

COMP5111

Database System and Management

Topic 1: Database Environment

**Dr Edward Hung
Department of Computing
Hong Kong Polytechnic University
Sep 7, 2005**



Outline

- Some basic terms
- Limitations of file processing systems
- Database management system
 - Approach
 - Advantages
 - Costs and risks

Terms

- Database: organized collection of logically related data
- Data: stored representations of meaningful objects and events
 - Structured: numbers, text, dates
 - Unstructured: images, video, documents

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

What do these data mean ??

Context helps users understand data !!

Class Roster

Course: MGT 500 Semester: Spring 200X
Business Policy

Section: 2

<u>Name</u>	<u>ID</u>	<u>Major</u>	<u>GPA</u>
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

- Information: data processed to increase knowledge in the person using the data



Class Roster

Course: MGT 500
Business Policy

Semester: Spring 200X

Section: 2

<u>Name</u>	<u>ID</u>	<u>Major</u>	<u>GPA</u>
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

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

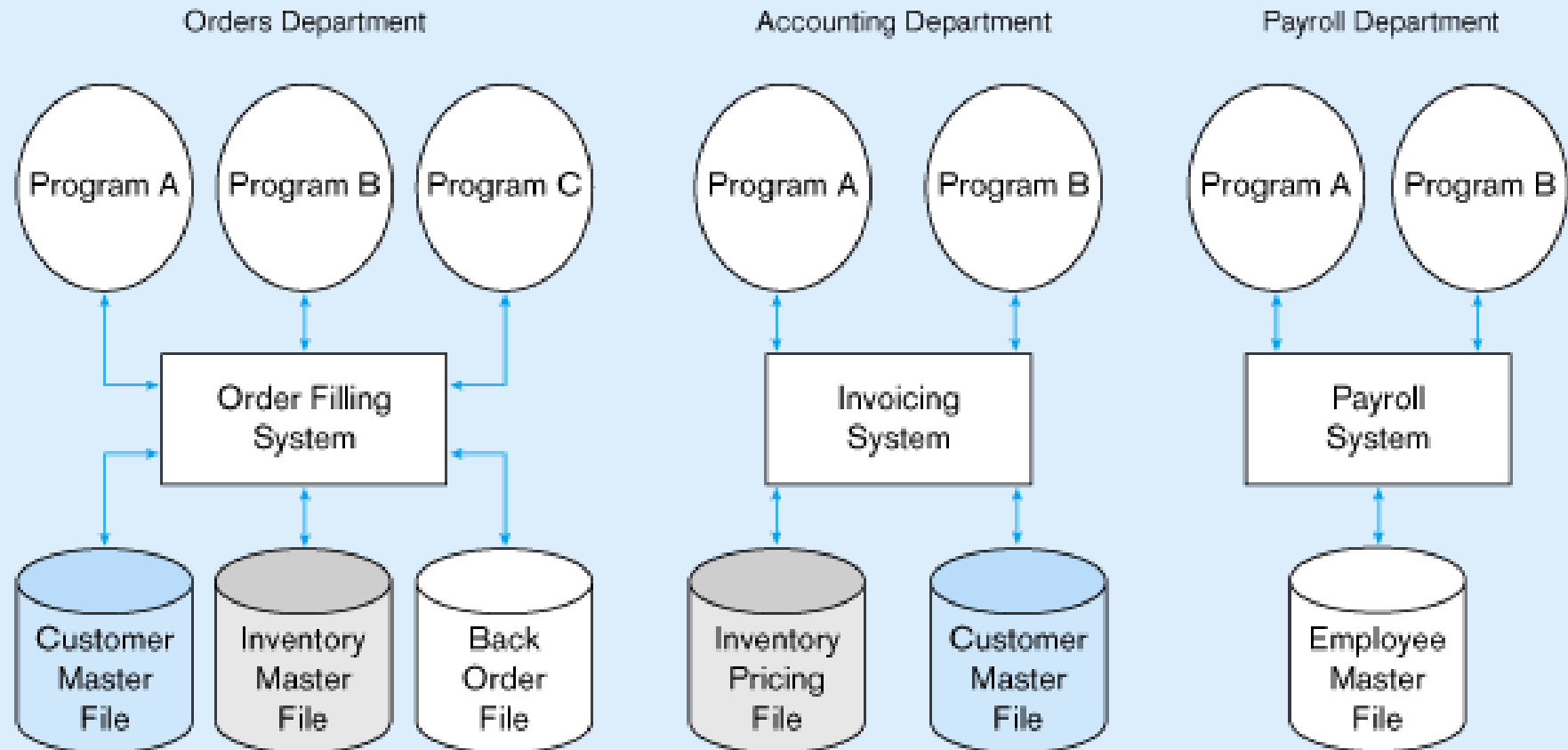
Table 1-1 Example Metadata for Class Roster

<i>Data Item</i>		<i>Value</i>				
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

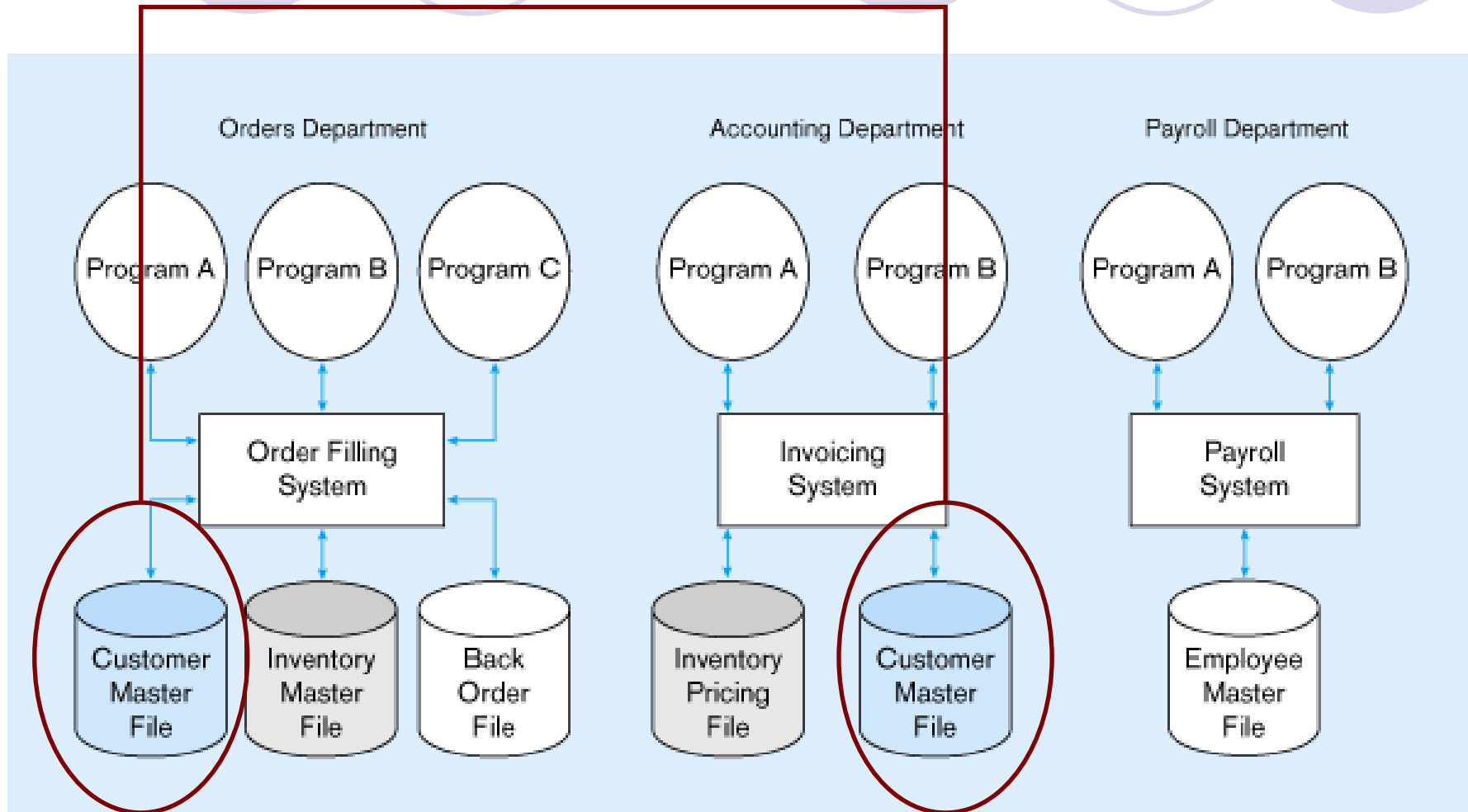
Disadvantages of File Processing System

Three file processing systems at Pine Valley Furniture Company



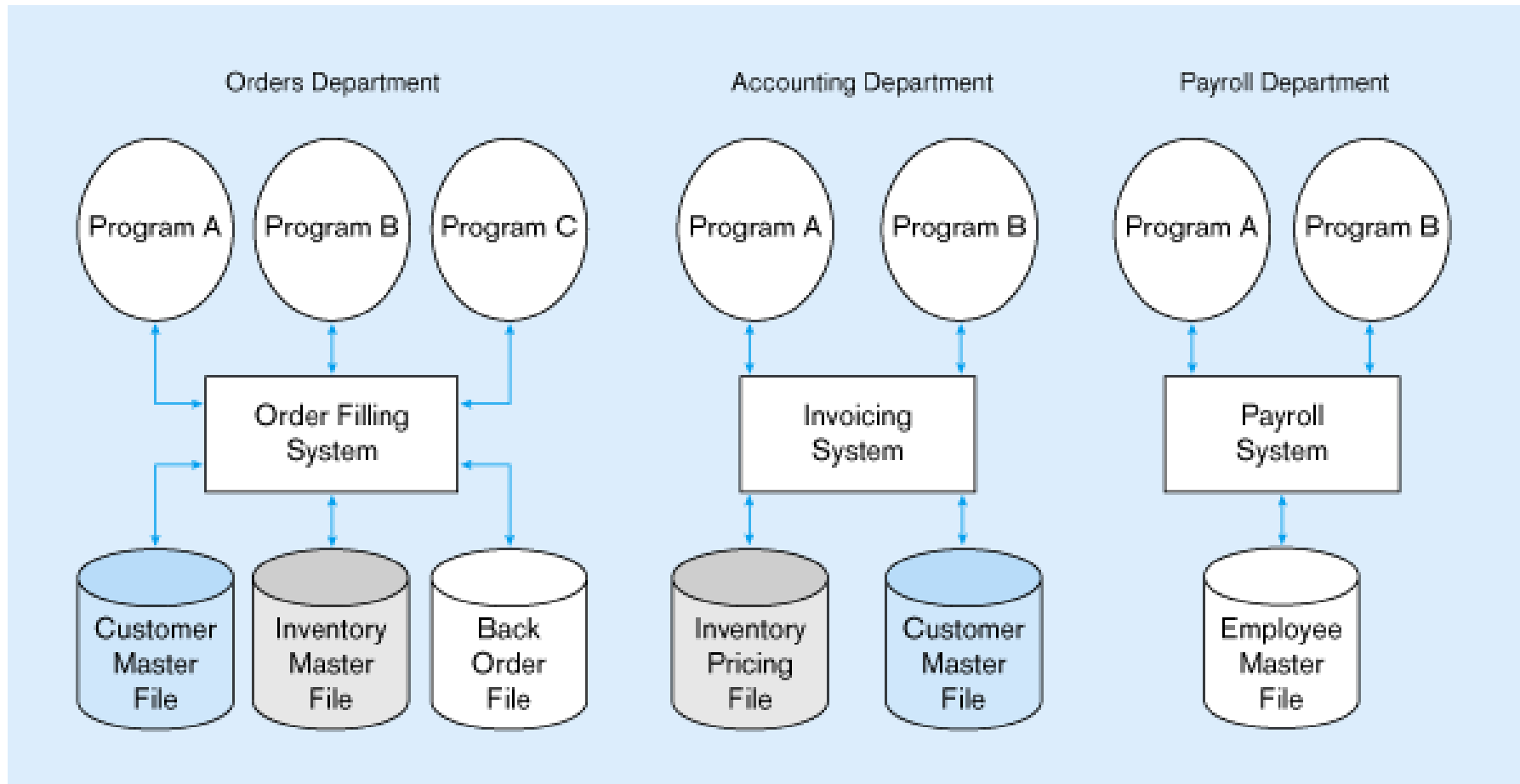
Disadvantages of File Processing System

Duplicate Data



Discussion (5 mins)

1. Form a group of 4-6 people with your neighbors (side, before and behind) and discuss the disadvantages
2. Each group will be asked to give one disadvantage not mentioned by previous groups



Disadvantages of File Processing System

● Program-Data Dependence

- All programs maintain their own data and metadata for each file they use; *e.g., when the file format is changed, all programs that access the file need to be changed*
- Programmers must design their own file formats
- Each application program must have its own processing routines for reading, inserting, updating and deleting data

→ Lengthy Development Times

- Lack of coordination and no centralized control of data

→ Limited Data Sharing

Disadvantages of File Processing System

● **Duplication of Data**

- Different systems/programs have separate copies of the same data
- Waste of space
- More maintenance headaches
- When data changes in one file, it could cause inconsistencies
- Loss of *data integrity*
 - Same data item has different names in different files
 - Different data items have same names in different files

● **Excessive Program Maintenance**

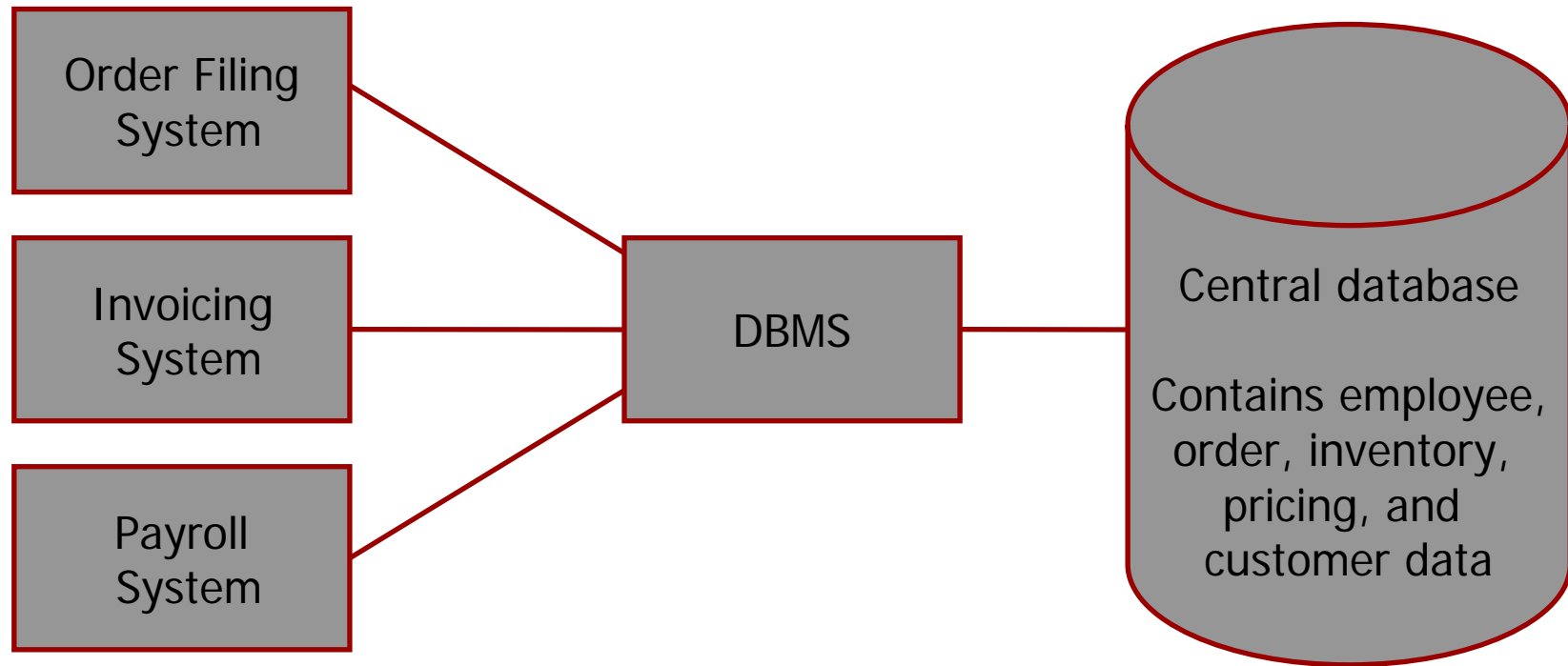
- 80% of information systems budget

SOLUTION:

Database Management System (DBMS)

- A software system that is used to create, maintain, and provide controlled access to user databases
 - Central repository of shared data
 - Data is managed by a controlling agent
 - Stored in a standardized, convenient form
 - *DBMS manages data resources like an operating system manages hardware resources*

Database Management System





Elements of the Database Approach

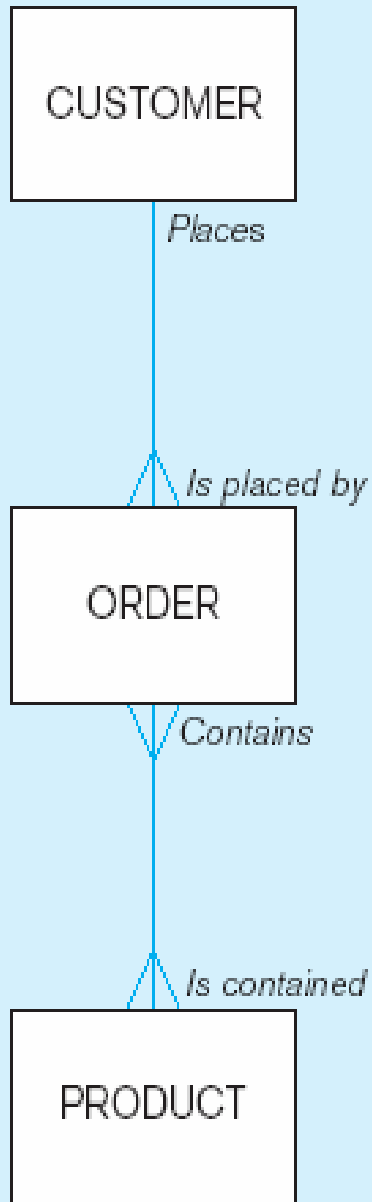
- Enterprise Data Model
 - Graphical model showing high-level entities and relationships for the organization
- Relational Databases
 - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships
- Use of Internet Technology
 - Networks and 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



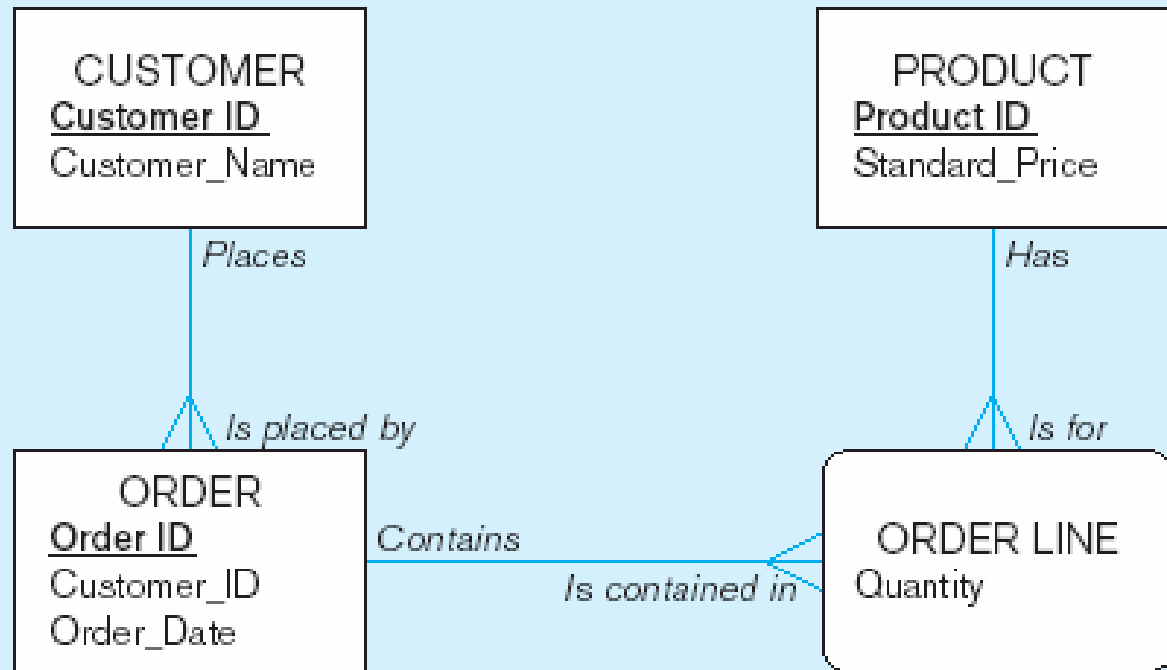
Enterprise Data Model

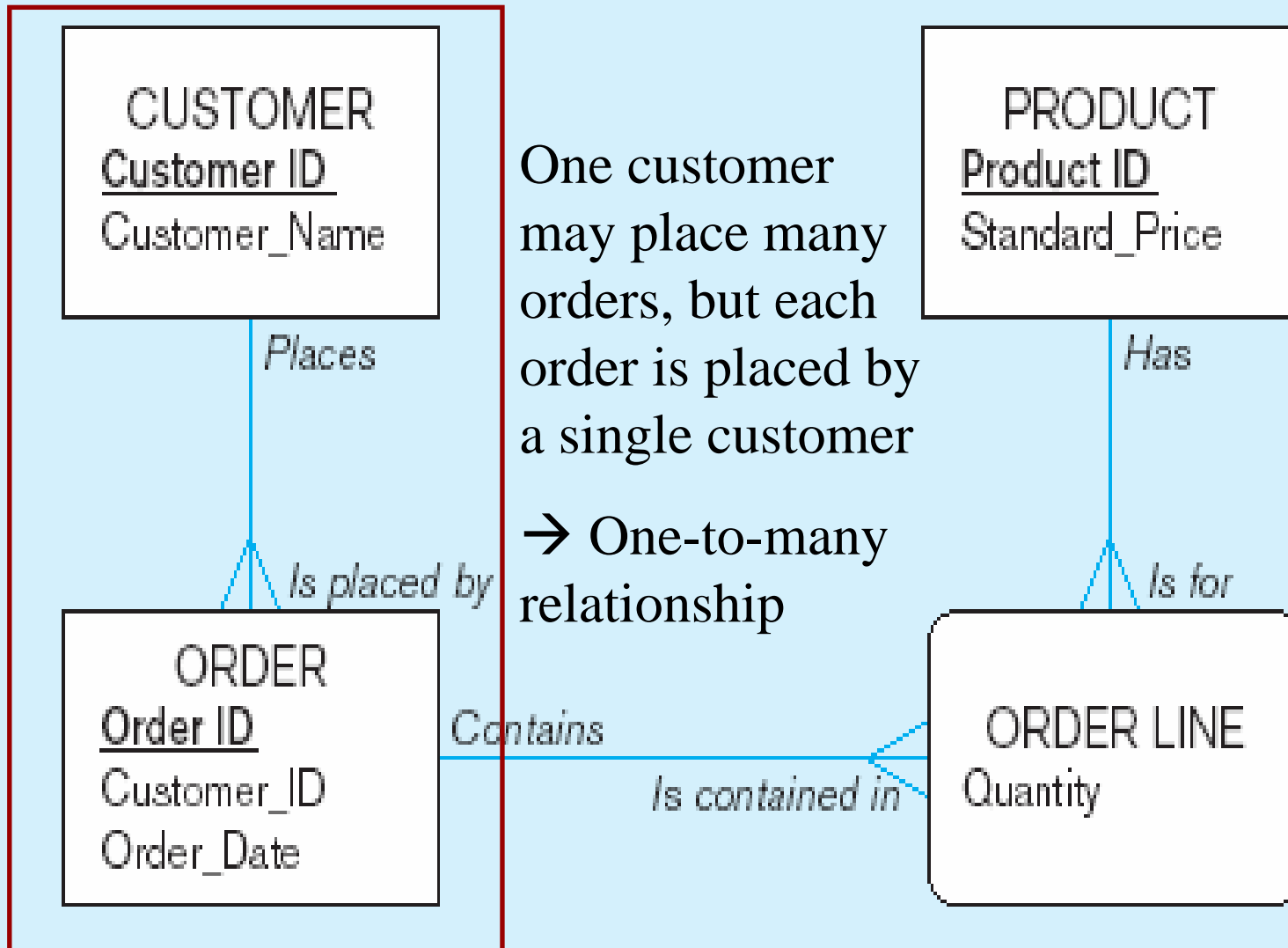
- If we are told that something about our **customers, orders and products ...**
 - One customer may place many orders, but each order is placed by a single customer
 - One order has many order lines; each order line is associated with a single order
 - One product can be in many order lines, each order line refers to a single product
- How to represent this in an easy-to-understand way?

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



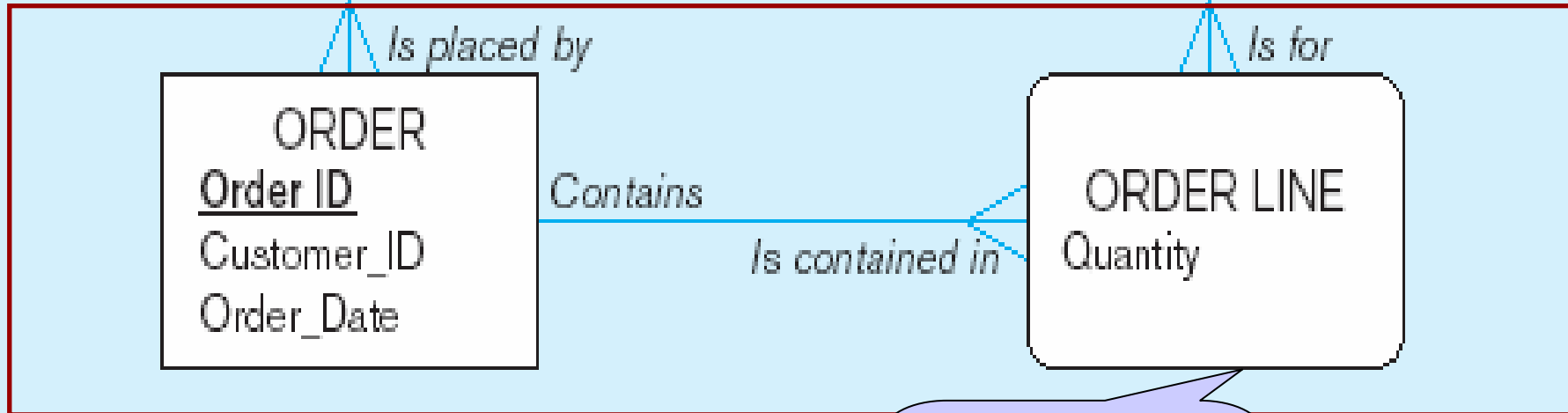
One order has many order lines; each order line is associated with a single order



→ One-to-many relationship

Places

Has

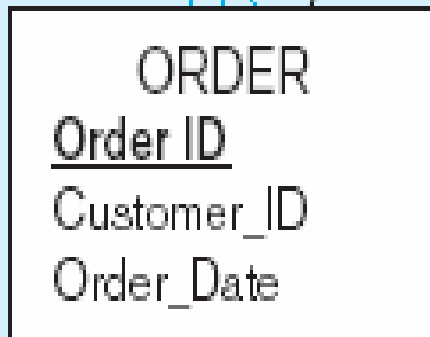


Details about each product sold on an order

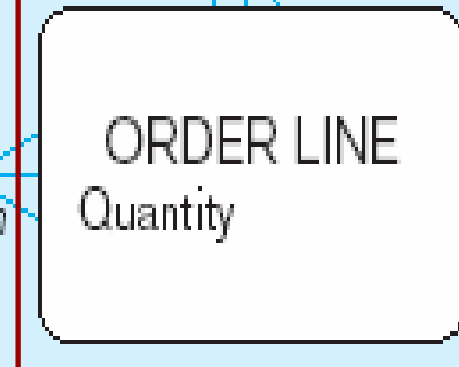


One product can be in many order lines, each order line refers to a single product

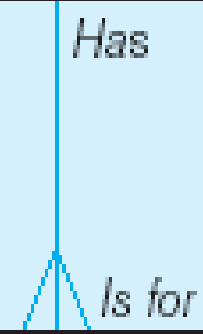
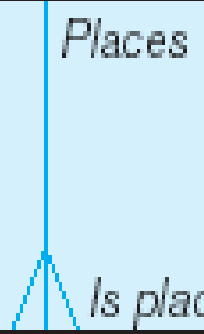
→ One-to-many relationship

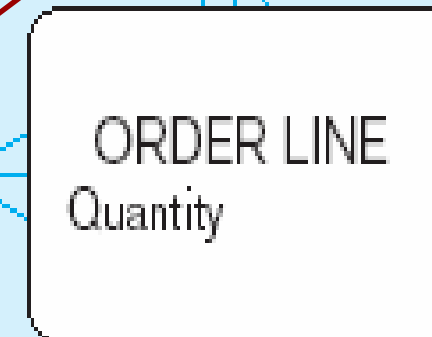
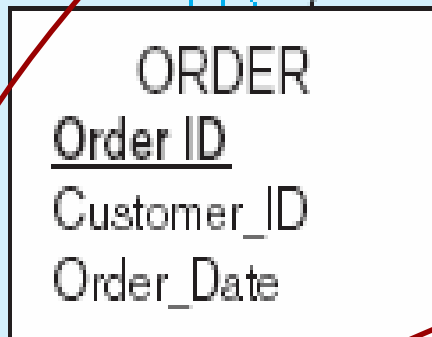


Contains



Is contained in





Places

Has

Is placed by

Is for

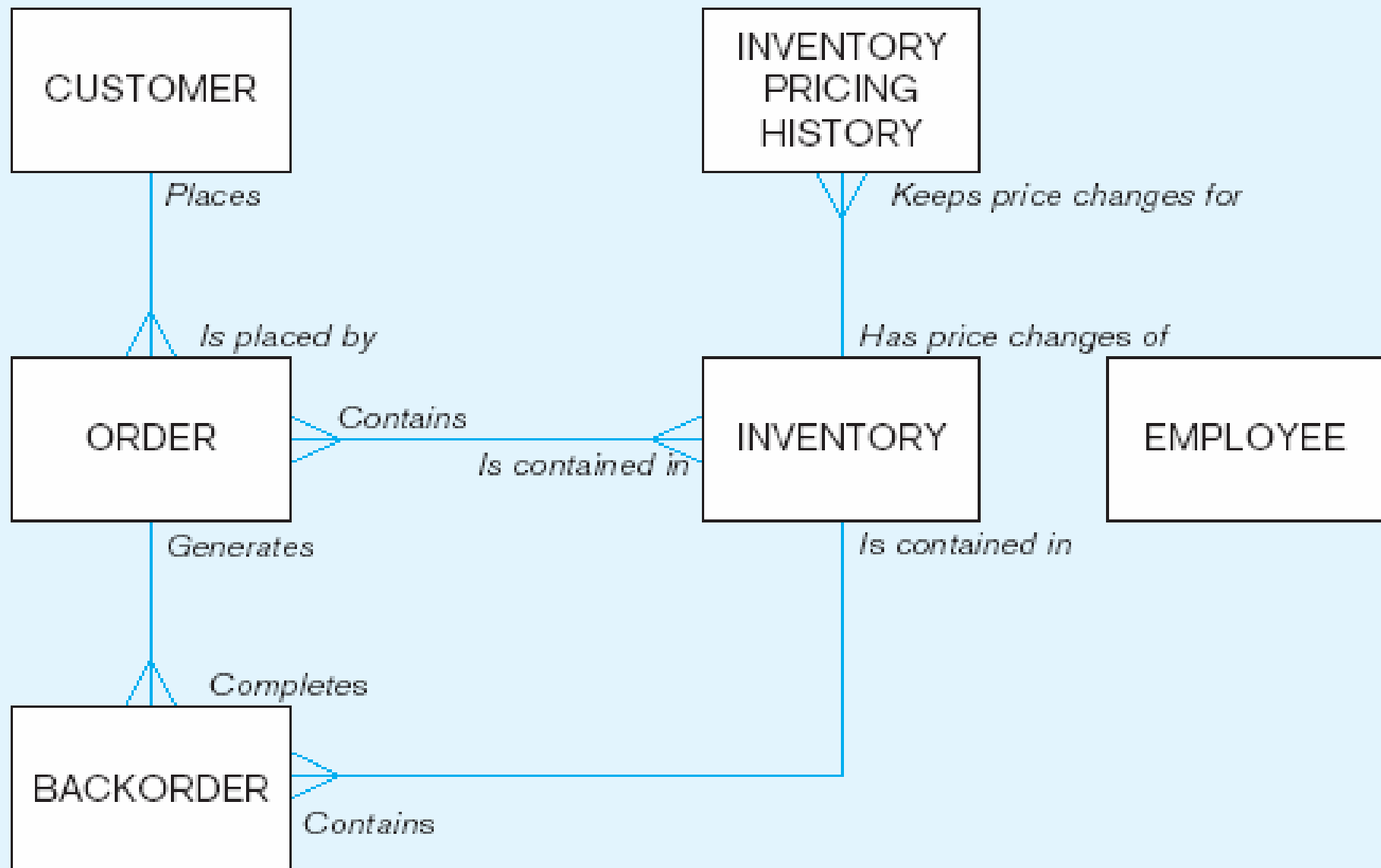
Contains

Is contained in

Therefore, one order involves many products and one product is involved in many orders

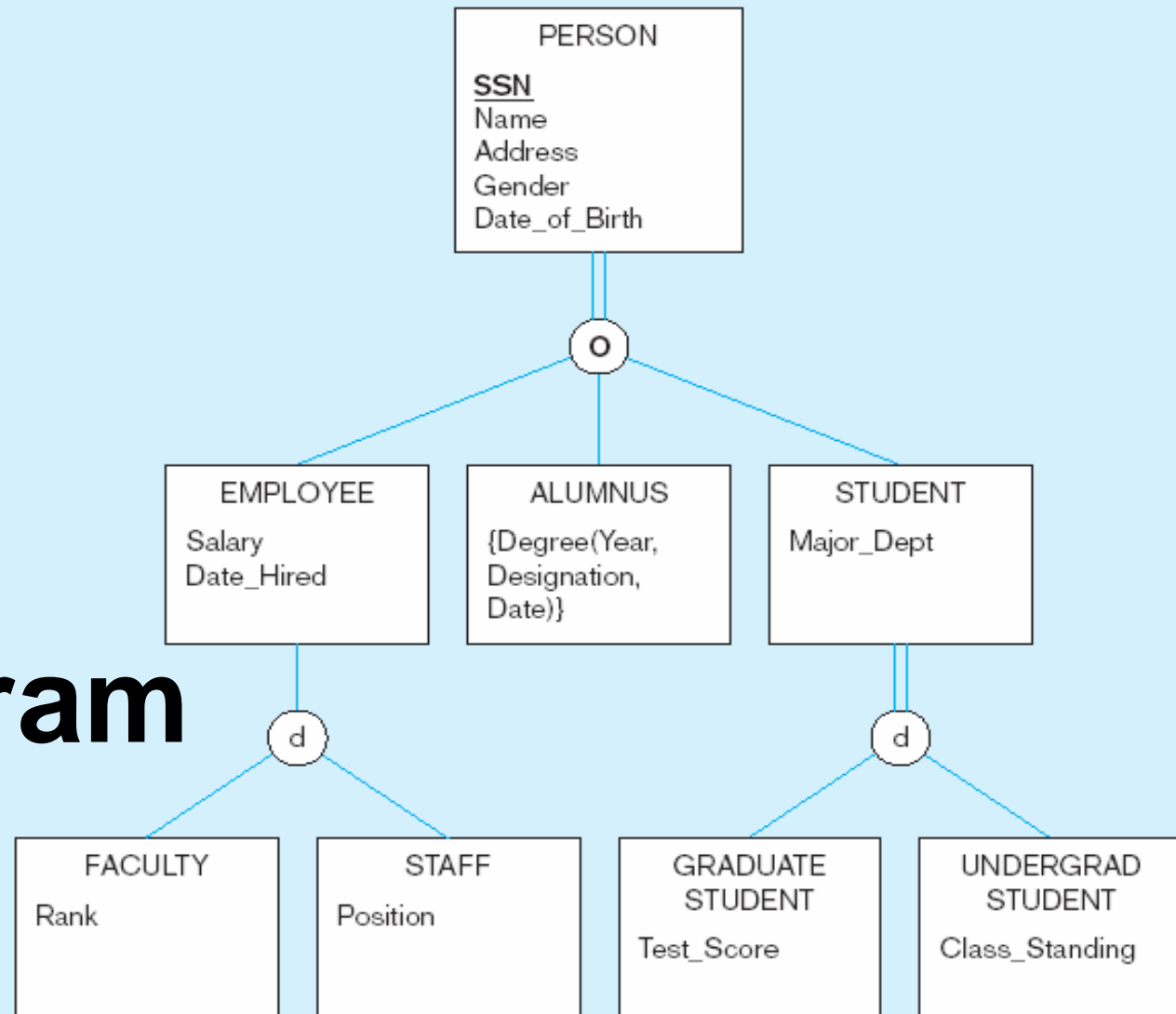
→ Many-to-many relationship

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



Topic 3

ER diagram





Elements of the Database Approach

- Enterprise Data Model
 - Graphical model showing high-level entities and relationships for the organization
- Relational Databases
 - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships
- 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

Relational Databases

Figure 1-5a Four relations (Pine Valley Furniture Company) - ORDER and Order_line tables

The screenshot shows two tables in Microsoft Access. The first table, 'ORDER_t', has columns 'Order ID', 'Order Date', and 'Customer ID'. The second table, 'Order_line_t', has columns 'Order_ID', 'Product_ID', and 'Ordered_Quantity'. Red boxes highlight the value '1007' in the 'Order_ID' column of both tables, with red lines connecting them to show the relationship.

Order ID	Order Date	Customer ID
1001	10/21/2004	1
1002	10/21/2004	8
1003	10/22/2004	15
1004	10/22/2004	5
1005	10/24/2004	3
1006	10/24/2004	2
1007	10/27/2004	11
1008	10/30/2004	12
1009	11/5/2004	4
1010	11/5/2004	1
0		0

Order_ID	Product_ID	Ordered_Quantity
1001	1	2
1001	2	2
1001	4	1
1002	3	5
1003	3	3
1004	6	2
1004	8	2
1005	4	4
1006	4	1
1006	5	2
1006	7	2
1007	1	3
1007	2	2
1008	3	3
1008	8	3
1009	4	2
1009	7	3
1010	8	10
0	0	0

Relationships established in special columns that provide links between tables

Figure 5-6 SQL table definitions

Domain constraints

Topic 4

Entity integrity constraints

```
CREATE TABLE CUSTOMER
(CUSTOMER_ID          VARCHAR(5)          NOT NULL,
CUSTOMER_NAME        VARCHAR(25)         NOT NULL,
CUSTOMER ADDRESS     VARCHAR(30)         NOT NULL,
CITY                  VARCHAR(20)         NOT NULL,
STATE                 CHAR(2)             NOT NULL,
POSTAL_CODE          CHAR(10)            NOT NULL,
PRIMARY KEY (CUSTOMER_ID);

CREATE TABLE ORDER
(ORDER_ID             CHAR(5)             NOT NULL,
ORDER DATE            DATE                 NOT NULL,
CUSTOMER_ID           VARCHAR(5)          NOT NULL,
PRIMARY KEY (ORDER_ID),
FOREIGN KEY (CUSTOMER_ID) REFERENCES CUSTOMER (CUSTOMER_ID);

CREATE TABLE ORDER_LINE
(ORDER_ID             CHAR(5)             NOT NULL,
PRODUCT_ID            CHAR(5)             NOT NULL,
ORDERED_QUANTITY      INT                  NOT NULL,
PRIMARY KEY (ORDER_ID, PRODUCT_ID),
FOREIGN KEY (ORDER_ID) REFERENCES ORDER (ORDER_ID),
FOREIGN KEY (PRODUCT_ID) REFERENCES PRODUCT (PRODUCT_ID);

CREATE TABLE PRODUCT
(PRODUCT_ID           CHAR(5)             NOT NULL,
PRODUCT_DESCRIPTION   VARCHAR(25),
PRODUCT_FINISH        VARCHAR(12),
STANDARD_PRICE        DECIMAL(8,2)        NOT NULL,
PRODUCT_LINE_ID       INT                  NOT NULL,
PRIMARY KEY (PRODUCT_ID);
```

Referential integrity constraints are implemented with foreign key to primary key references

Topic 5: Functional Dependencies and Normal Forms

- With the relational model you created, find out all the partial FD in each relations. Discuss whether your relations are in the BCNF.
- If some of your relations are not in the BCNF, convert them into the BCNF.
- What anomalies did you identify before the conversion? Do they still exist in your this version of model?



Topic 6: SQL

```
SELECT CUSTOMER_NAME FROM CUSTOMER_T  
WHERE CUSTOMER_ID NOT IN  
(SELECT CUSTOMER_ID  
FROM ORDER_T, ORDER_LINE_T, PRODUCT_T  
WHERE  
ORDER_T.ORDER_ID =ORDER_LINE_T.ORDER_ID  
AND  
ORDER_LINE_T.PRODUCT_ID = PRODUCT_T.PRODUCT_ID  
AND  
PRODUCT_DESCRIPTION = 'Computer Desk');
```

Topic 7: Relational Algebra

RESEARCH_DEPT $\leftarrow \sigma_{\text{DNAME}='Research'}(\text{DEPARTMENT})$

RESEARCH_EMPS

$\leftarrow (\text{RESEARCH_DEPT} \bowtie_{\text{DNUMBER}=\text{DNO}} \text{EMPLOYEE})$

RESULT $\leftarrow \pi_{\text{FNAME}, \text{LNAME}, \text{ADDRESS}}(\text{RESEARCH_EMPS})$

Topic 8: Physical Database Design

- Data fields
- Physical record, denormalization, partitioning
- Physical files
- Indexes
- Optimizing for query performance



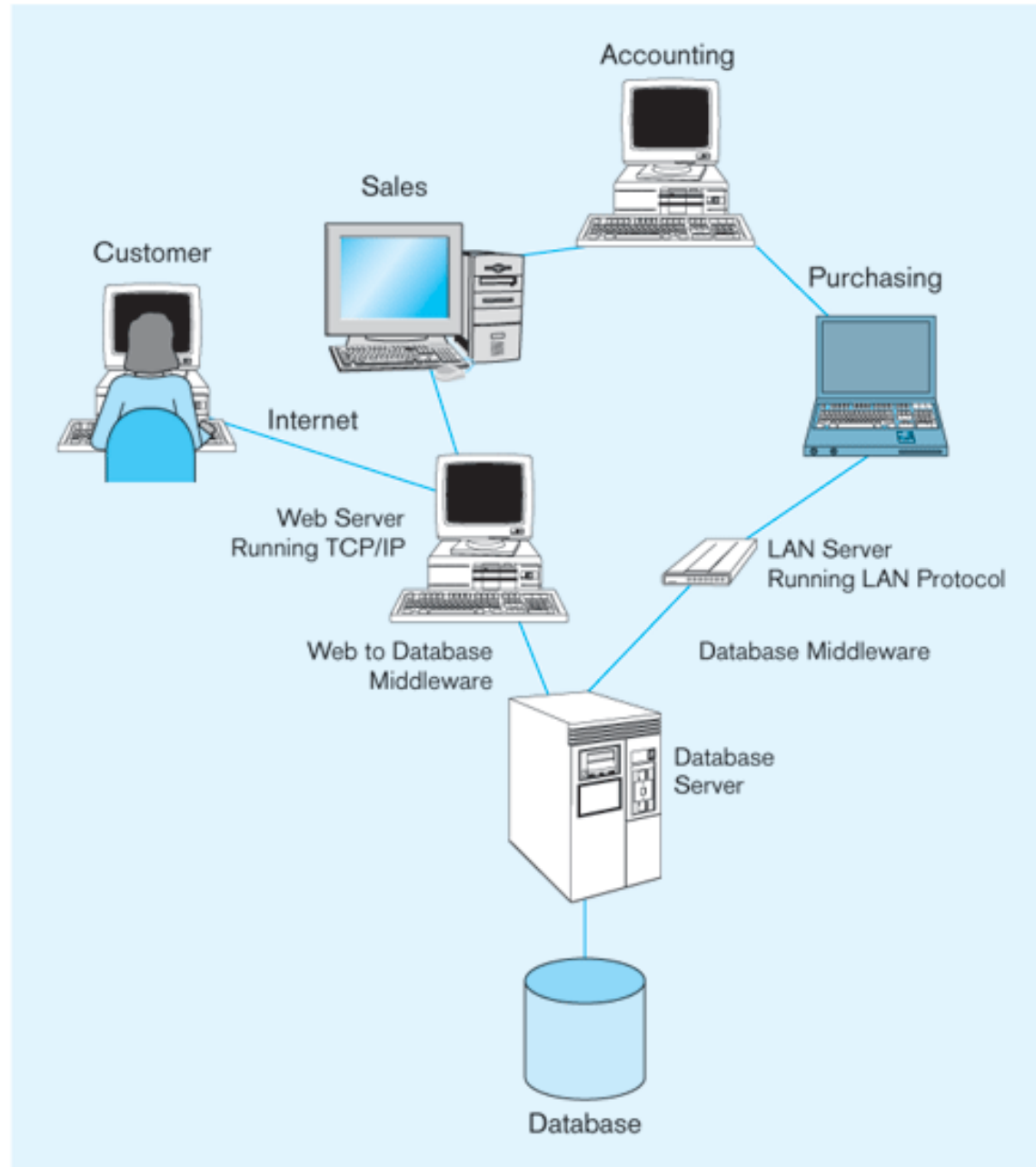
Elements of the Database Approach

- Enterprise Data Model
 - Graphical model showing high-level entities and relationships for the organization
- Relational Databases
 - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships
- 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

Topic 9

Client/server system architecture

Figure 1-6 Computer system for Pine Valley Furniture Company





Elements of the Database Approach

- Enterprise Data Model
 - Graphical model showing high-level entities and relationships for the organization
- Relational Databases
 - Database technology involving tables (relations) representing entities and primary/foreign keys representing relationships
- 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

Topic 10: Database Administration

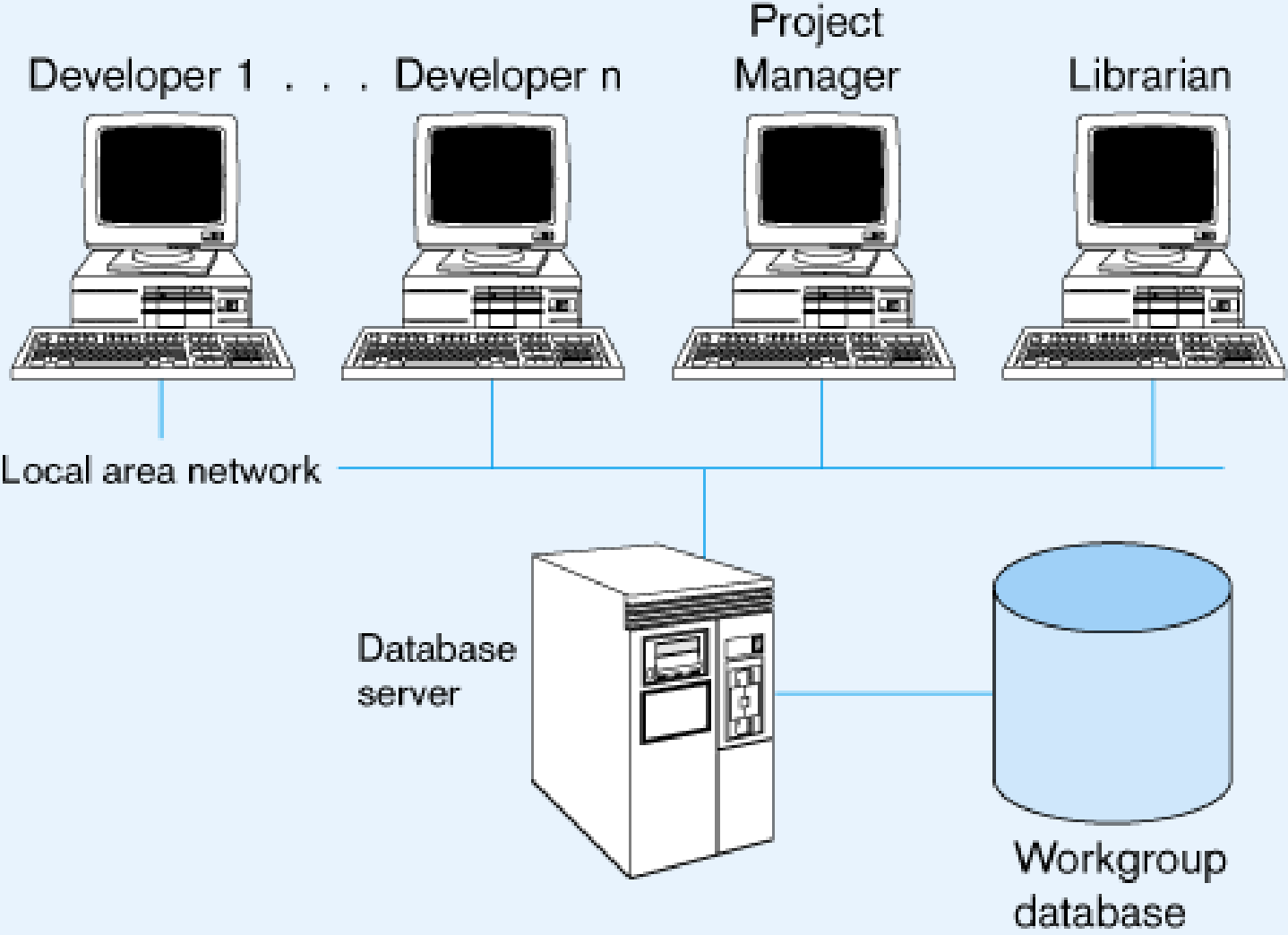
- Data and database administrations
- Data security
- Data backup and recovery
- Concurrency control



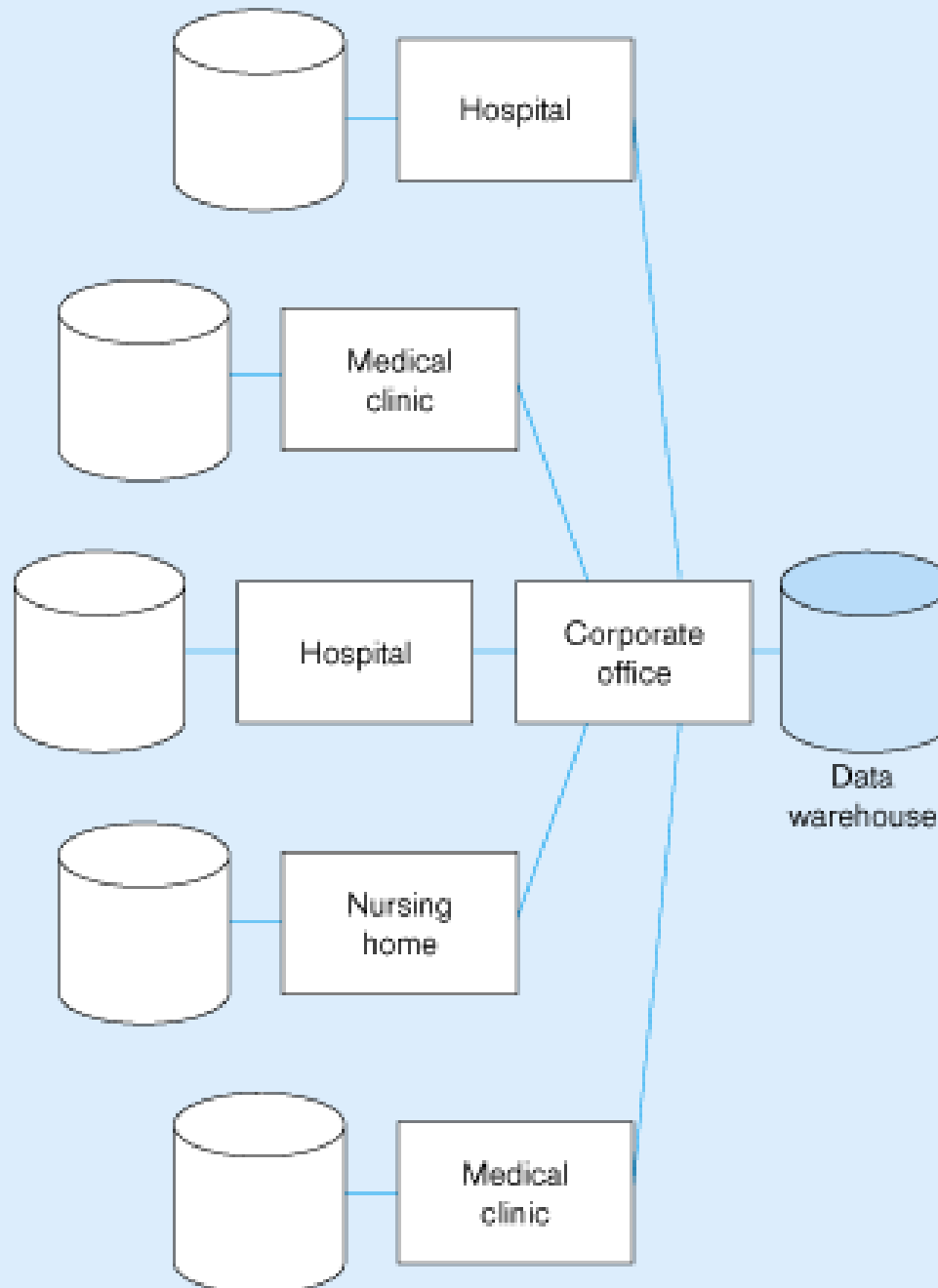
Table 1-4 Summary of Database Applications (adapted from White, 1995)

<i>Type of Database</i>	<i>Typical Number of Users</i>	<i>Typical Architecture</i>	<i>Typical Size of Database</i>
Personal	1	Desktop/laptop computer, PDA	Megabytes
Workgroup	5–25	Client/server (two-tier)	Megabytes–gigabytes
Department	25–100	Client/server (three-tier)	Gigabytes
Enterprise	>100	Client/server (distributed or parallel server)	Gigabytes–terabytes
Internet	>1000	Web server and application servers	Megabytes–gigabytes

Workgroup database with local area network



An enterprise data warehouse



Advantages of DBMS



- Program-data independence
 - Data change without changing the programs
- Minimal data redundancy
 - Avoid wasted storage space
- Improved data consistency
 - If an address is stored only in one place, it can't be inconsistent; it is also simplified to update
- Improved data sharing
 - A **user view** is a logical description of some portion of the database that is required by a user to perform some task, e.g., a table, a form or a report

Views

Figure 1-5a Four relations (Pine Valley Furniture Company) - ORDER and Order_line tables

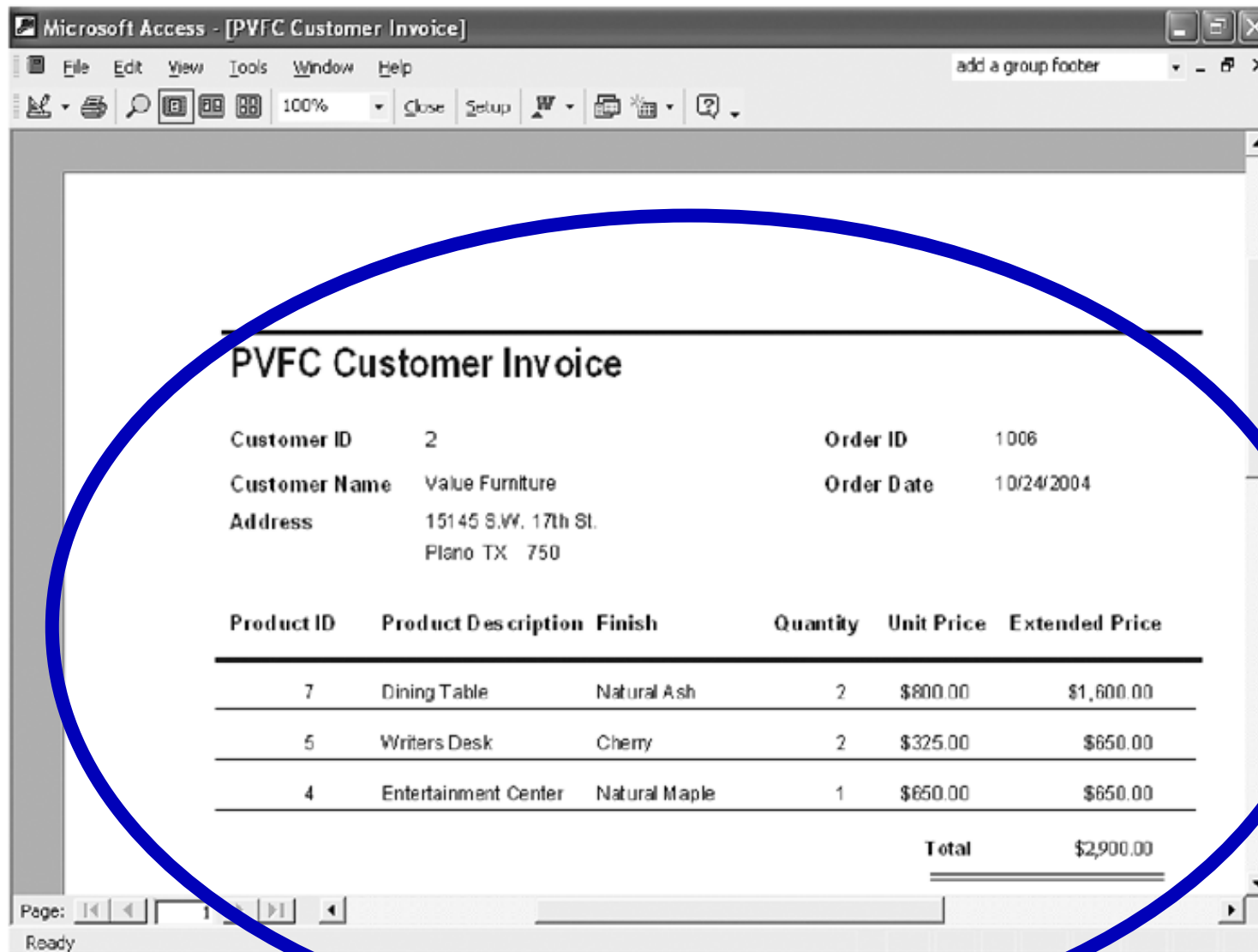
The screenshot shows the Microsoft Access interface with two tables displayed in Datasheet View. The left table, 'ORDER_t', contains 10 records with columns Order ID, Order Date, and Customer. The right table, 'Order_line_t', contains 18 records with columns Order_ID, Product_ID, and Ordered_Quantity. Both tables are circled in blue.

Order ID	Order Date	Customer
1001	10/21/2004	1
1002	10/21/2004	8
1003	10/22/2004	15
1004	10/22/2004	5
1005	10/24/2004	3
1006	10/24/2004	2
1007	10/27/2004	11
1008	10/30/2004	12
1009	11/5/2004	4
1010	11/5/2004	1
0		0

Order_ID	Product_ID	Ordered_Quantity
1001	1	2
1001	2	2
1001	4	1
1002	3	5
1003	3	3
1004	6	2
1004	8	2
1005	4	4
1006	4	1
1006	5	2
1006	7	2
1007	1	3
1007	2	2
1008	3	3
1008	8	3
1009	4	2
1009	7	3
1010	8	10
0	0	0

Views

Figure 1-7 Customer invoice (Pine Valley Furniture Company)



Microsoft Access - [PVFC Customer Invoice]

File Edit View Tools Window Help

add a group footer

100% Close Setup

PVFC Customer Invoice

Customer ID 2 Order ID 1006
Customer Name Value Furniture Order Date 10/24/2004
Address 15145 S.W. 17th St.
Plano TX 750

Product ID	Product Description	Finish	Quantity	Unit Price	Extended Price
7	Dining Table	Natural Ash	2	\$800.00	\$1,600.00
5	Writers Desk	Cherry	2	\$325.00	\$650.00
4	Entertainment Center	Natural Maple	1	\$650.00	\$650.00
Total					\$2,900.00

Page: 1

Ready

Advantages of DBMS (cont.)

- Increased productivity of application development
 - Programmers can concentrate on specific functions without worrying about low-level implementation details
- Enforcement of standards
 - Naming conventions, data quality standards, uniform procedures for data access/update...
- Improved data quality
 - A constraint is a rule that cannot be violated by database users
 - Operational data are cleaned up before placed in the data warehouse.



Advantages of DBMS (cont.)

- Improved data accessibility and responsiveness
 - End users without programming experience can retrieve and display data
- Reduced program maintenance
 - Due to program-data independence
- Improved decision support

Costs and Risks



- New, specialized personnel
 - For DB design, implementation and admin
- Installation and management cost and complexity
 - High initial cost of hardware and software
 - Costs for trained personnel to install and maintain
- Conversion costs (from older applications)
- Need for explicit backup and recovery
 - Ensure accuracy and availability at all times
- Organization conflict
 - On data definitions, data formats and coding, rights to update shared data...

Evolution of DB Systems

- Flat files - 1960s - 1980s
 - Hierarchical – 1970s - 1990s
 - Network – 1970s - 1990s
 - Relational – 1980s - present
 - Object-oriented – 1990s - present
 - Object-relational – 1990s - present
 - Data warehousing – 1990s - present
 - Web-enabled – 1990s - present
- Very limited data independence
- Easy access via SQL
- Unstructured data (e.g. multimedia)
- Structured and Unstructured data
- Decision support DB whose content derived from operational DBs

Summary



- Limitations of file processing systems
 - Program-Data Dependence
 - Lengthy Development Times
 - Limited Data Sharing
 - Duplication of Data
 - Excessive Program Maintenance
- Database management system
 - Approach
 - Enterprise Data Model
 - Relational Databases
 - Use of Internet Technology
 - Database Applications

Summary (cont)

○ Advantages

- Improved data consistency
- Enforcement of standards
- Improved data quality
- Improved data accessibility and responsiveness
- Improved decision support
- + overcome limitations of file processing systems

○ Costs and risks

- New, specialized personnel
- Installation and management cost and complexity
- Conversion costs (from older applications)
- Need for explicit backup and recovery
- Organization conflict