Systems Analysis and Design

Alan Dennis, Barbara Haley Wixom, and Roberta Roth
John Wiley & Sons, Inc.

Slides by Candace S. Garrod
Red Rocks Community College

WILEY

Data Bases

Chapter 11

WILEY

Key Definitions

- The data storage function manages how data is stored and handled by programs that run the system
- Goals of data storage design
  - Efficient data retrieval (good response time)
  - Access to the information users need

WILEY

DATA STORAGE FORMATS

WILEY

Types of Data Storage Formats

- **Files:** electronic lists of data optimized to perform a particular transaction

- **Database:** a collection of groupings of information the relate to each other in some way.

- A **Database Management System (DBMS)** is software that creates and manipulates databases.

WILEY

Appointment File

Appointment Date	Appointment Time	Duration	Reason	Patient ID	First Name	Last Name	Phone Number	Doctor ID	Doctor Last Name
11/23/2006	2:30	.25 hour	Flu	758843	Patrick	Dennis	548-9456	V524625587	Vroman
11/23/2006	2:30	1 hour	Physical	136136	Adelaide	Kin	548-7887	T445756225	Tantalo
11/23/2006	2:45	.25 hour	Shot	544822	Chris	Pullig	525-5464	V524625587	Vroman
11/23/2006	3:00	1 hour	Physical	345344	Felicia	Marston	548-9333	B544742245	Brousseau
11/23/2006	3:00	.5 hour	Migraine	236454	Thomas	Bateman	667-8955	V524625587	Vroman
11/23/2006	3:30	.5 hour	Muscular	887777	Ryan	Nelson	525-4772	V524625587	Vroman
11/23/2006	3:30	.25 hour	Shot	966233	Peter	Todd	667-2325	T445756225	Tantalo
11/23/2006	3:45	.75 hour	Muscular	951657	Mike	Morris	663-8944	T445756225	Tantalo
11/23/2006	4:00	1 hour	Physical	223238	Ellen	Whitener	525-8874	B544742245	Brousseau
11/23/2006	4:00	.5 hour	Flu	365548	Jerry	Starsia	548-9887	V524625587	Vroman
11/23/2006	4:30	1 hour	Minor surg	398633	Susan	Perry	525-6632	V524625587	Vroman
11/23/2006	4:30	.5 hour	Migraine	222577	Elizabeth	Gray	667-8400	T445756225	Tantalo
11/24/2006	8:30	.25 hour	Shot	858756	Elias	Awad	663-6364	T445756225	Tantalo
11/24/2006	8:30	1 hour	Minor surg	232158	Andy	Ruppel	525-9888	V524625587	Vroman
11/24/2006	8:30	.25 hour	Flu	244875	Rick	Grenci	548-2114	B544742245	Brousseau
11/24/2006	8:45	.5 hour	Muscular	655683	Eric	Meier	667-0254	T445756225	Tantalo
11/24/2006	8:45	1 hour	Physical	447521	Jane	Pace	548-0025	B544742245	Brousseau
11/24/2006	9:30	.5 hour	Flu	554263	Trey	Maxham	663-8547	V524625587	Vroman

Appointment Database

Appointment Date	Appointment Time	Duration	Reason	Patient ID	Doctor ID
11/23/2006	2:30	.5 hour	Flu	758843	V524625587
11/23/2006	2:30	1 hour	Physical	136136	T445756225
11/23/2006	2:45	.25 hour	Shot	544822	V524625587
11/23/2006	3:00	1 hour	Physical	345344	B544742245
11/23/2006	3:00	.5 hour	Migraine	236454	V524625587
11/23/2006	3:30	.5 hour	Muscular	887777	V524625587
11/23/2006	3:30	.25 hour	Shot	966233	T445756225
11/23/2006	3:45	.75 hour	Muscular	951657	T445756225
11/23/2006	4:00	1 hour	Physical	223238	B544742245
11/23/2006	4:00	.5 hour	Flu	365548	V524625587
11/23/2006	4:30	1 hour	Minor surg	398633	V524625587
11/23/2006	4:30	.5 hour	Migraine	222577	T445756225
11/24/2006	8:30	.25 hour	Shot	858756	T445756225
11/24/2006	8:30	1 hour	Minor surg	232158	V524625587
11/24/2006	8:30	.25 hour	Flu	244875	B544742245
11/24/2006	8:45	.5 hour	Muscular	655683	T445756225
11/24/2006	8:45	1 hour	Physical	447521	B544742245
11/24/2006	9:30	.5 hour	Flu	554263	V524625587

Tables related using patient id

Tables related using doctor id

Patient ID	First Name	Last Name	Phone Number
136136	Adelaide	Kin	548-7887
222577	Elizabeth	Gray	667-8400
223238	Ellen	Whitener	525-8874
232158	Andy	Ruppel	525-9888
236454	Thomas	Bateman	667-8955
244875	Rick	Grenci	548-2114
345344	Felicia	Marston	548-9333
365548	Jerry	Starsia	548-9887
398633	Susan	Perry	525-6632
447521	Jane	Pace	548-0025
544822	Chris	Pullig	525-5464
554263	Trey	Maxham	663-8547
655683	Eric	Meier	667-0254
758843	Patrick	Dennis	548-9456
858756	Elias	Awad	663-6364
887777	Ryan	Nelson	525-4772
951657	Mike	Morris	663-8944
966233	Peter	Todd	667-2325

Doctor ID	Last Name
B544742245	Brousseau
T445756225	Tantalo
V524625587	Vroman

File Attributes

- ☑ Files contain information formatted for a particular transaction
- ☑ Typically organized sequentially
- ☑ Pointers used to associate records with other records
- ☑ Linked Lists are files with records linked together using pointers

WILEY

File Types

- **Master files** – store core, important information

- **Look-up files** – store static values

- **Transaction files** – store information that updates a master file

- **Audit files** – record before and after versions of data

- **History (archive) files** – store past information

WILEY

Database Types

- **Legacy database**
  - Hierarchical (depict parent-child relationships using inverted trees)
  - Network (depict nonhierarchical associations using pointers)
- **Relational database**
- **Object database**
- **Multidimensional database**

Hierarchical Database Example
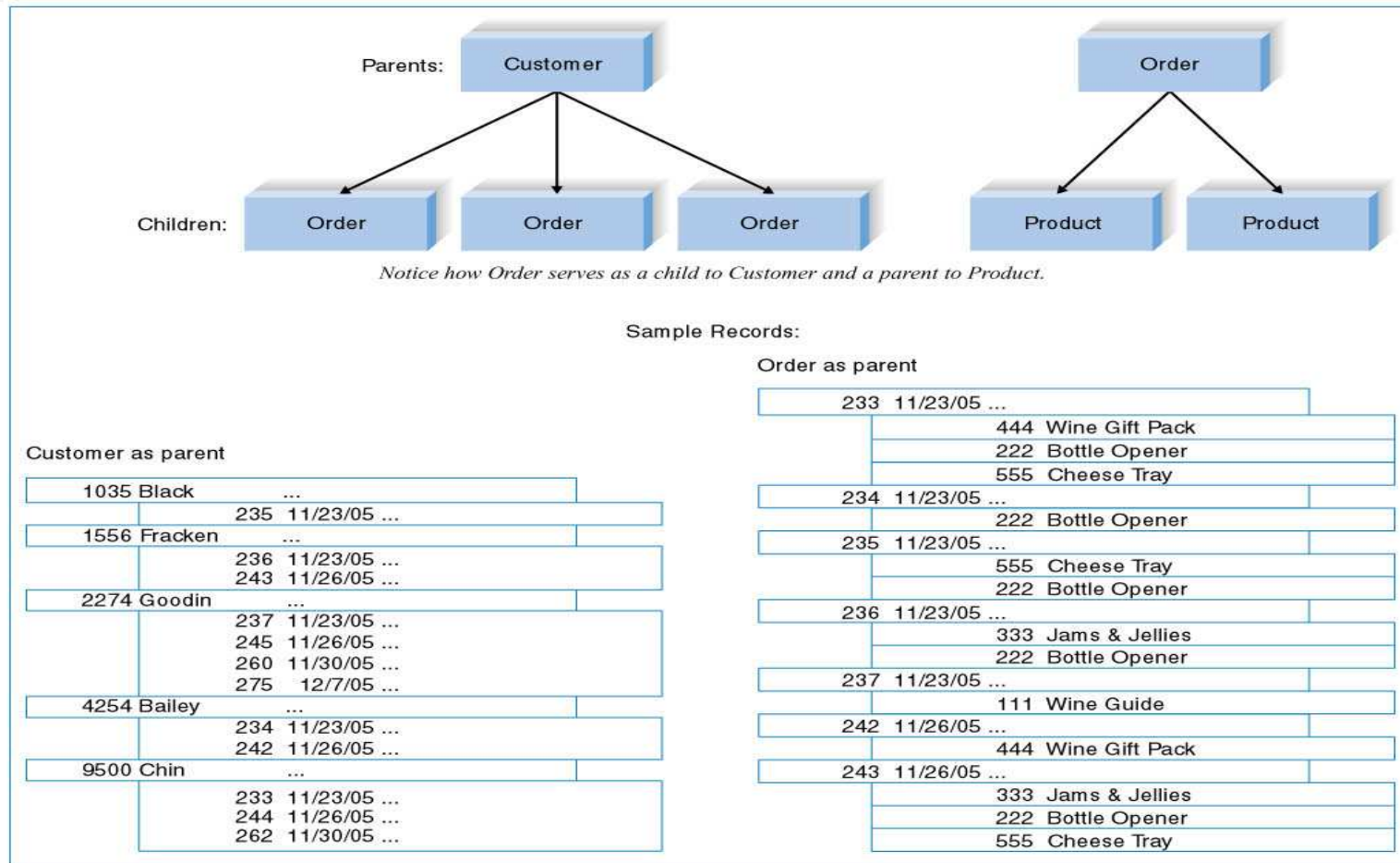
Network Database Example
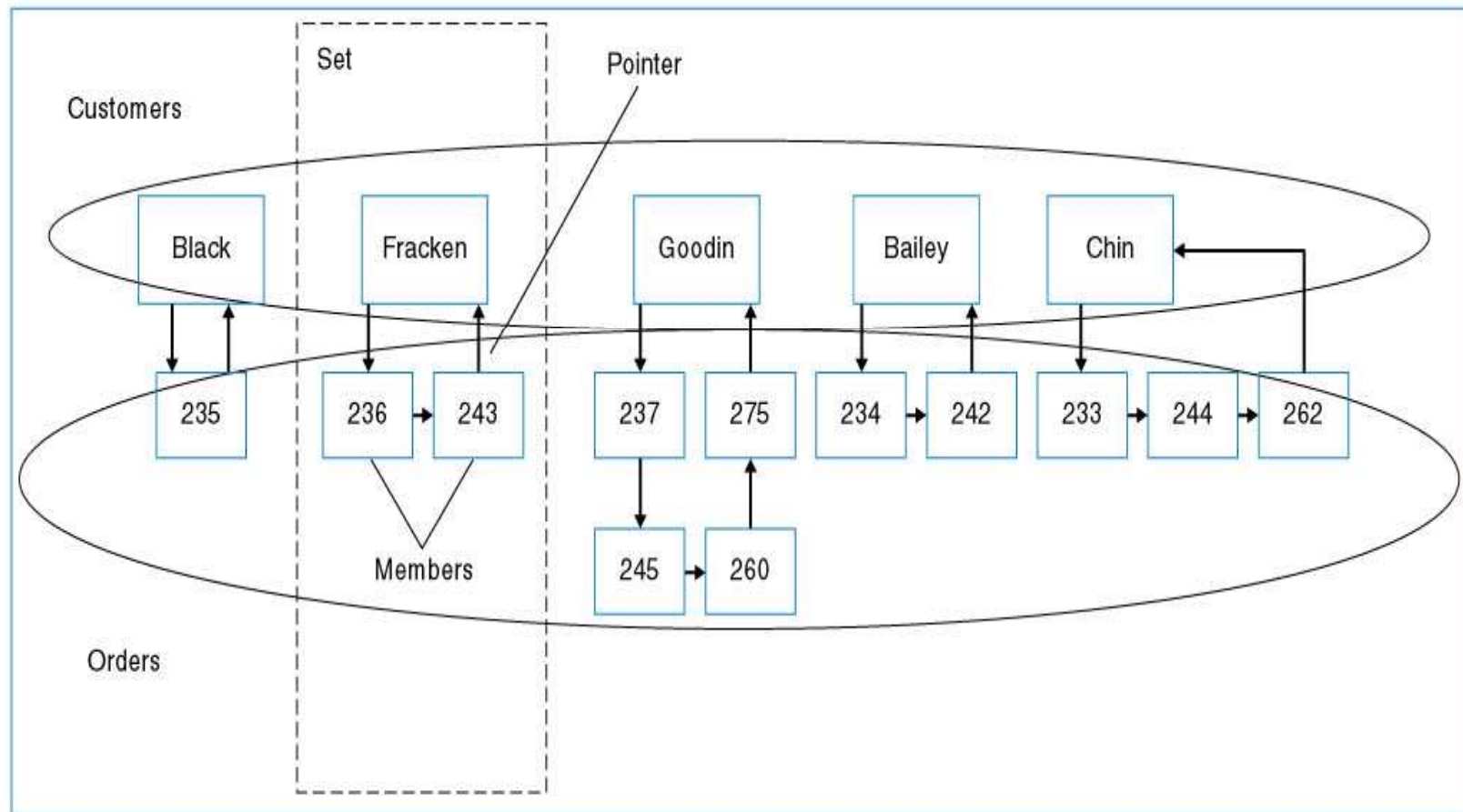
Relational Database Concepts

- Popular; easy for developers to use
- Primary and foreign keys used to identify and link tables
- Referential integrity ensures correct and valid table synchronization
- Structured Query Language (SQL)-standard language for accessing data

WILEY

Relational Database Example
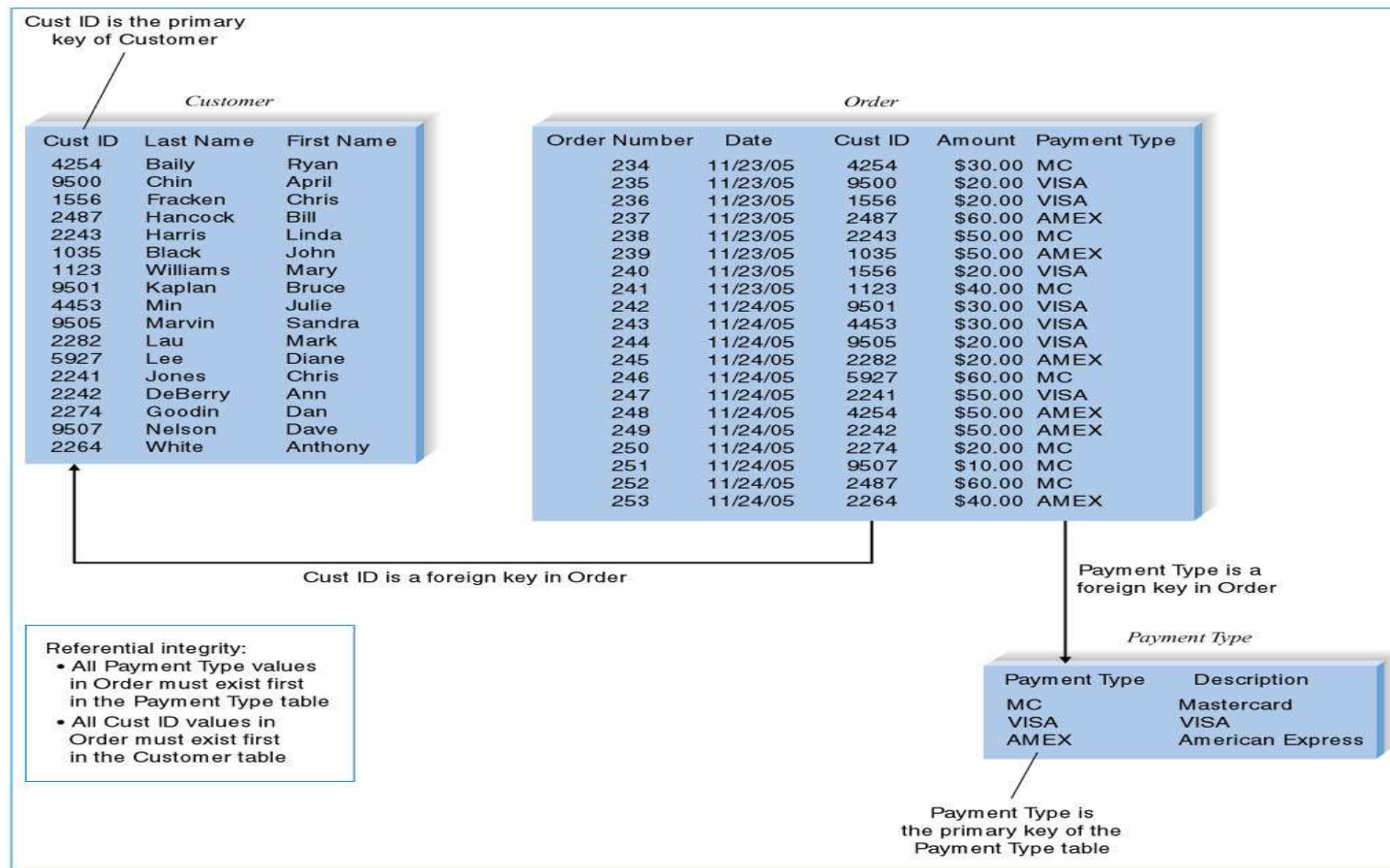
Object Database Concepts

- Built around objects consisting of both data and processes

- Objects are encapsulated (self-contained)

- Object classes – major object categories

- OODBMS – used primarily for applications with multimedia or complex data

- Hybrid OODBMS – both object and relational features

Object Database Example
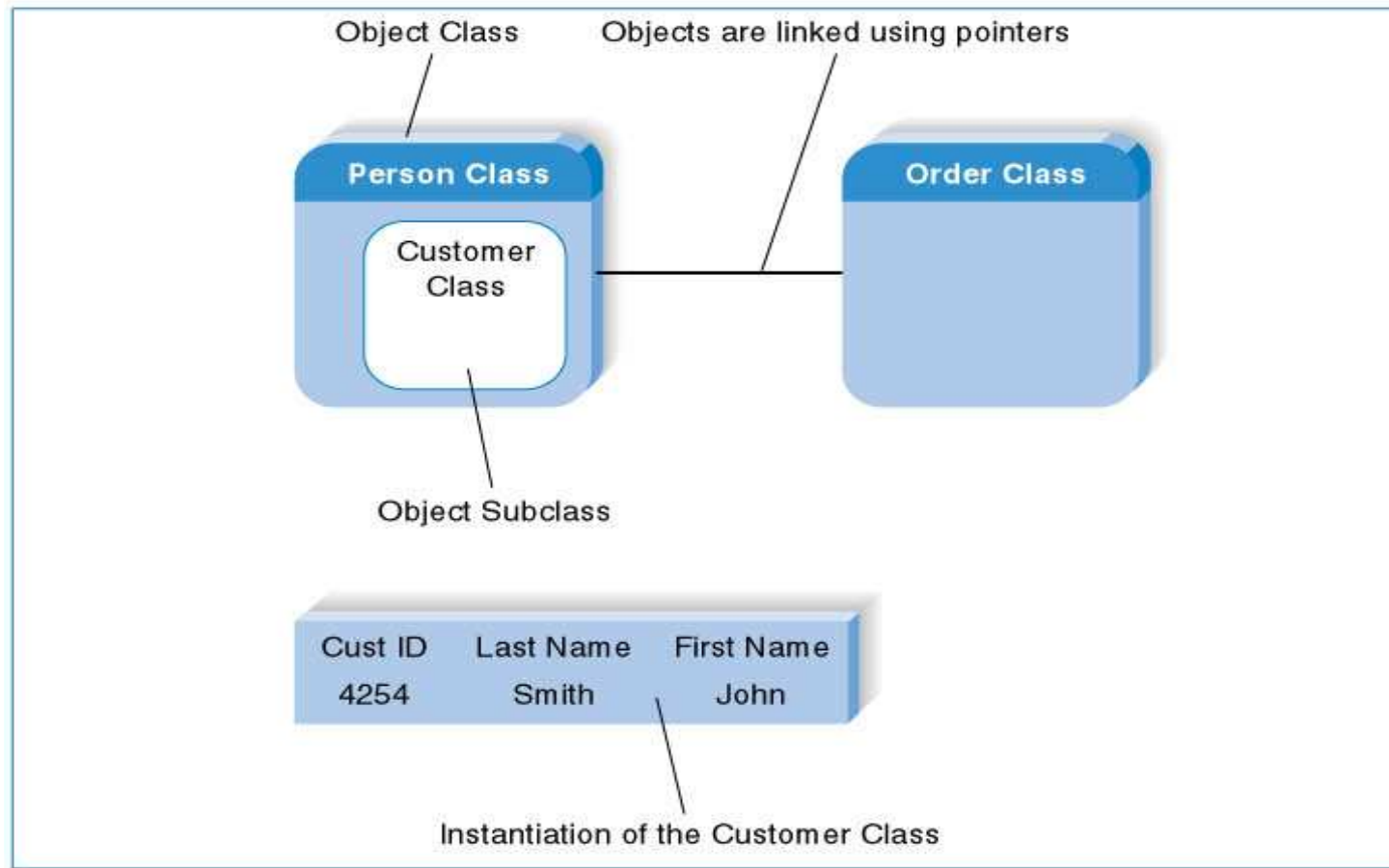
Multidimensional Database Concepts

- Stores data for easy aggregation and manipulation across many dimensions

- Used for data warehouses and data marts

- Summary data is pre-calculated and stored for fast access

WILEY

Multidimensional Database Example

Last quarter, how many customers placed more than one order using an American Express card?
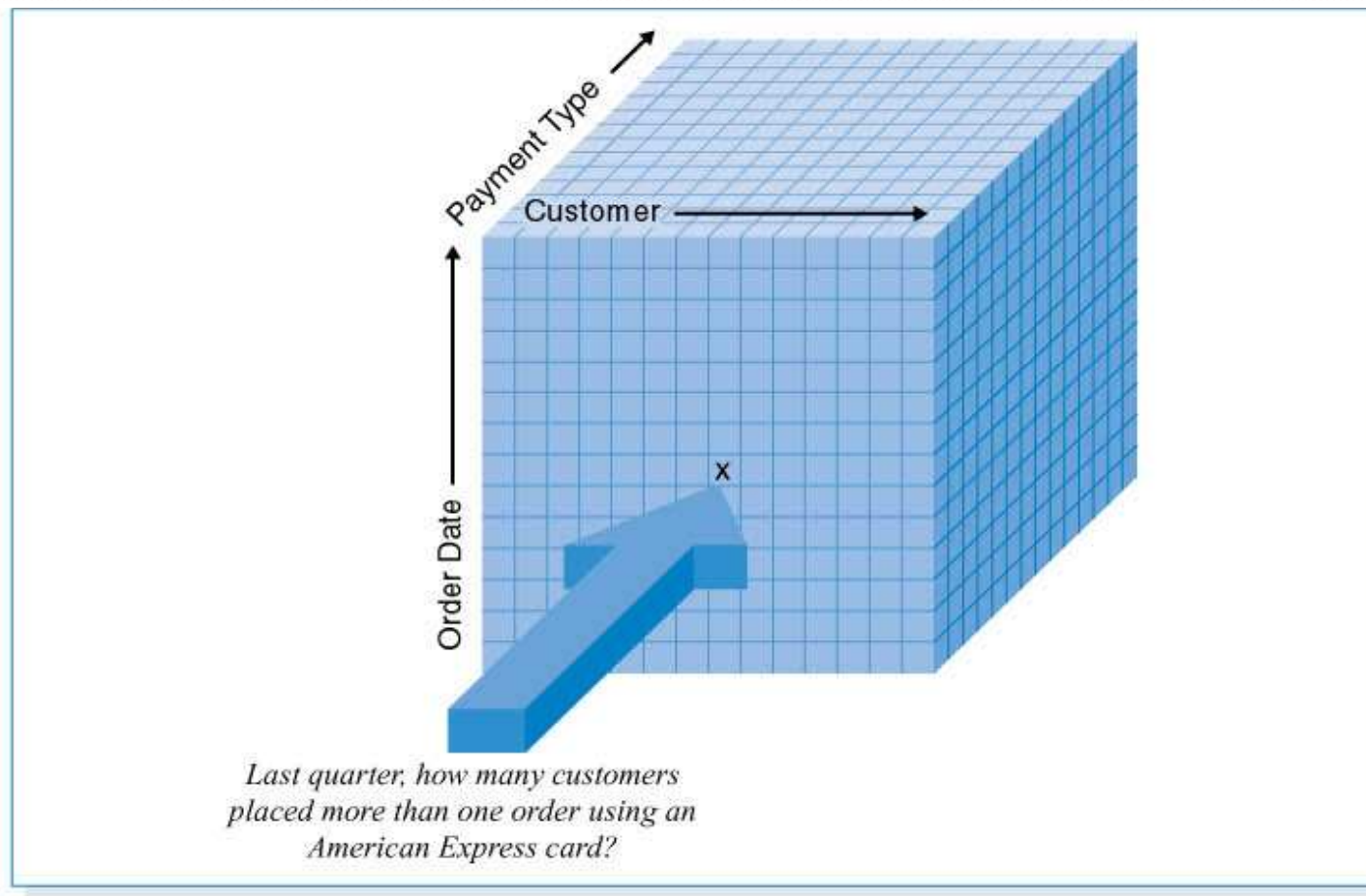
SELECTING A STORAGE FORMAT

Comparing Data Storage Formats

	Files	Legacy DBMS	Relational DBMS	Object-Oriented DBMS	Multi-dimensional DBMS
Major strengths	Files can be designed for fast performance; good for short-term data storage	Very mature products	Leader in the data-base market; can handle diverse data needs	Able to handle complex data	Configured to answer decision support questions quickly
Major weaknesses	Redundant data; data must be updated using programs	Not able to store data as efficiently; limited future	Cannot handle complex data	Technology is still maturing; skills are hard to find	Highly specialized use; skills are hard to find
Data types supported	Simple	Not recommended for new systems	Simple	Complex (e.g., video, audio, images)	Aggregated
Types of application systems supported	Transaction processing	Not recommended for new systems	Transaction processing and decision making	Transaction processing	Decision making
Existing data formats	Organization dependent	Organization dependent	Organization dependent	Organization dependent	Organization dependent
Future needs	Limited future prospects	Poor future prospects	Good future prospects	Uncertain future prospects	Uncertain future prospects

DBMS = database management system.

Selecting a Storage Format (continued)

- ☑ **Type of Application System:** the best choices for these systems usually are relational databases and multidimensional databases as the formats can be configured

- ☑ **Existing Storage Formats:** Data storage format should be selected primarily on the basis of the kind of data and application system being developed.

WILEY

Types of Data in Internet Systems

Data	Type	Use	Suggested Format
Customer information	Simple (mostly text)	Transactions	Relational
Order information	Simple (text and numbers)	Transactions	Relational
Marketing information	Both simple and complex (eventually the system will contain audio clips, video, etc.)	Transactions	Object add-on?
Information that will be exchanged with the special order system	Simple text, formatted specifically for importing into the special order system	Transactions	Transaction file
Temporary information	The Web component will likely need to hold information for temporary periods of time (e.g., the shopping cart will store order information before the order is actually placed)	Transactions	Transaction file

WILEY

MOVING FROM LOGICAL TO PHYSICAL DATA MODELS

The Physical Entity Relationship Diagram: Five Steps

Step	Explanation
Change entities to tables or files	Beginning with the logical entity relationship diagram, change the entities to tables or files and update the metadata.
Change attributes to fields	Convert the attributes to fields and update the metadata.
Add primary keys	Assign primary keys to all entities.
Add foreign keys	Add foreign keys to represent the relationships among entities.
Add system-related components	Add system-related tables and fields.

Five Step Process

- ☑ Step 1: Change Entities to Tables or Files

- ☑ Step 2: Change Attributes to Fields

- ☑ Step 3: Add Primary Keys

- ☑ Step 4: Add Foreign Keys

- ☑ Step 5: Add System-Related Components

Doctor's Office System Physical ERD
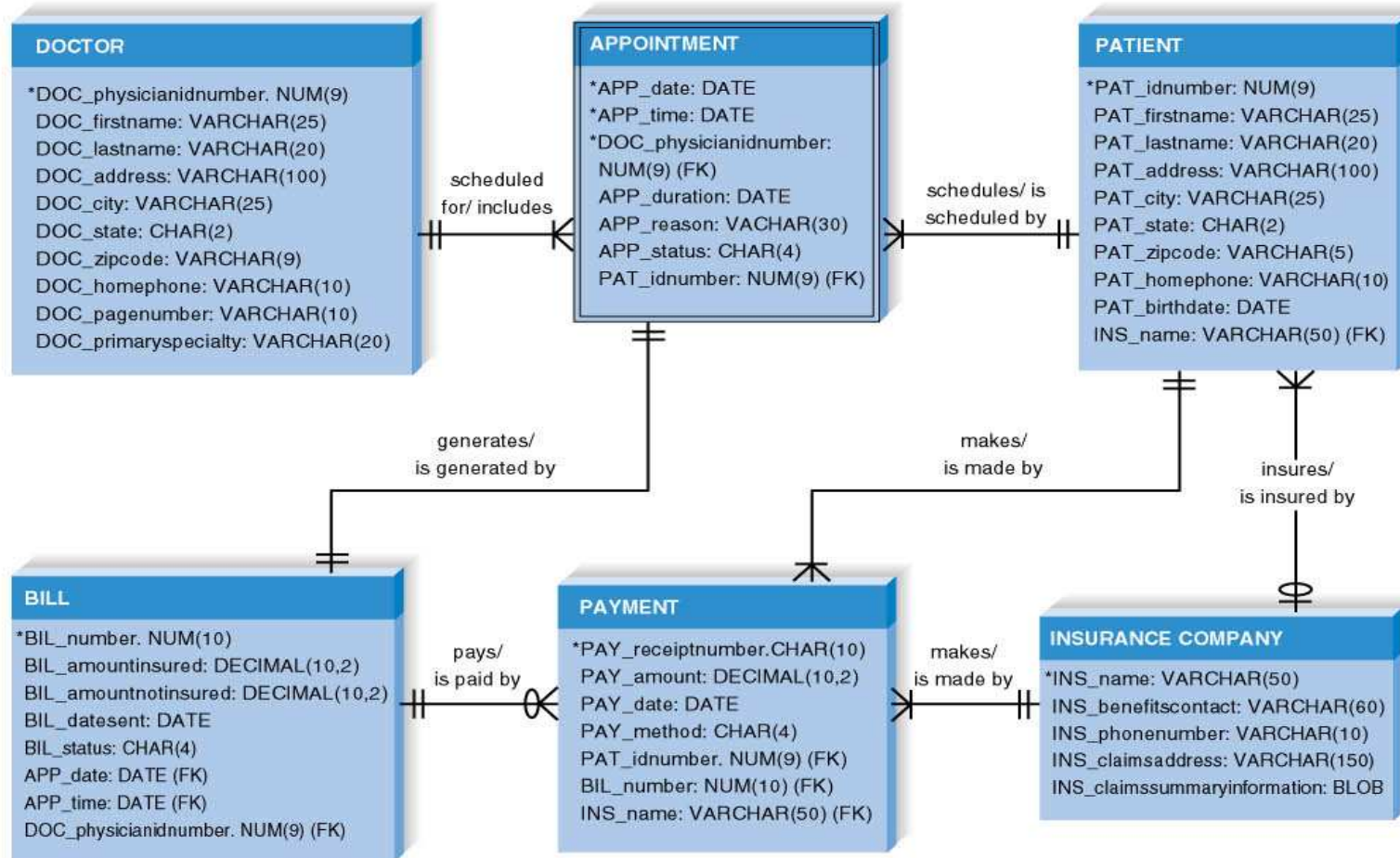
CRUD Matrix for Take Request Processes

	1.1 Find CDs	1.2 Provide CD Information	1.3 Find Stores	1.4 Place CD in Shopping Cart	1.5 Check out
CD table					
CD_sku	R	R			
CD_title	R	R			
CD_artist	R	R			
CD_category	R	R			
CD_salestatus	R	R			
CD_lastupdated					
MKTMAT table					
MAR_type		R			
VEN_id					
MAR_description					
MAR_email					
MAR_content		R			
CD_sku		R			
MAR_lastupdated					
INVENTORY table					
INV_id			R		
INV_store			R		
INV_zipcode			R		
CD_sku			R		
HOL_id					
INV_lastupdated					
HOLD table					
HOL_id					CRUD
HOL_date					CRUD
CUS_email					CRUD
HOL_lastupdated					CRUD

WILEY

OPTIMIZING DATA STORAGE

Optimizing Storage Efficiency

PowerPoint Presentation for Dennis, Wixom, & Roth Systems Analysis and Design, 4th Edition
Copyright 2009 © John Wiley & Sons, Inc. All rights reserved.

Optimizing Access Speed

- After you have optimized your data model design for data storage efficiency, the end result is data that is spread out across a number of tables

- There are several techniques that the project team can use to try to speed up access to data:
  - Denormalization
  - Clustering
  - Indexing
  - Estimating the size of data for hardware planning purposes

WILEY

Denormalized Data Model

CUSTOMER

*CUS_id: CHAR(9)
CUS_firstname: VARCHAR(15)
CUS_lastname: VARCHAR(20)
CUS_address: VARCHAR(50)
CUS_city: VARCHAR(25)
CUS_state: CHAR(2)
CUS_zipcode: VARCHAR(9)
CUS_phone: VARCHAR(10)

places/
is placed by

ORDER

*ORD_number: CHAR(18)
CUS_id: CHAR(9)(FK)
ORD_date: DATE
ORD_amount: NUM(6,2)
CUS_firstname: VARCHAR(15)
CUS_lastname: VARCHAR(20)

Customer name will be stored
in both tables

WILEY
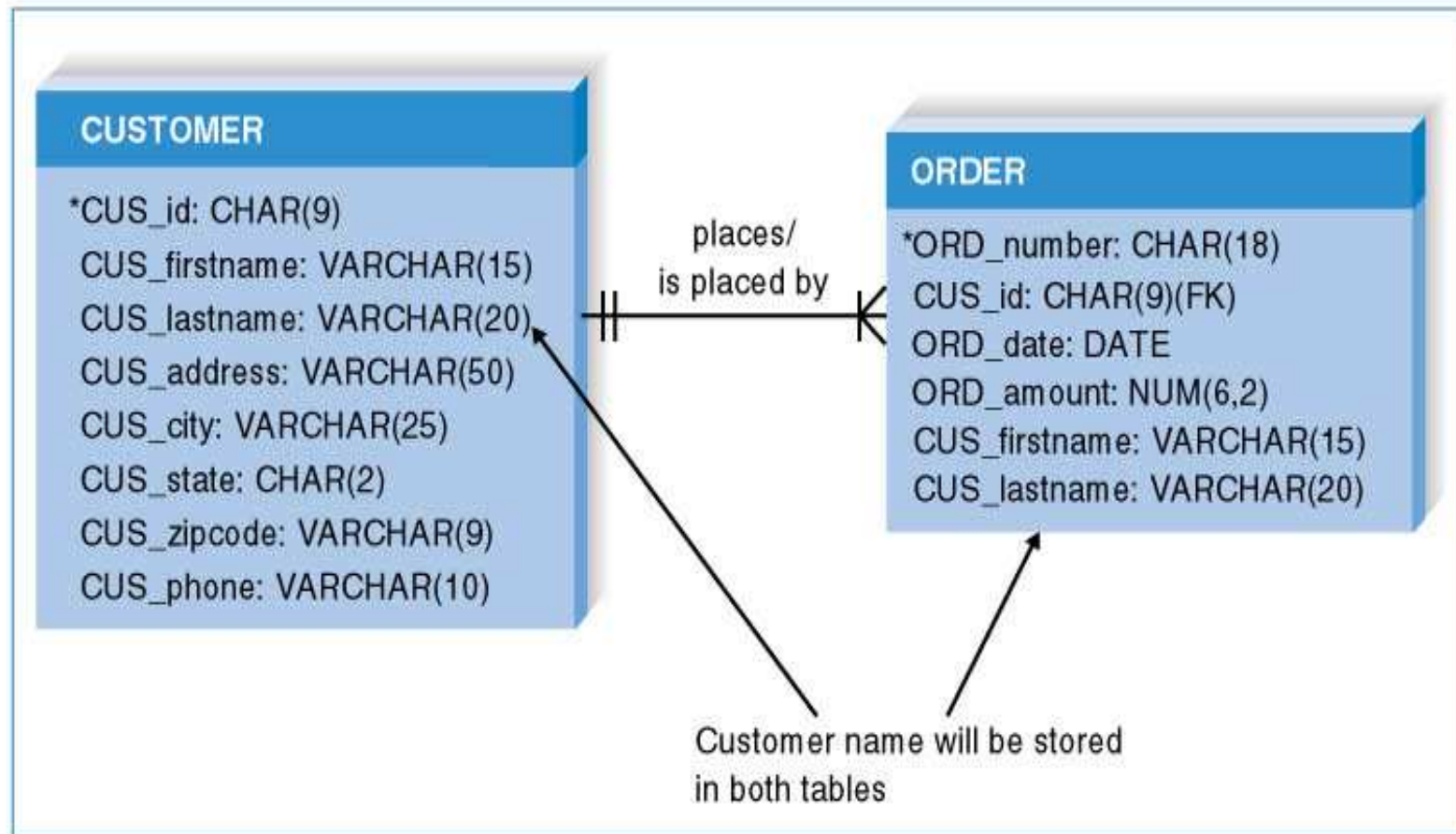
Clustering

- Reduce the number of times storage must be accessed by physically placing like records close together.
  - Intrafile clustering – similar records in a table are stored together
  - Interfile clustering – combine records from more that one table that are typically retrieved together

WILEY

Indexing

- A minitable that contains values from one or more fields in a table and the location of the values within the table
- Similar to the index of a book.

WILEY

Payment Type Index

Guidelines for Creating Indexes

- Use indexes sparingly for transaction systems.
- Use many indexes to improve response times in decision support systems.
- For each table, create a unique index that is based on the primary key.
- For each table, create an index that is based on the foreign key to improve the performance of joins.
- Create an index for fields that are used frequently for grouping, sorting, or criteria.

Estimating Storage Size

- ☑ Raw data – sum of the average widths of all fields in a table.

- ☑ Calculate overhead requirements based on DBMS vendor recommendations

- ☑ Estimate initial number of records

- ☑ Estimate growth rate of records

WILEY

Volumetrics – Estimating Data Storage Size

- Raw data – sum of the average widths of all fields in a table.
- Calculate overhead requirements based on DBMS vendor recommendations
- Estimate initial number of records
- Estimate growth rate of records

WILEY

Calculating Volumetrics

Field	Average Size (Characters)
Order number	8
Date	7
Cust ID	4
Last name	13
First name	9
State	2
Amount	4
Tax rate	2
Record size	49
Overhead	30%
Total record size	63.7
Initial table size	50,000
Initial table volume	3,185,000
Growth rate/month	1,000
Table volume @ 3 years	5,478,200

WILEY

Summary

- Files are electronic lists of data generally of five types: master, look-up, transaction, audit, and history.

- A database is a collection of groupings of information and a DBMS is software that creates and manipulates these.

- There are a number of methods for optimizing data access speed and data storage efficiency, though the designers may have to make tradeoffs between these goals.

- Physical ERDs contain references to how data will be stored in a file or database table.

WILEY

Copyright © 2009
John Wiley & Sons, Inc.