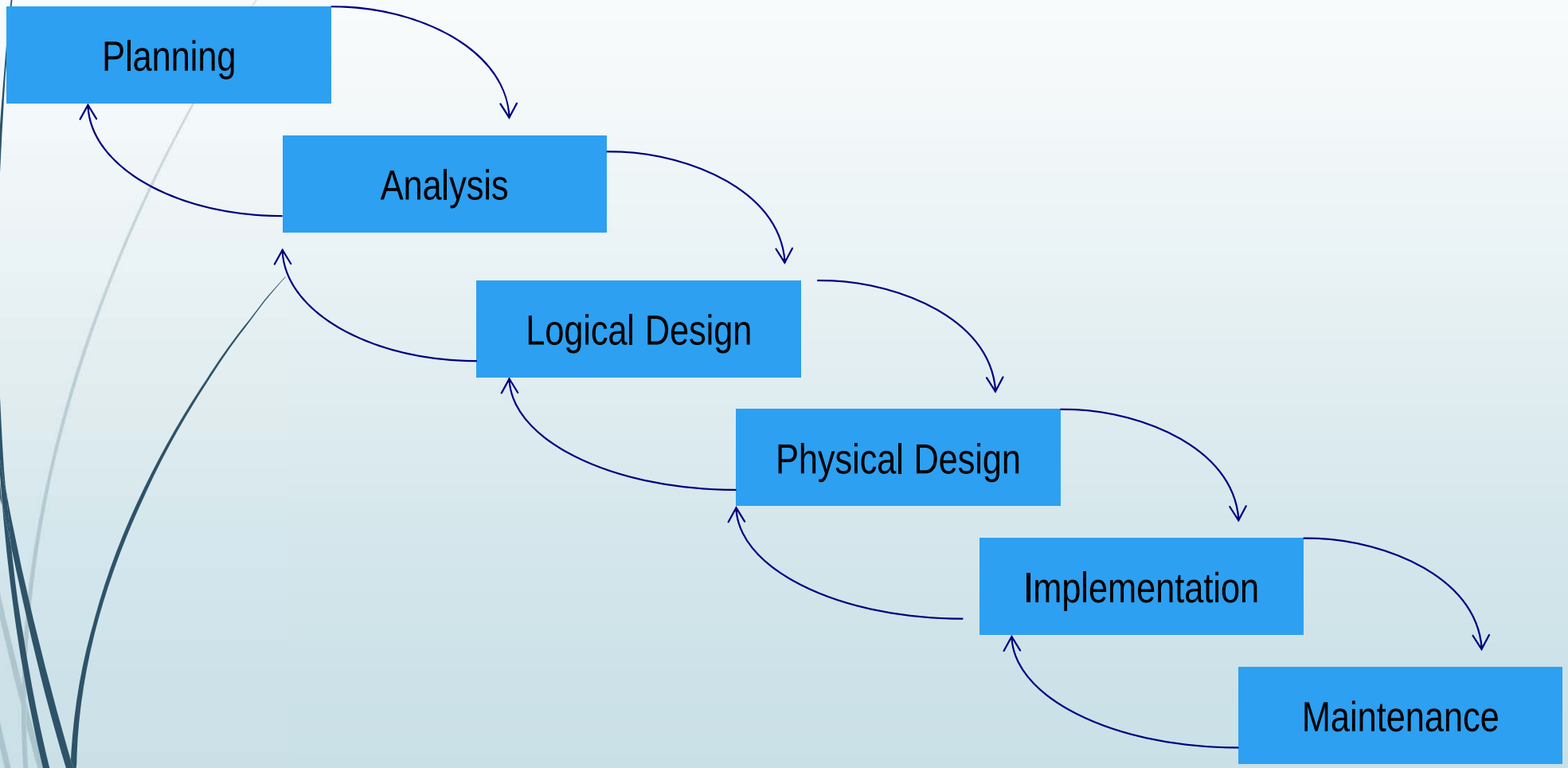
Evolution of DB Systems



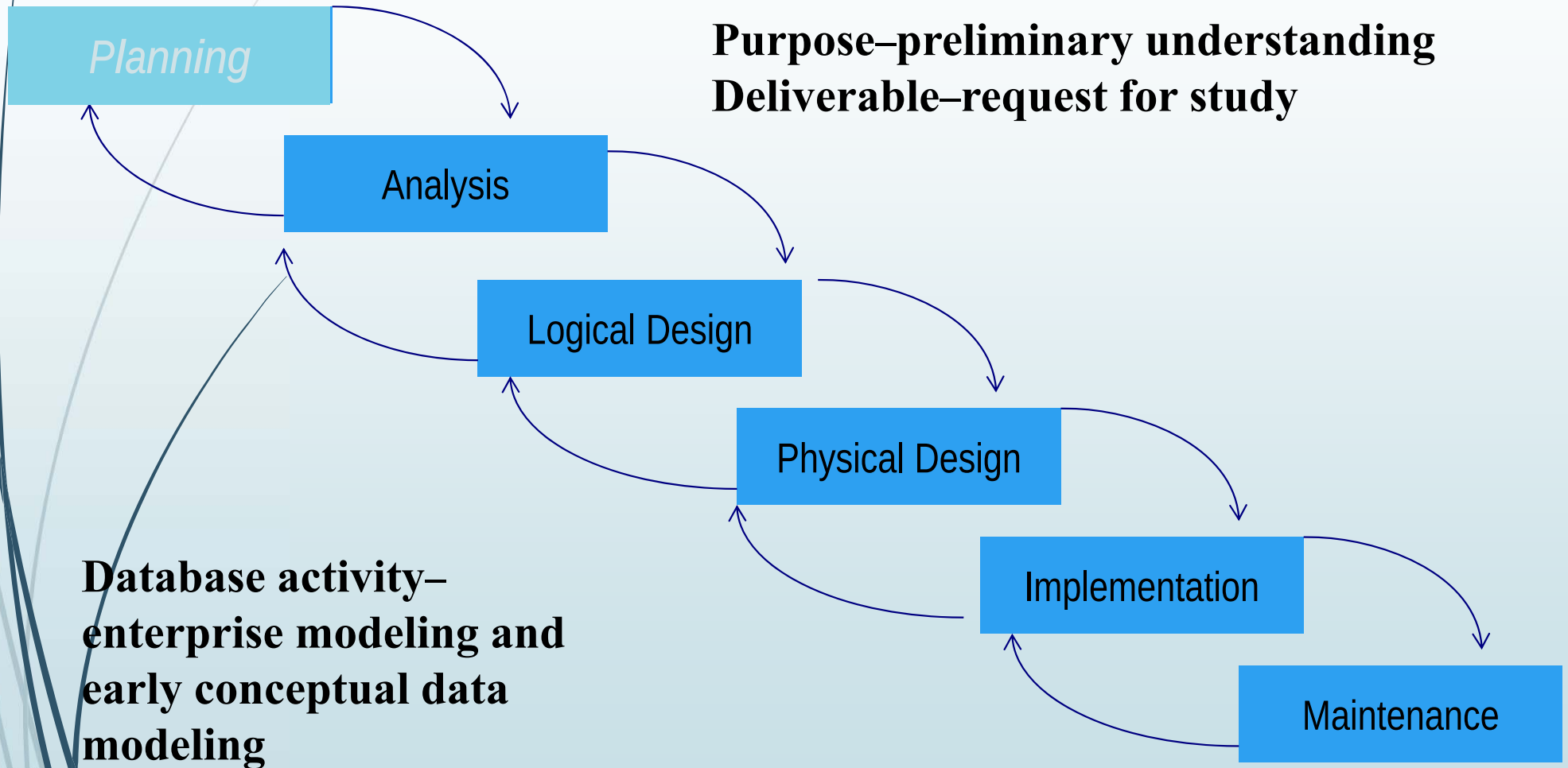
Two Major Approaches to Database and System Development

- 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



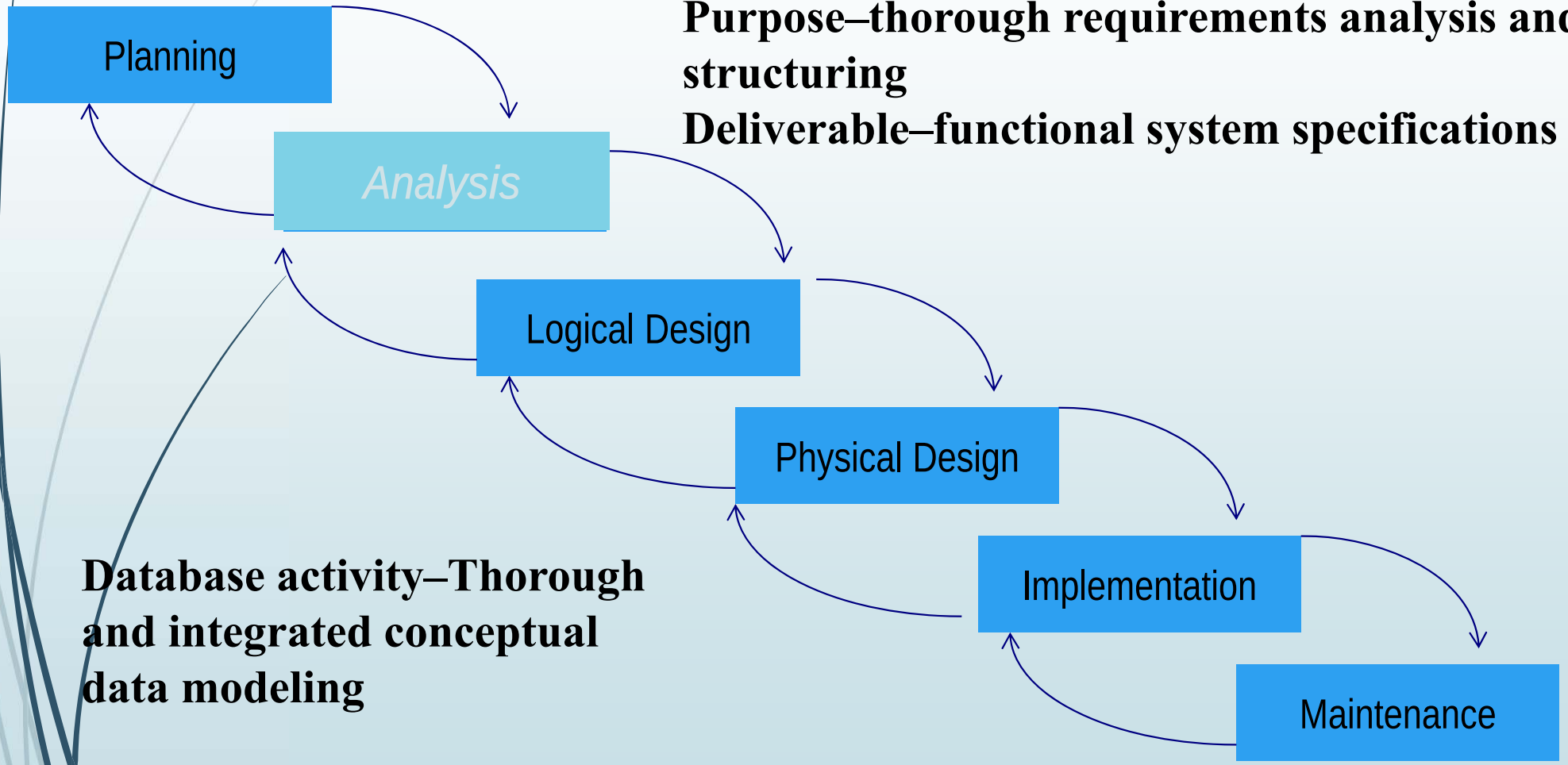
Systems Development Life Cycle



Purpose–preliminary understanding
Deliverable–request for study

Database activity–enterprise modeling and early conceptual data modeling

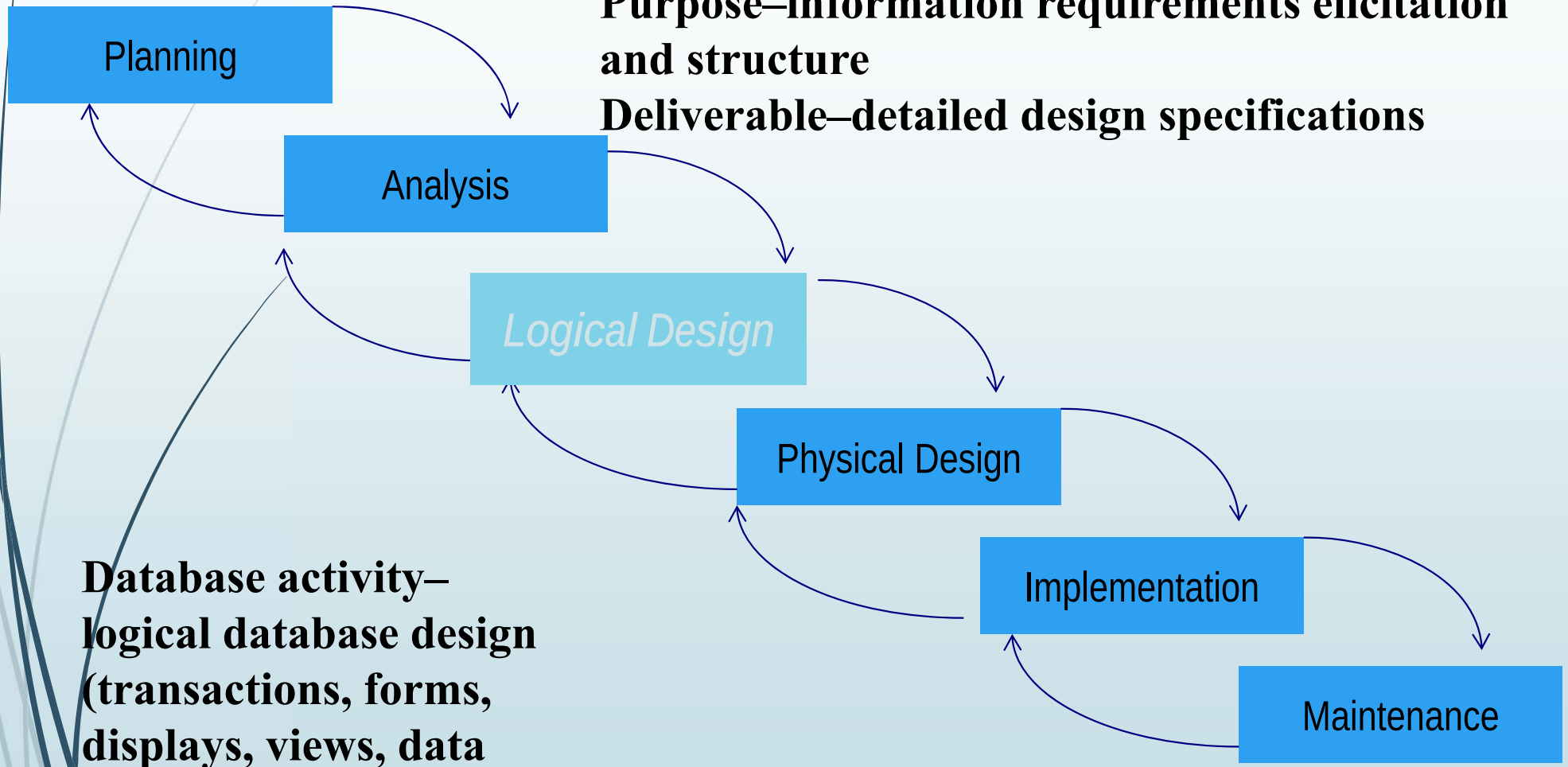
Systems Development Life Cycle



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

Database activity–Thorough and integrated conceptual data modeling

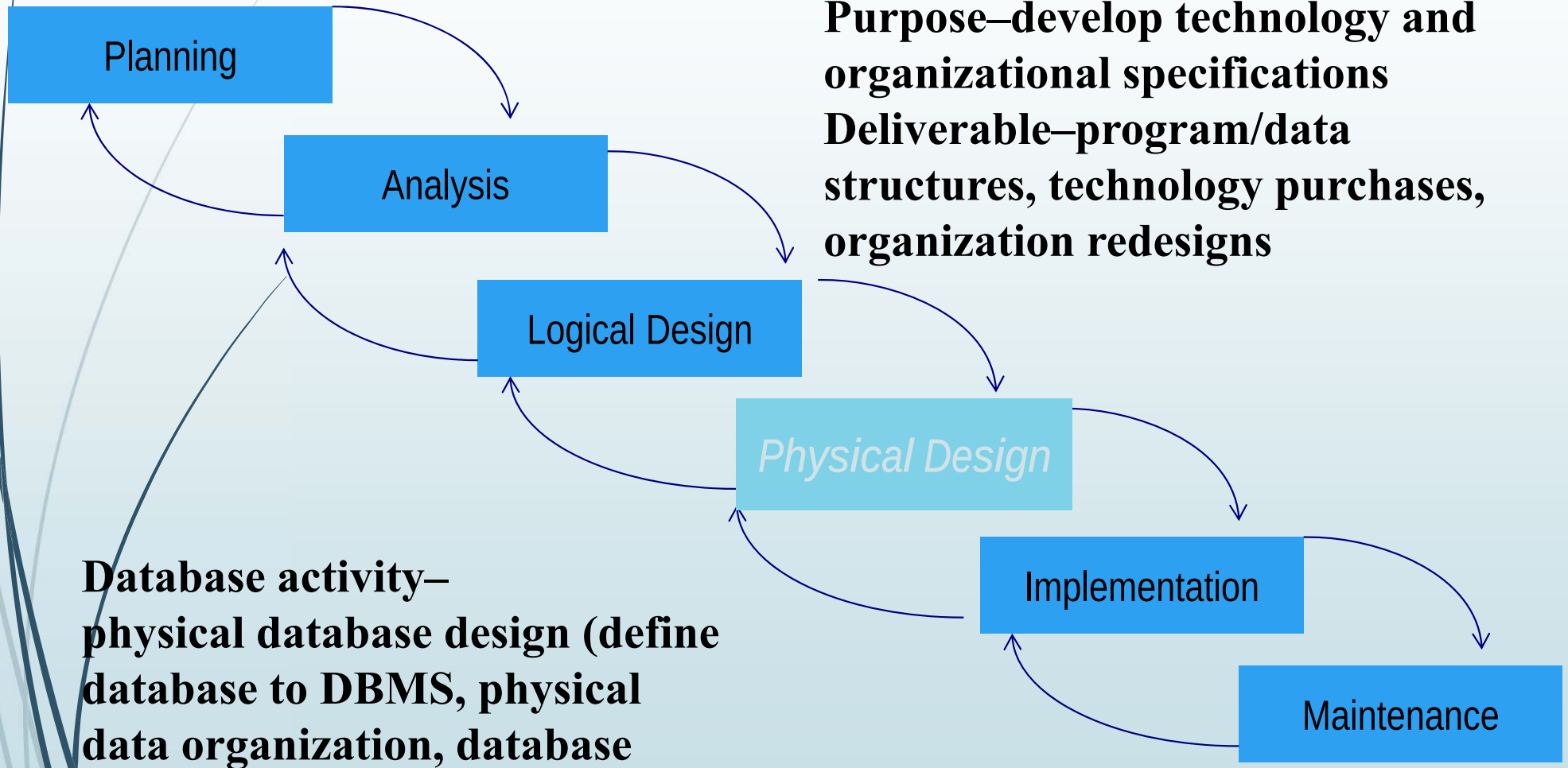
Systems Development Life Cycle



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

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

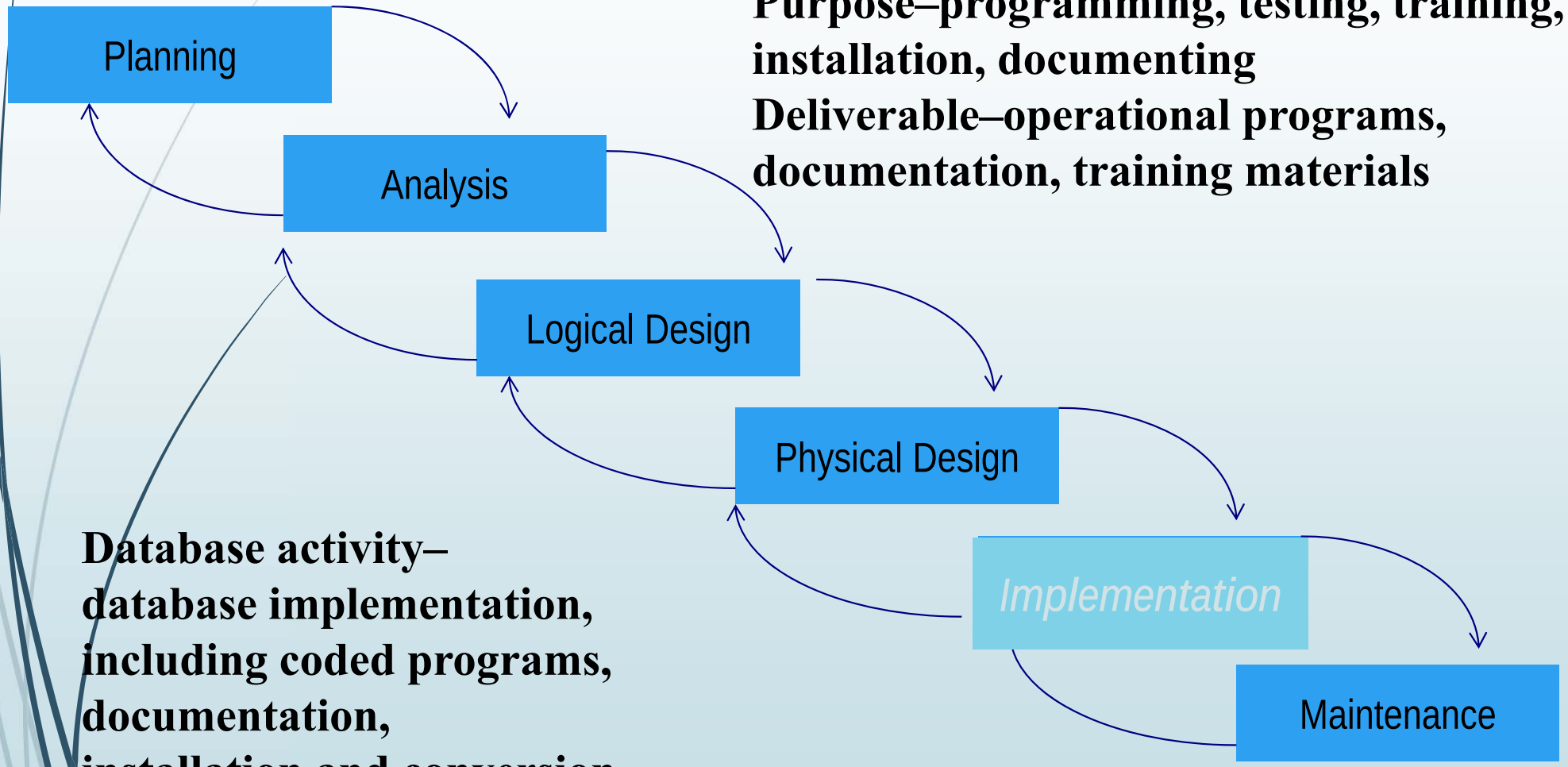
Systems Development Life Cycle



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)

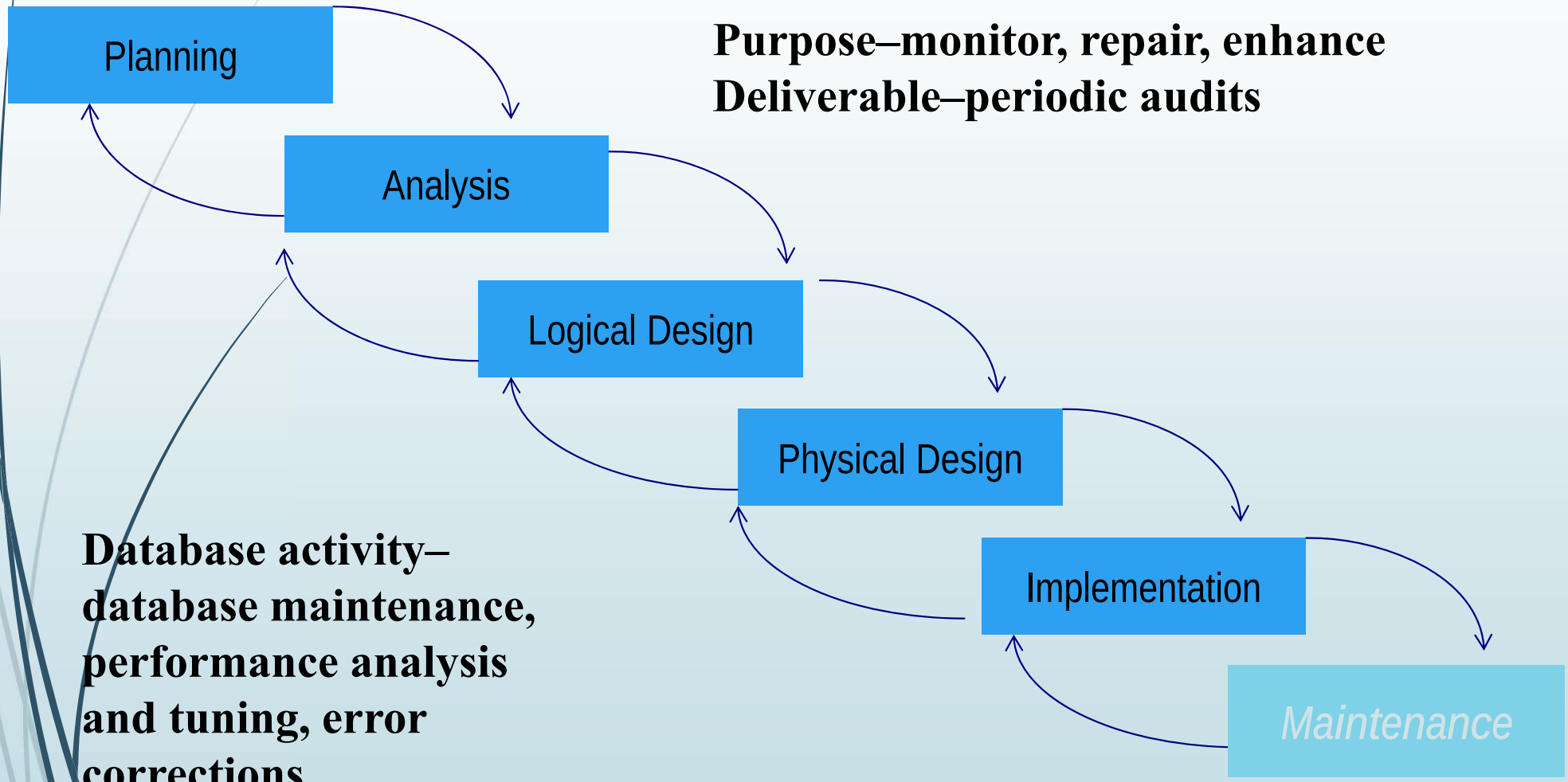
Systems Development Life Cycle



Purpose–programming, testing, training, installation, documenting
Deliverable–operational programs, documentation, training materials

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

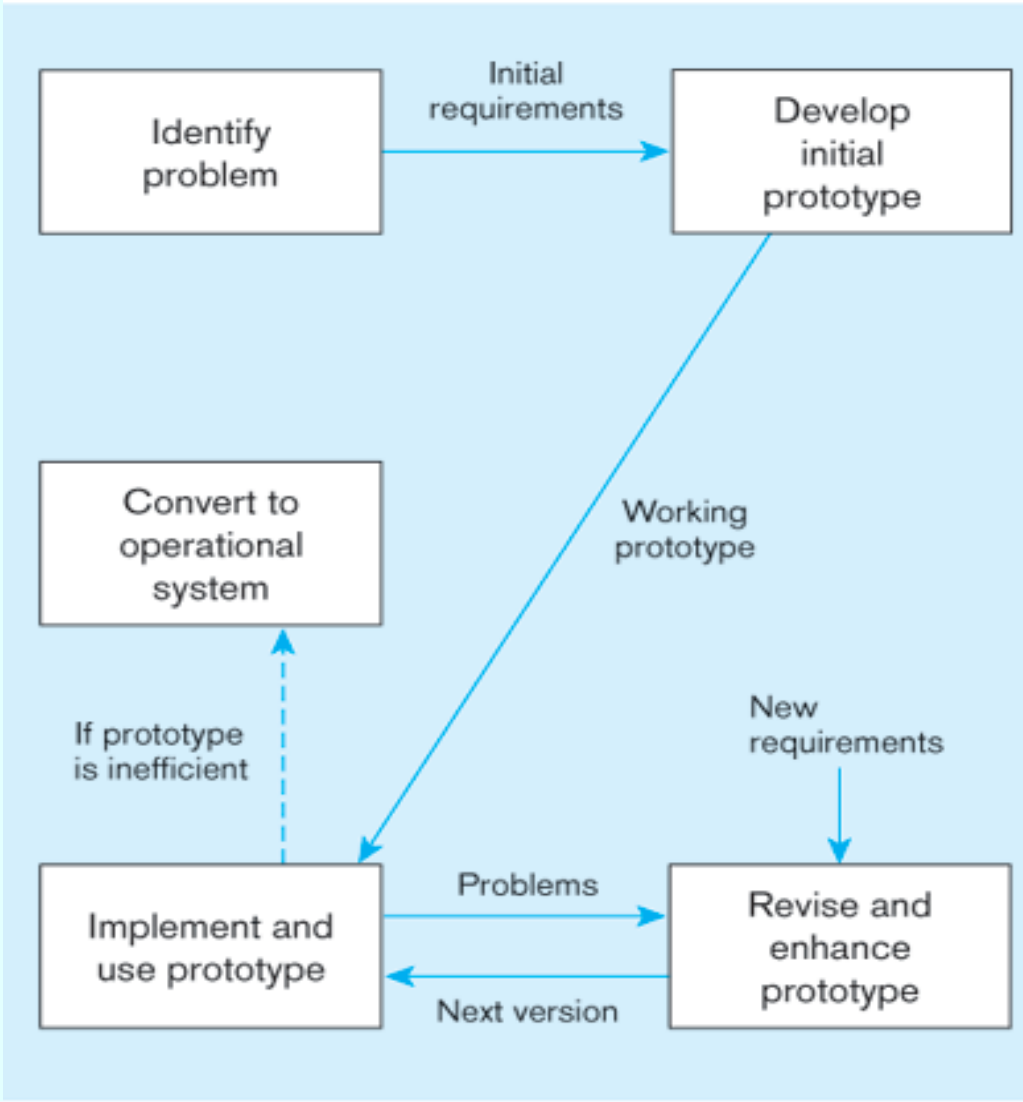
Systems Development Life Cycle



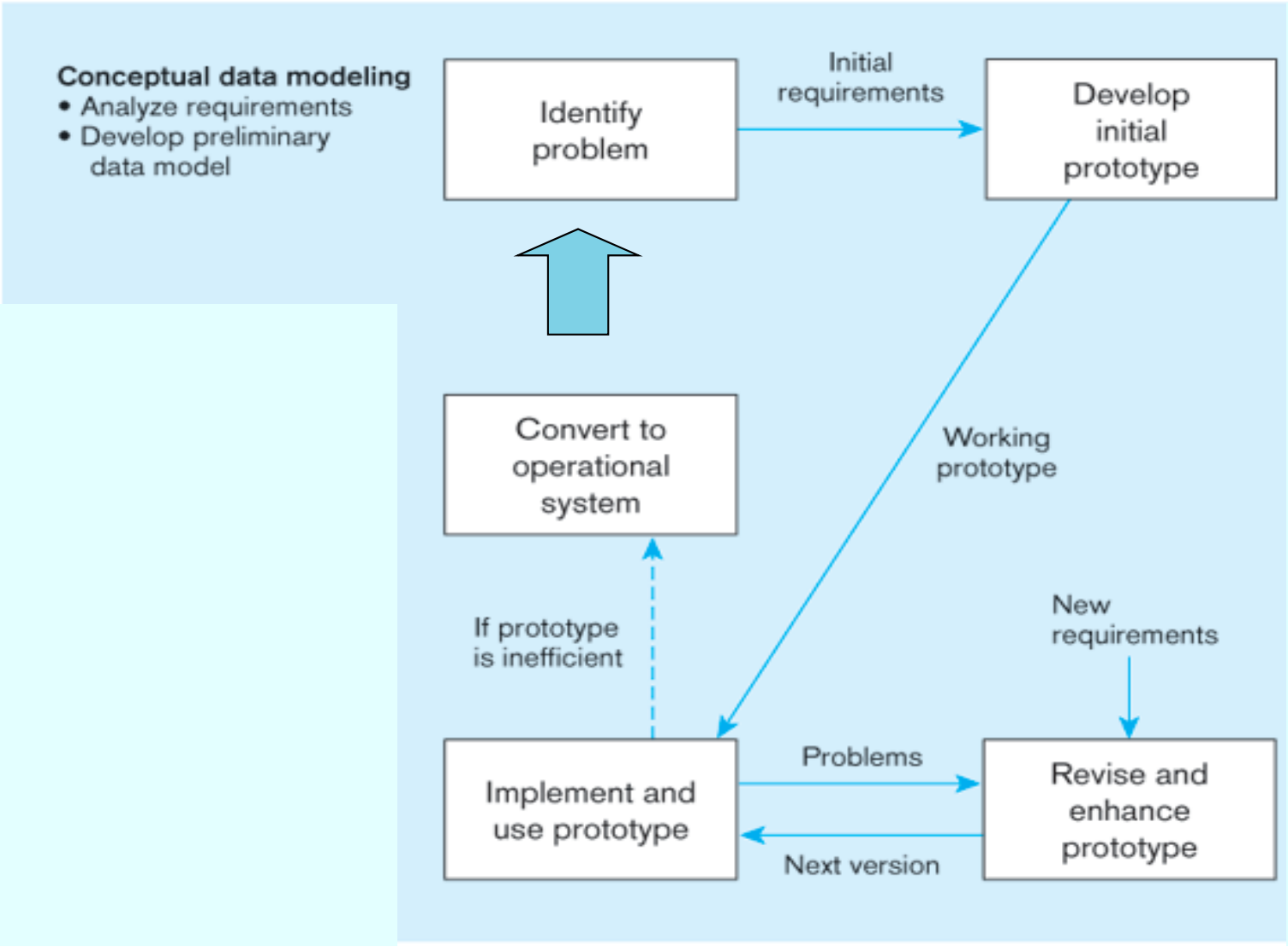
Purpose–monitor, repair, enhance
Deliverable–periodic audits

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

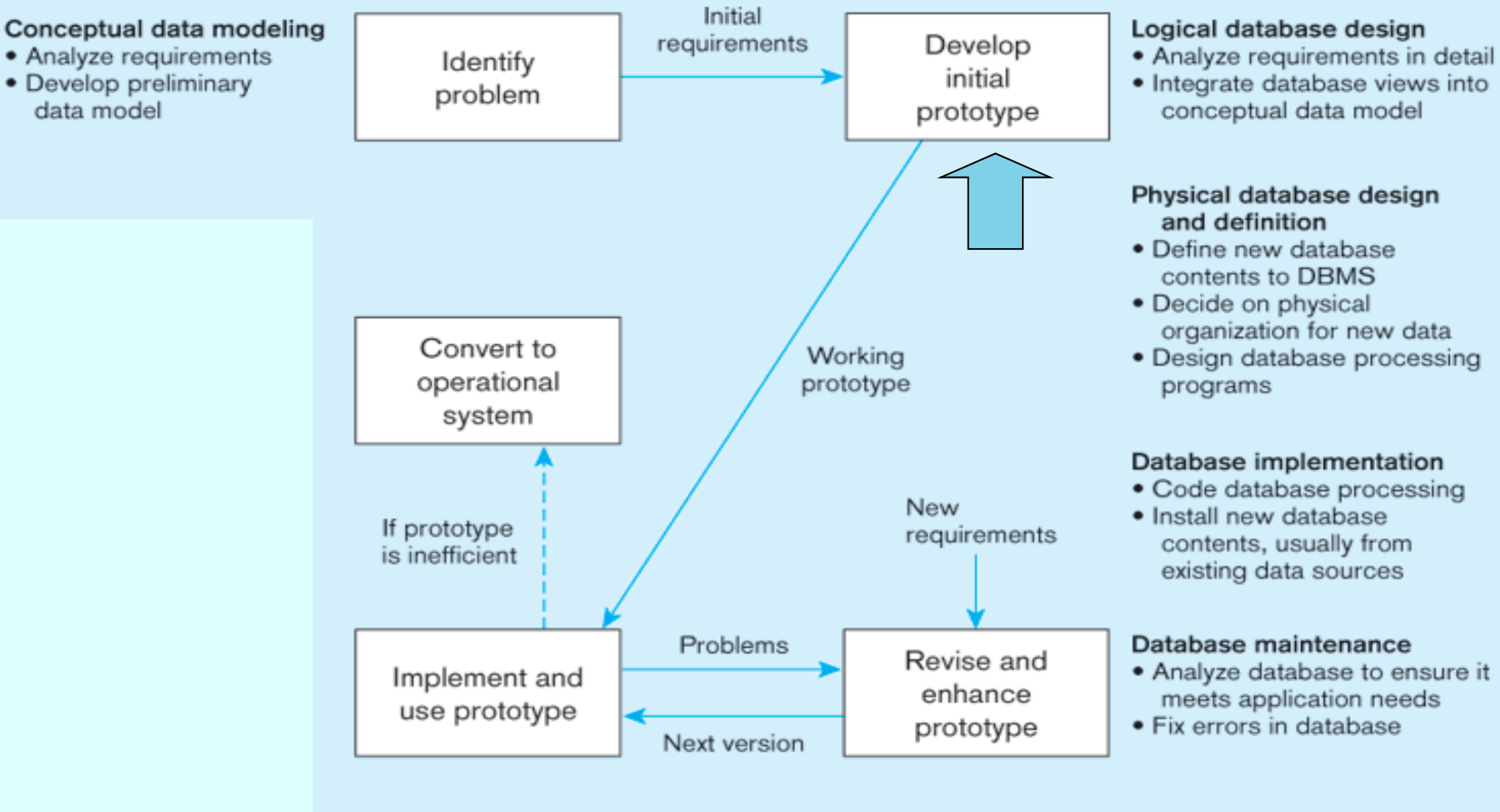
Prototyping Database Methodology



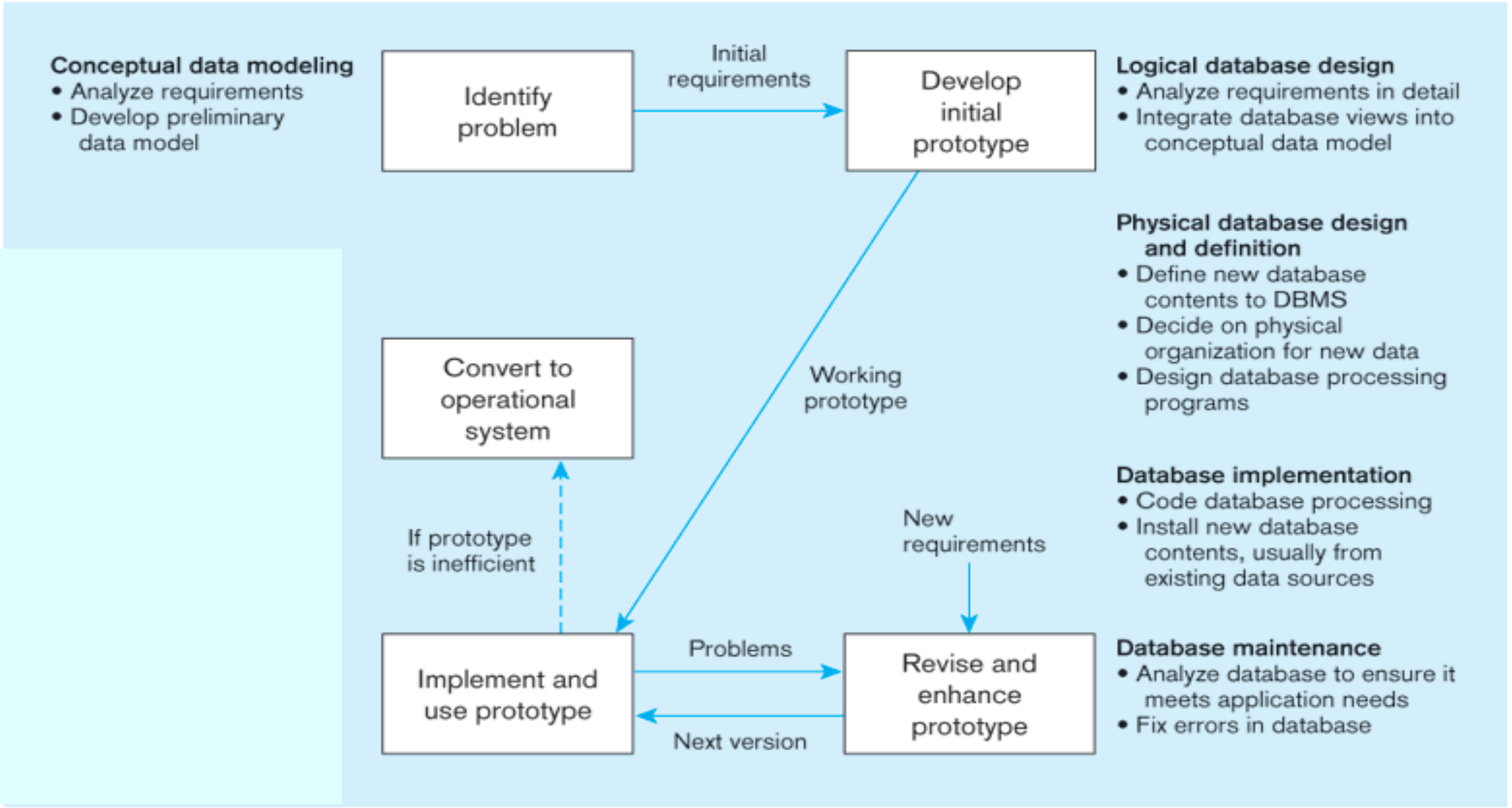
Prototyping Database Methodology



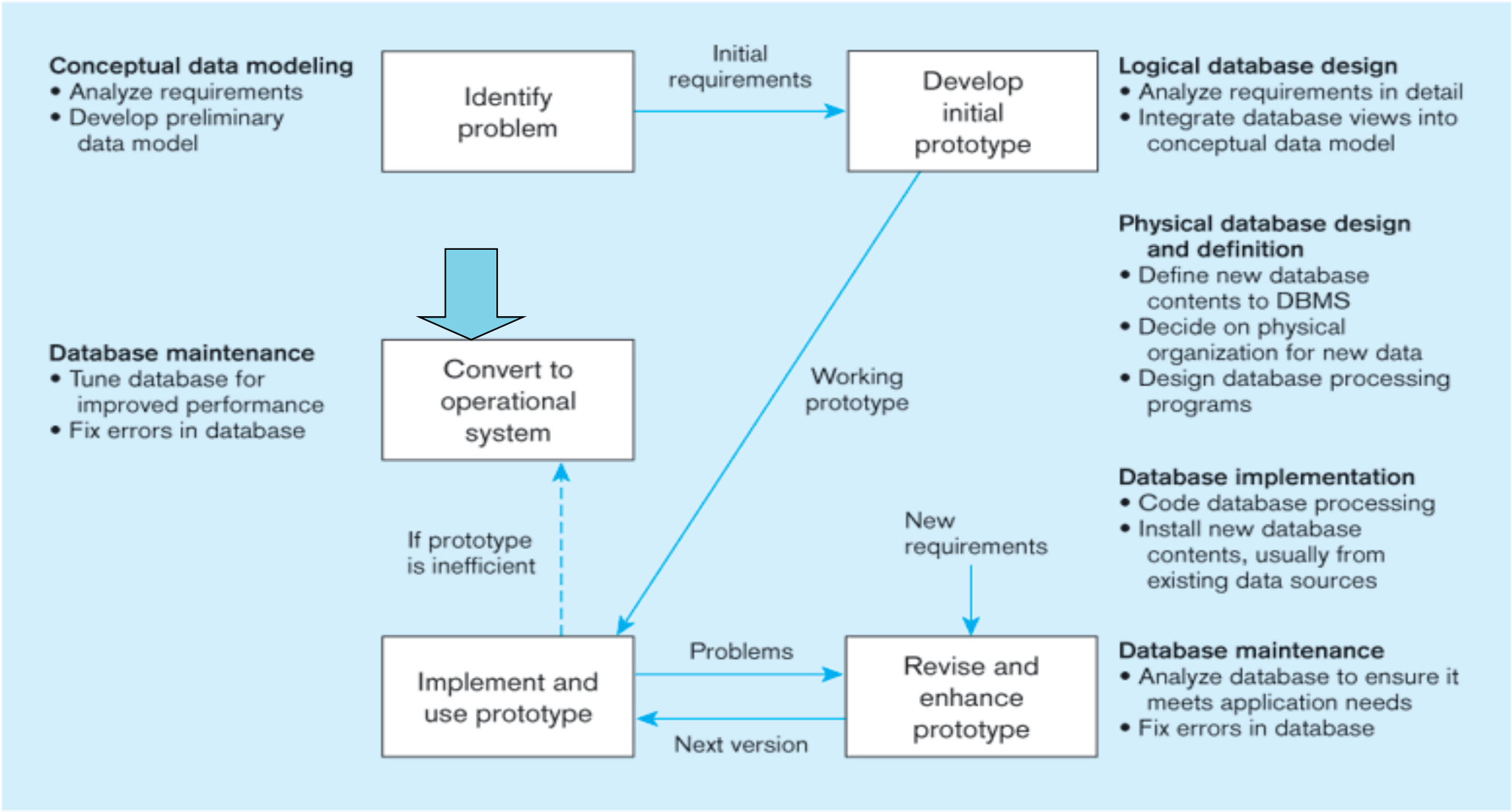
Prototyping Database Methodology



Prototyping Database Methodology



Prototyping Database Methodology



Focus of this course

- Relational model
  - Conceptual design (E-R model)
  - Logical design
  - Physical design
  - Implementation / SQL
  - MySQL
- Data warehousing
  - Architecture
  - Star schema
  - Google BigQuery
- NoSQL model
  - Document model
  - MongoDB