Lecture 1: Part I: Course Overview

CSX3006 DATABASE SYSTEMS

ITX3006 DATABASE MANAGEMENT SYSTEMS

Outline

- ❑ Course Overview and Objectives
- ❑ Basic Definitions
- ❑ Brief Introduction to Database Management Systems
- ❑ File-based Approach VS Database Approach
- ❑ Project
- ❑ Database Development Process
- ❑ Workshop1

Objectives

- ❑ Definition of terms
- ❑ Explain importance of databases
- ❑ Name limitations of conventional file processing
- ❑ Explain advantages of databases
- ❑ Identify costs and risks of databases
- ❑ List components of database environment

Basic Definitions: Data

❑ Stored representations of meaningful objects and events

❑ Raw facts supporting business operations and decision making

CONVENTIONAL FACTS

(STRUCTURED – *HIGHLY ORGANIZED*)

- Name
- Address
- Date of birth
- Purchased amount
- Selling price

UNCONVENTIONAL FACTS

(UNSTRUCTURED -- *NOT INCLUDED IN THIS COURSE*)

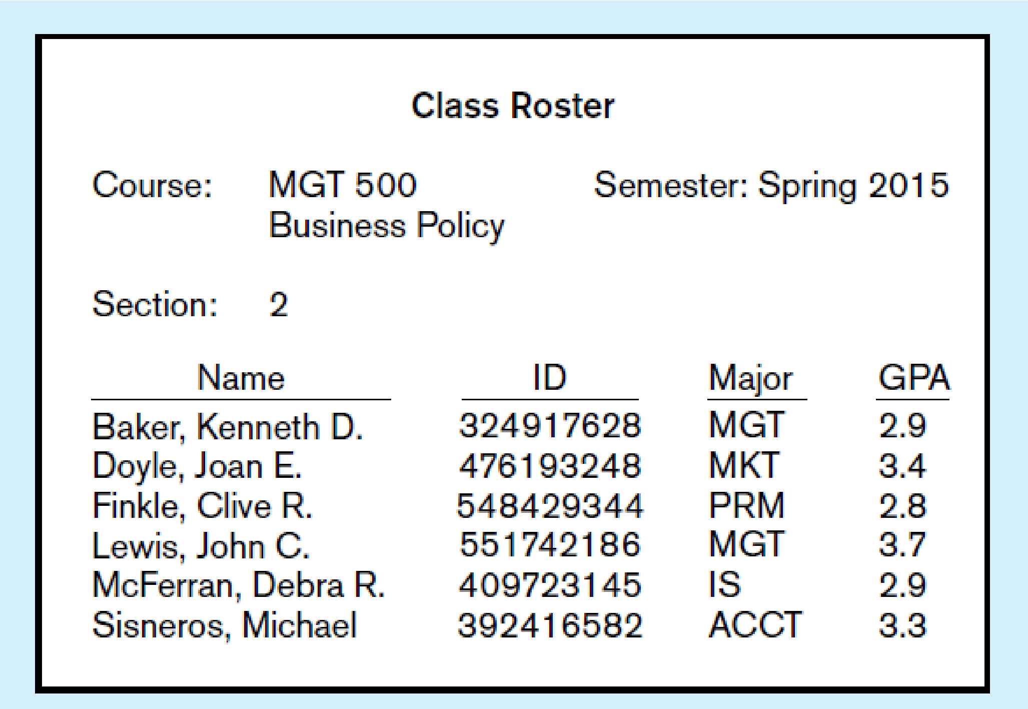
- Map
- Picture
- Blueprint
- Fingerprint
- Video
- Document

Data vs Information (1/2)

STUDENT'S DATA

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145

INFORMATION (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

Information = processed data such that the knowledge of the person who uses the data is increased.

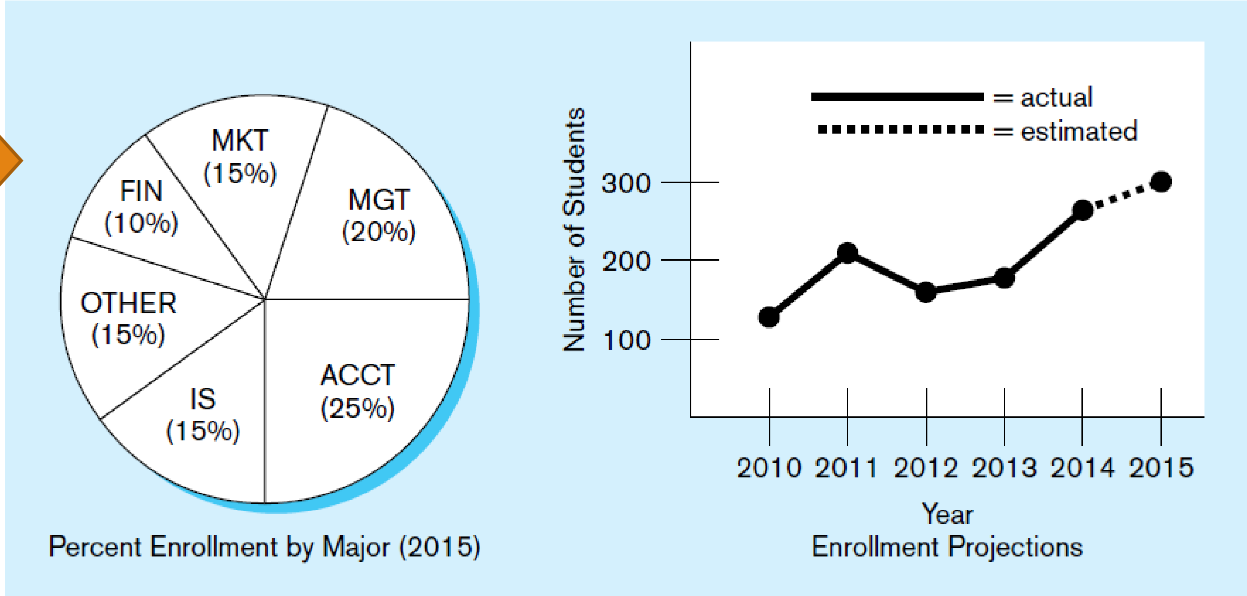
Data vs Information (2/2)

STUDENTS' DATA

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145

…

INFORMATION (SUMMARIZED DATA)



Percent Enrollment by Major (2015)

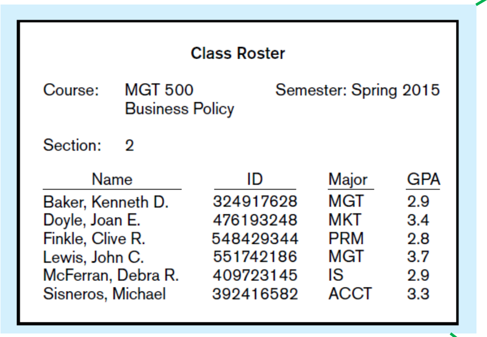
Enrollment Projections

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

Meta Data (Data Description)

Data about Data

- Describe the **properties** or **characteristics** of the data, including data types, field sizes, allowable values, and *data context*
  - Metadata describing *data context* include the source of the data, where the data are stored, ownership, and usage.



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

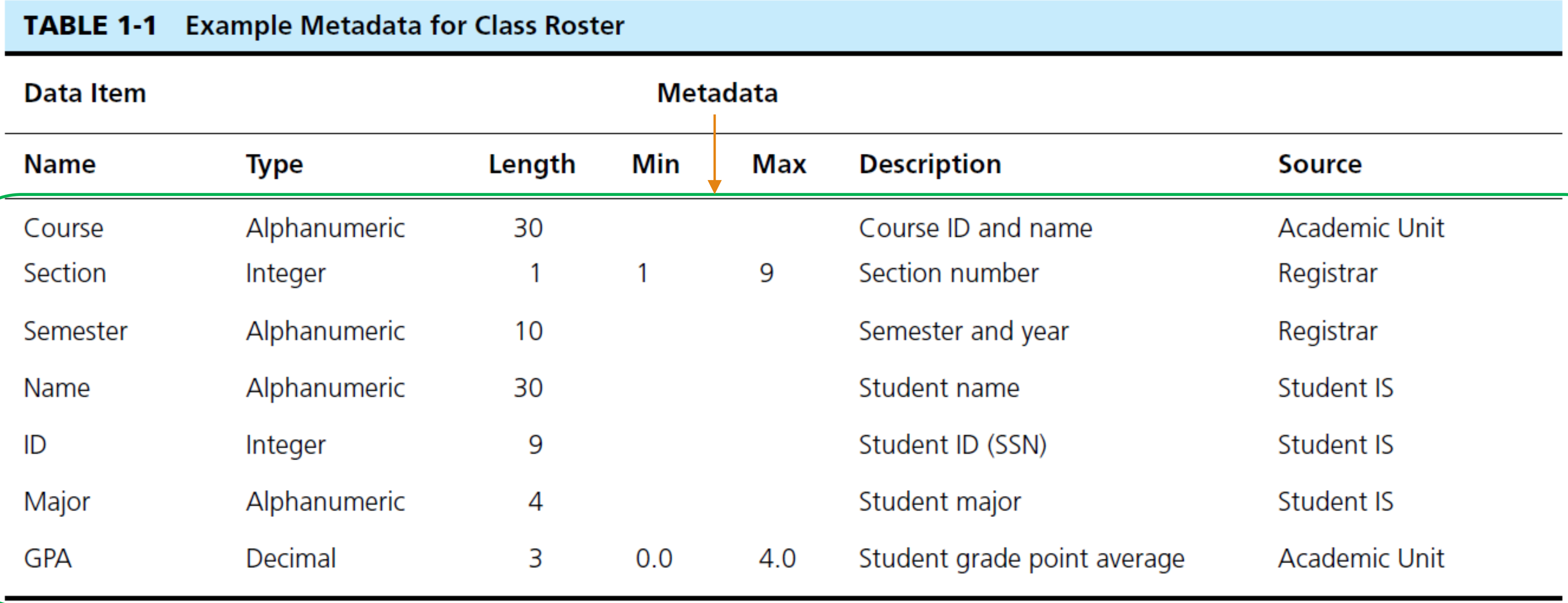


TABLE 1-1 Example Metadata for Class Roster

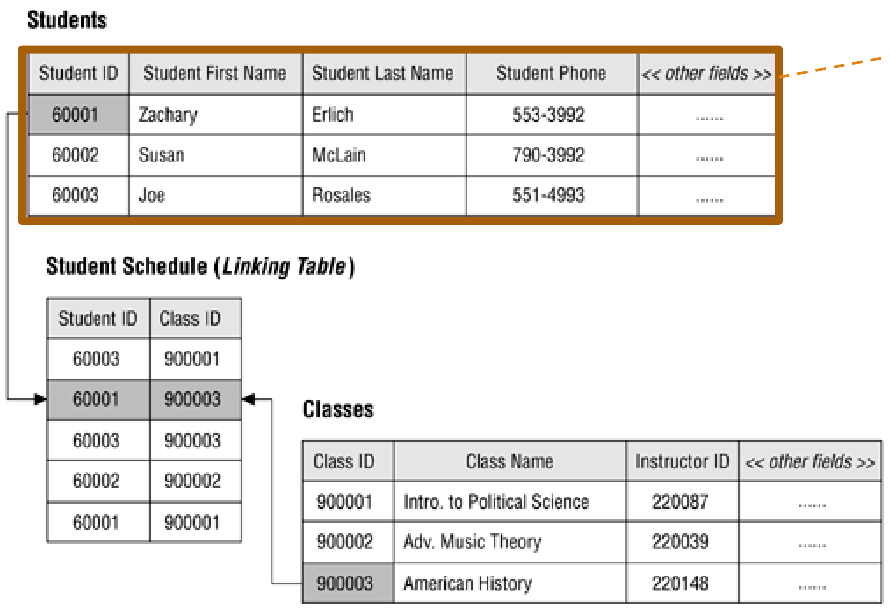
Data Item	Metadata					
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID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

Meta Data *(Cont.)*

- ❑ Metadata describe the properties of data *but* are separate from that data.
- ❑ Metadata enable database designers and users to understand the meaning of data and how to distinguish between data items that at first glance look similar.
- ❑ Managing metadata is at least as crucial as managing the associated data;
  - ❑ Data ***without*** clear meaning can be ***confusing, misinterpreted,*** or ***erroneous***.

Basic Definitions: Database

- ❑ Structured collection of logically related data as to yield useful information



Students

Student ID	Student First Name	Student Last Name	Student Phone	<< other fields >>
60001	Zachary	Erlich	553-3992	......
60002	Susan	McLain	790-3992	......
60003	Joe	Rosales	551-4993	......

Student Schedule (*Linking Table*)

Student ID	Class ID
60003	900001
60001	900003
60003	900003
60002	900002
60001	900001

Classes

Class ID	Class Name	Instructor ID	<< other fields >>
900001	Intro. to Political Science	220087	......
900002	Adv. Music Theory	220039	......
900003	American History	220148	......

Img. Source: http://etutorials.org/SQL/Database+design+for+mere+mortals/Part+I+Relational+Database+Design/Chapter+3.+Terminology/Relationship-Related+Terms/

Data Dictionary for the table **Students**

Column	Data Type	Descrip-tion	Data Size	Key Type	…
Student ID	Integer	Identification number of a student	7	PK	…
Student First Name	Varchar	First name of a student	20		…
Student Last Name	Varchar	Last name of a student	20		…
Student Phone	Integer	Phone number of a student	10		…
…	…	…	…	…	…

Basic Definitions:
Database Management System (DBMS)

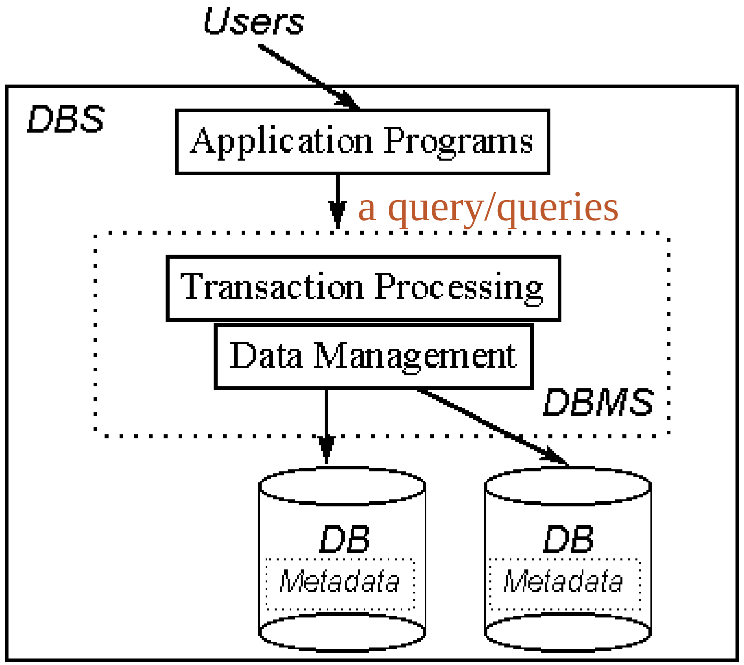
- **A collection of interrelated data** and **a software system** that is used to create, maintain, and provide controlled access to use databases
  - DBMS manages data resources like an operating system manages hardware resources
- **Primary Goal**: provide a **convenient** and **efficient way** of **managing and accessing information**
  - Provide a systematic method of creating, updating, storing, and retrieving the data stored in a database.
- Examples of well-known Relational DBMSs (RDBMSs):

ORACLE MySQL™ Microsoft® SQL Server™ PostgreSQL

A Database Management System (DBMS) – *Cont.*

- DBMS is a software system that enables the use of a database approach.
- It enables data sharing for multiple users/programmers and applications;
  - Use a single shared database rather than propagate and store new files for every new application.
- Facilities of DBMS aid controlling data access, enforcing data integrity, managing concurrency control, and restoring a database.

Basic Definitions:
Database System (DBS)



Img source: http://holowczak.com/database-management-systems-course-notes/

- ❑ Consists of **a DBMS, a database** stored and managed by the DBMS, and **all related applications** accessing the database via the DBMS that are *designed to provide necessary functionalities required by a particular application domain.*

- ❑ Database systems are used to manage collections of data that:
  - are highly valuable,
  - are relatively large, and
  - are accessed by multiple users and applications, often at the same time.

File Processing Systems

at Pine Valley Furniture Company

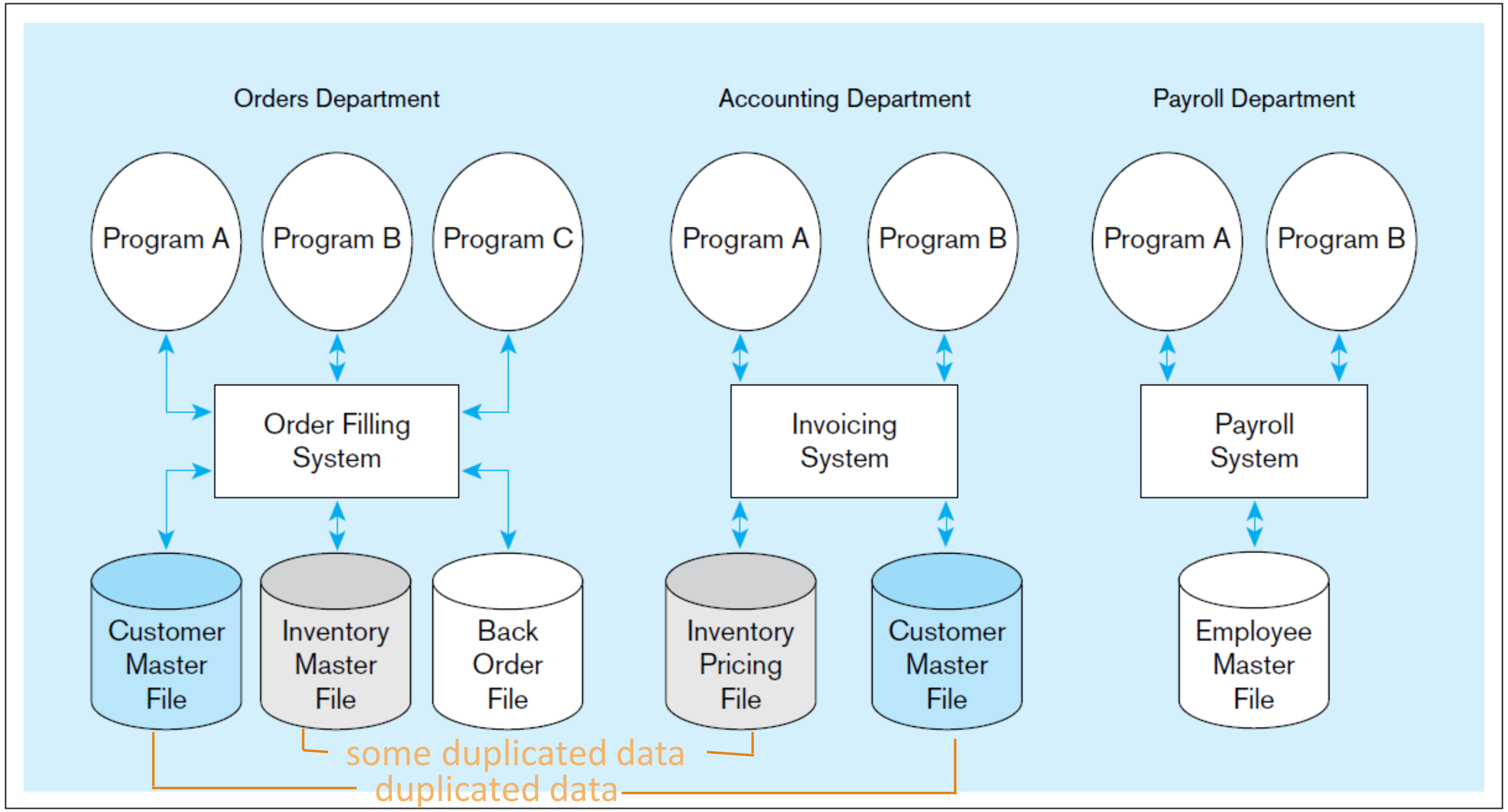


FIGURE 1-2 **Old file processing systems at Pine Valley Furniture Company**

Disadvantages of File Processing Systems

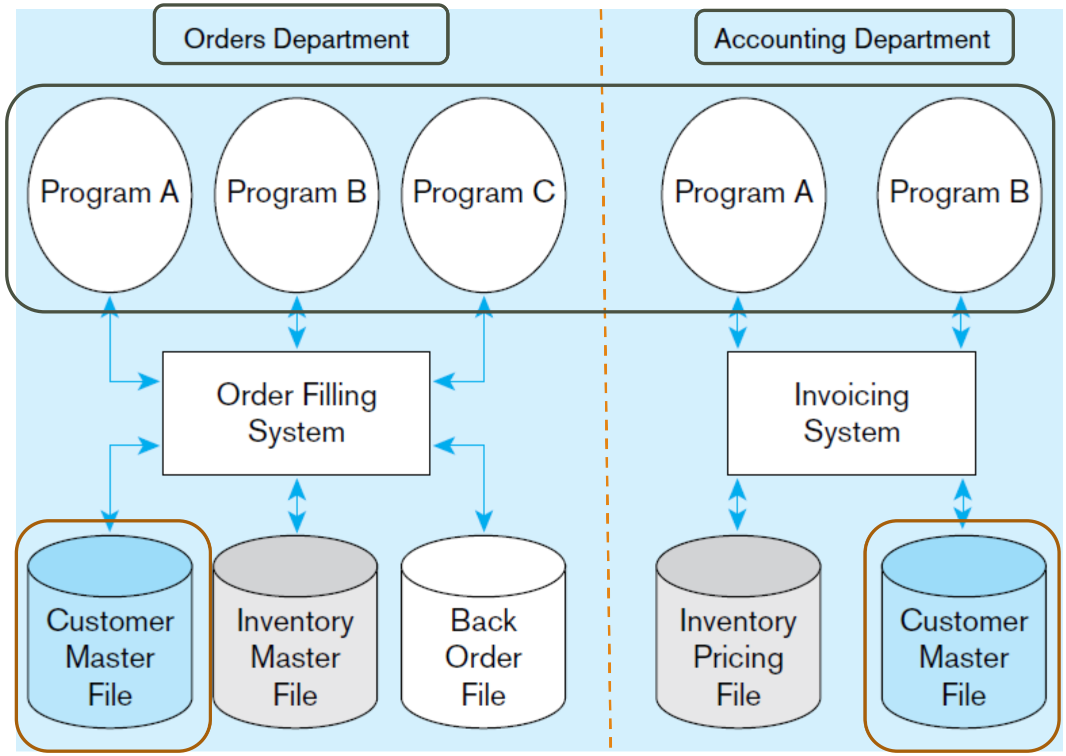
- ❑ Program-Data Dependence
  - All programs maintain metadata for each file they use
- ❑ Duplication of Data
  - Different systems/programs have separate copies of the same data
- ❑ Limited Data Sharing
  - No centralized control of data
- ❑ Lengthy Development Times
  - Programmers must design their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

Problems with Data Dependency

- ❑ Each application programmer (in each department) must maintain his/her own data
  - Lack of coordination and central control

- ❑ Non-standard file formats

- ❑ **E.g.,** changing length of `customer address` field (from 30 to 40 characters) will affect up the codes upto 5 programs



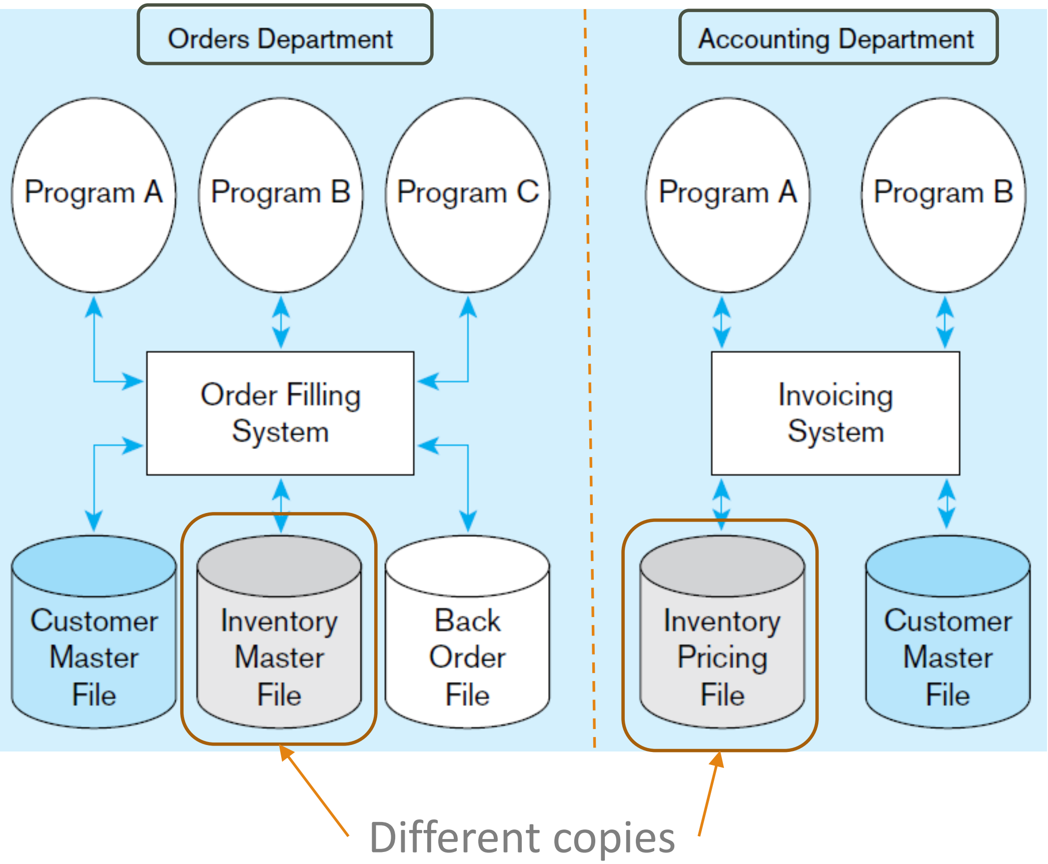
Problems with Data Redundancy

- ❑ Waste of space to have duplicate data

E.g., Inventory Master File and Inventory Pricing File store data of Valley Furniture Company's products:

- Product code,
- Product description,
- Unit price,
- Available quantity,
- Etc.

- ❑ Difficult to maintain
  - ❑ Data changes in one file could cause inconsistencies
    - ❑ E.g., Changing a value of a product's unit price



Different copies

Limited Data Sharing

E.g., accounting department’s manager must request data from other department;

- Programmers must create a scripted program to collect data from several (incompatible) files (in separate system).

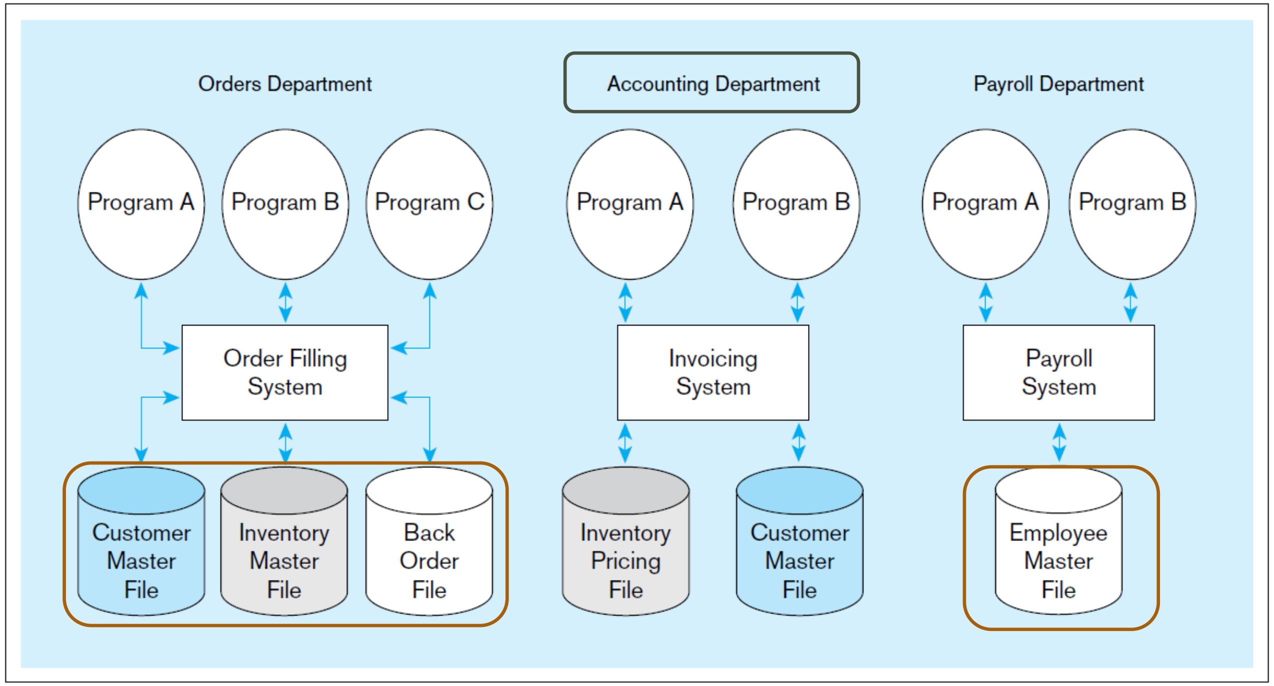


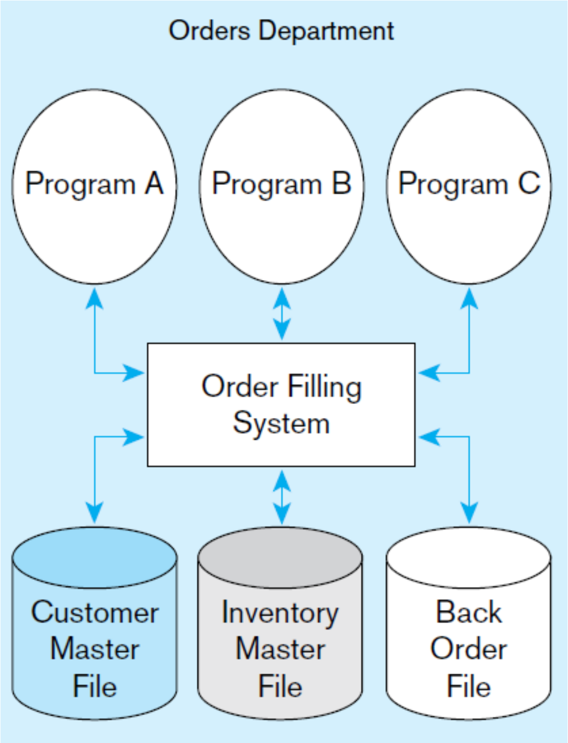
FIGURE 1-2 **Old file processing systems at Pine Valley Furniture Company**

Lengthy Development Times

Take times to write programs:

- Show top-5 best selling products

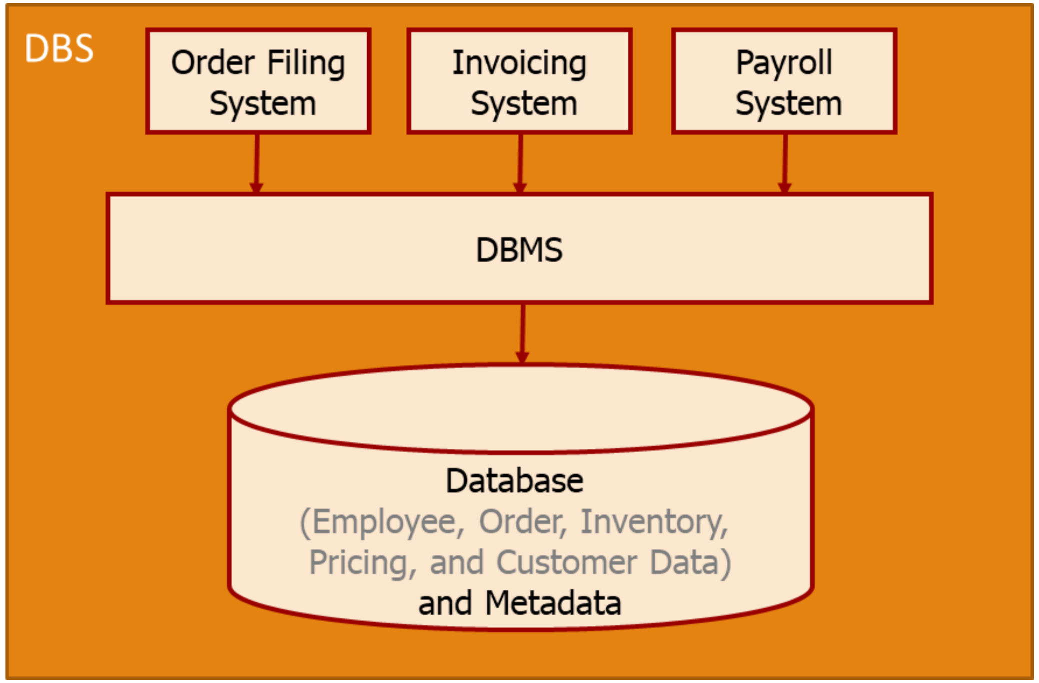
- (what if adjust to) Show top-5 best selling products of June 2021 in category ‘Kitchen’



Excessive Program Maintenance

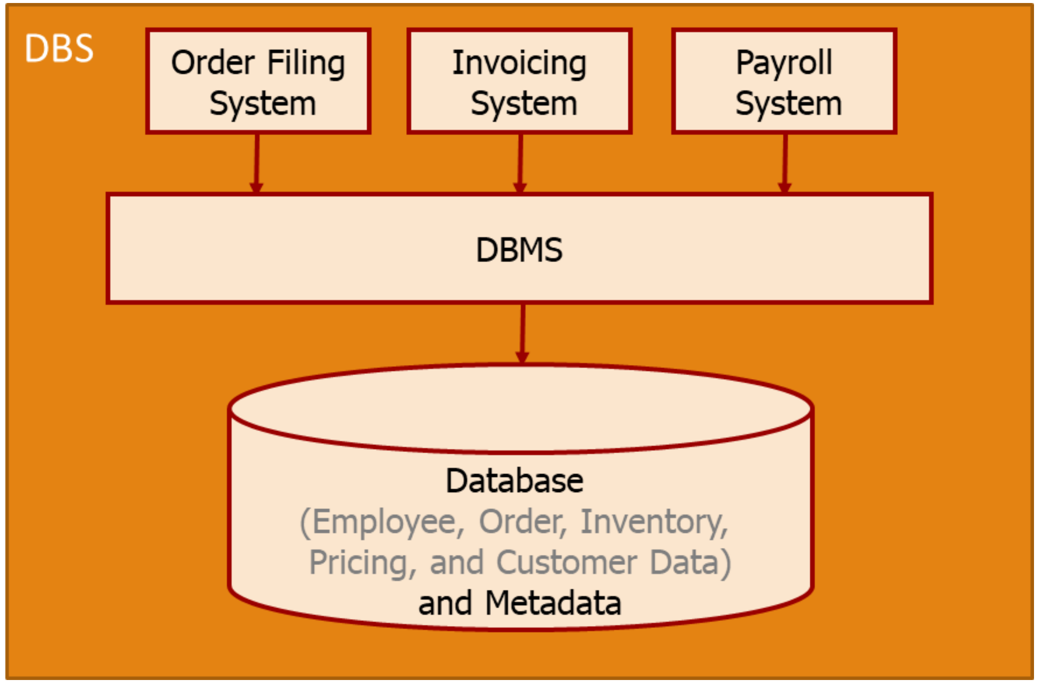
- ❑ IF an organization develops ***many separately*** managed databases with ***little or no*** coordination of the metadata, it will cause a lot of maintenance time and money.

SOLUTION: The DATABASE Approach



- ❑ Use a central repository of shared data
- ❑ Manage data by a controlling agent
- ❑ Store data in a standardized, convenient form

Elements of the Database Approach



- ❑ Data models (ER model) *-- design*
- ❑ Relational Databases
- ❑ Database Applications
- ❑ Use of Internet Technology

Data Models

(Entity-Relationship Models)

- **Graphical systems** capturing nature and relationship of data understood by end users, systems analysts, and database designers.
- **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
- Main components including Entity, Attributes, Relationships, and associating constraints

Entity, Attribute and Instance

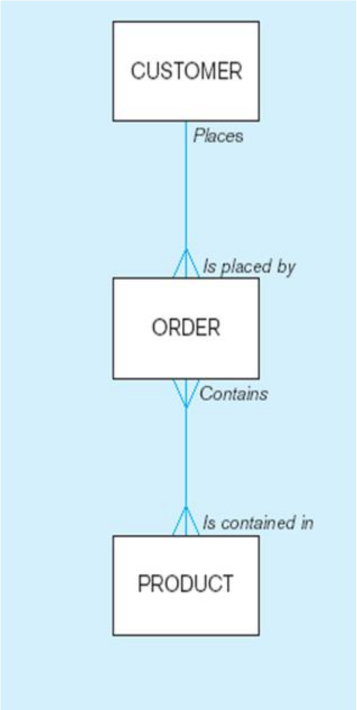
❑ **Entity:** a noun to describe a person, a place, an object, an event, or a concept in the business environment for which information ***must be recorded and retained***

❑ **Attribute:** The data (value) you are interested in capturing about the entity

- E.g., Customer's name

❑ **Instance**: a group of relating data used to identify an object

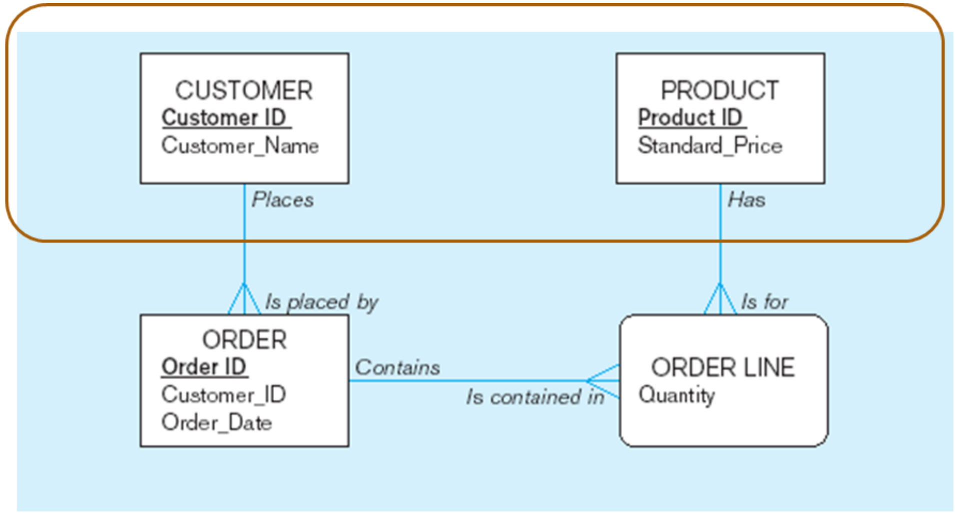
Examples of Entity



A segment of an **enterprise** data model

- 3 Entities:
  - Customer
  - Order
  - Product

Examples of Attribute of Entity



A segment of a **project-level** data model

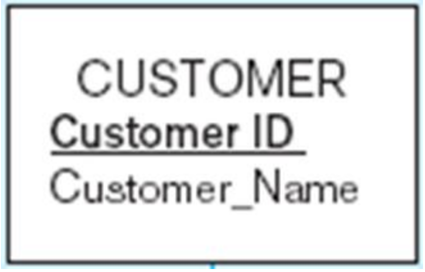
❑ Attributes of the entity Customer:

- Customer_ID
- Customer_Name

❑ Attributes of the entity Product:

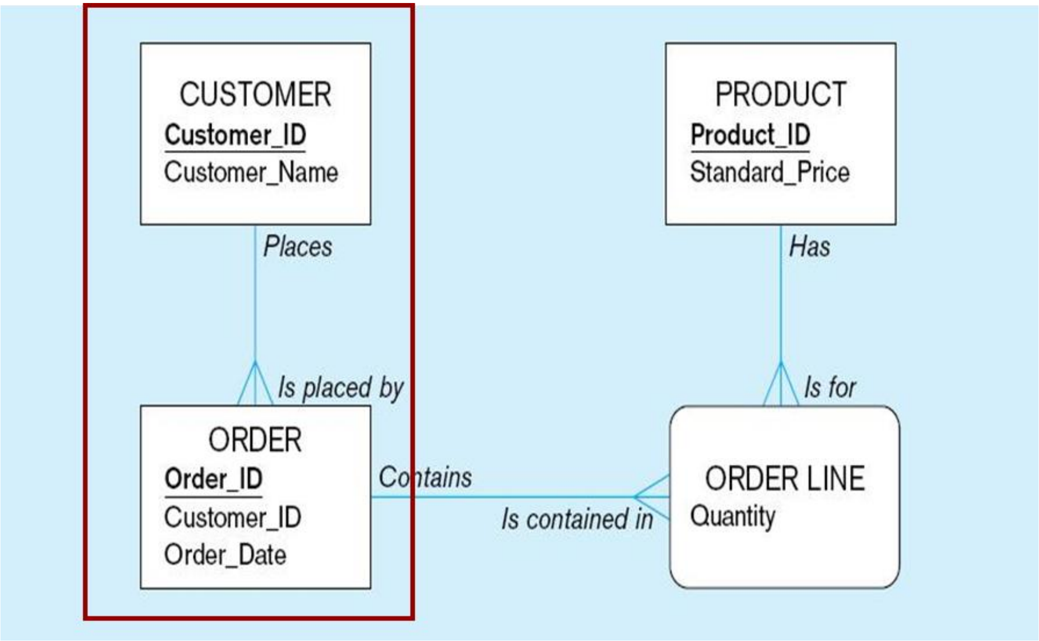
- Product_ID
- Standard_Price

Examples of Instance of Entity



- 2 instances of the entity Customer:
  - Instance 1:
    - Customer_ID: 1001
    - Customer_Name: John Doe
  - Instance 2:
    - Customer_ID: 1002
    - Customer_Name: Scott Tiger

Relationship and Constraint



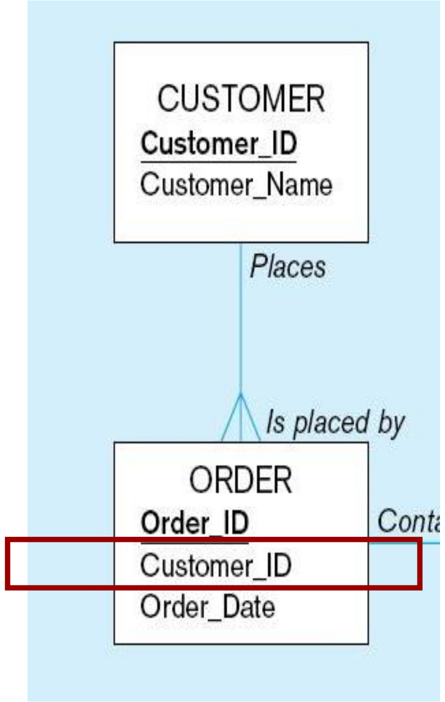
- ❑ To retrieve desired information, a well-structured database establishes the relationships between entities that exist in organizational data.
- ❑ Almost all relationships are one-to-many (1:M) or many-to-many (M:N).

Example:

- **Relationship:** `Places`
- **Constraint:** *One-to-many* relationship
- ***Detail:*** One customer may place *many* orders, but each order is placed by *a single* customer

Relational Database

- Represents data as a **collection of tables** *in which all* ***data relationships*** *are represented by* ***common attributes*** *(values) in related tables (relations).*



Relation: CUSTOMER

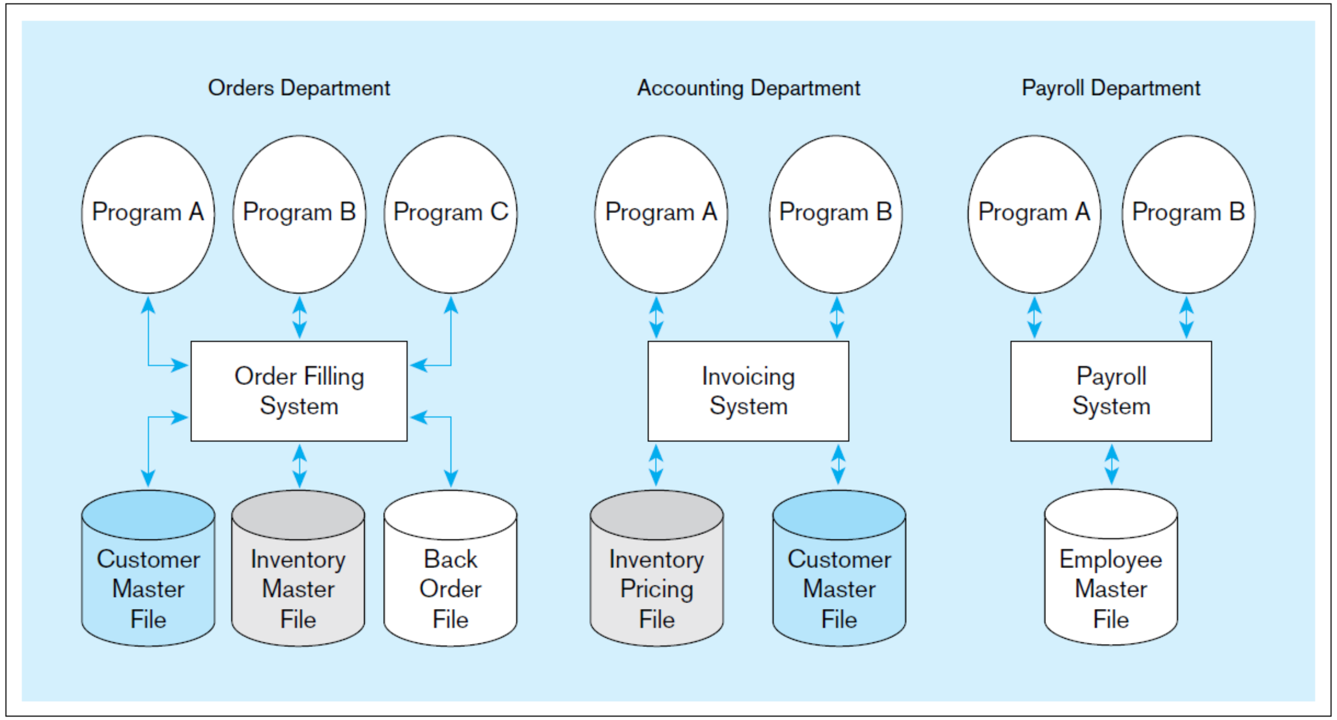
Customer_ID	Customer_Name
1001	John Doe
1002	Scott Tiger
…	…

Relation: ORDER

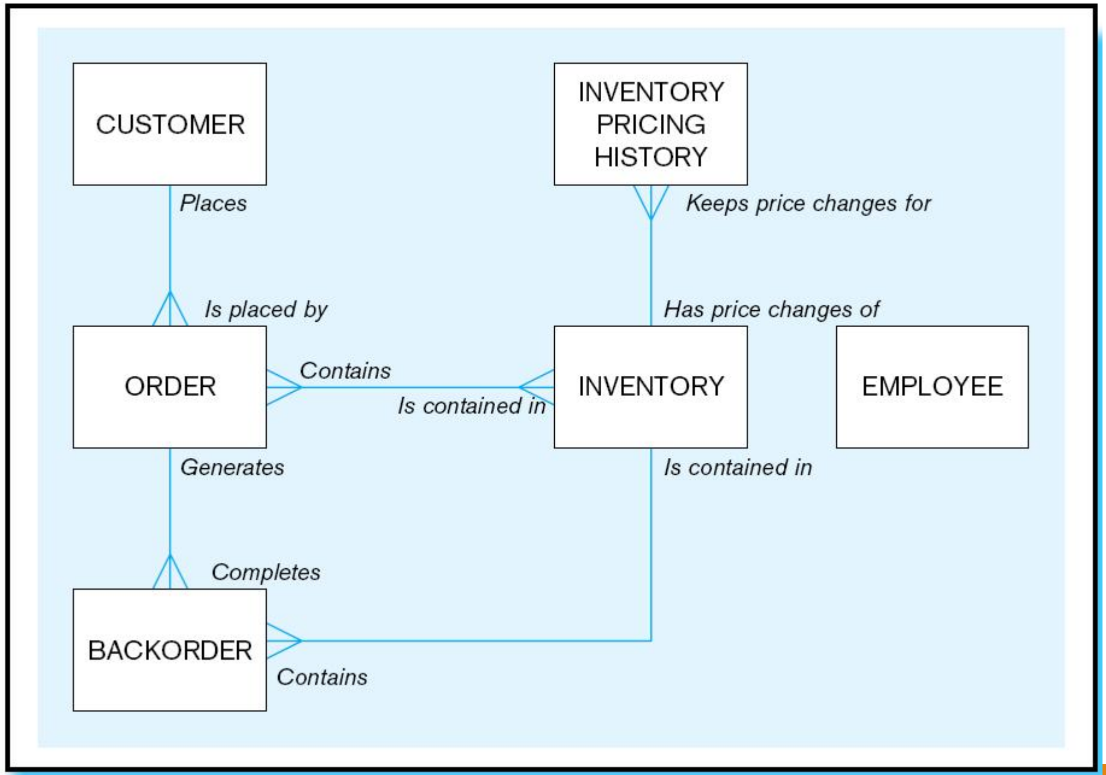
Order_ID	Customer_ID	Order_Date
201	1001	12/6/2021
202	1002	12/6/2021
203	1002	13/6/2021
…	…	…

File-based Approach VS Database Approach

FILE-BASED APPROACH



DATABASE APPROACH

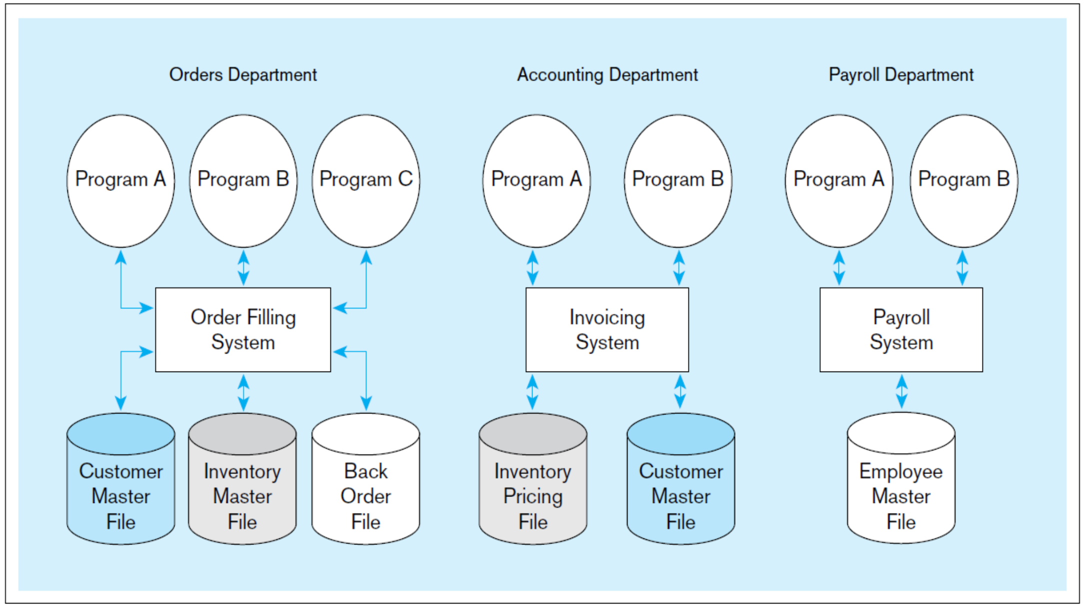


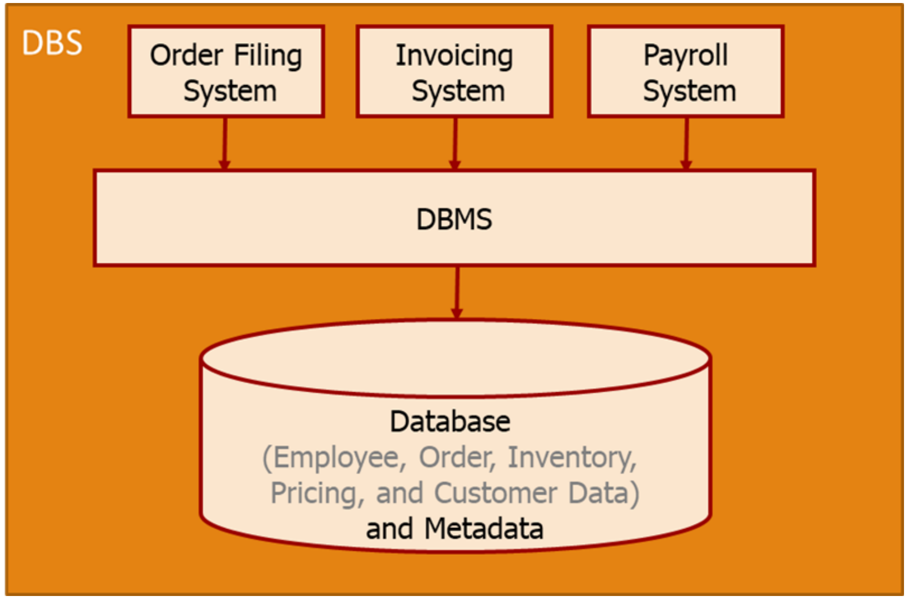
Advantages of the Database Approach

- ❑ Program-data **in**dependence
- ❑ Handling 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

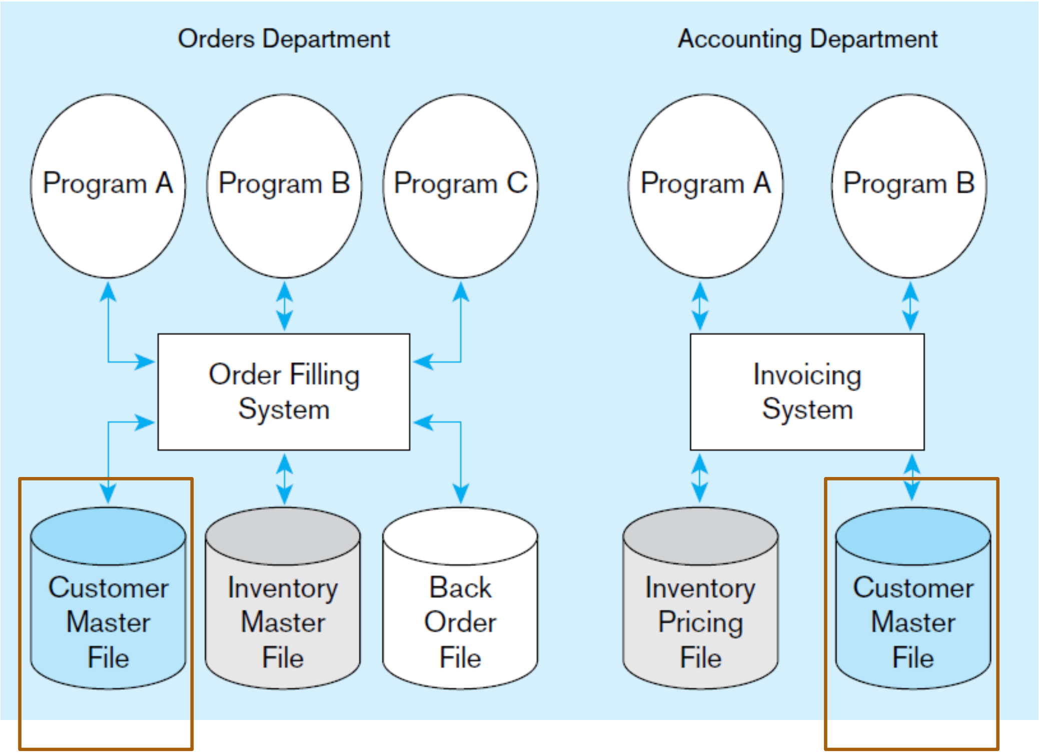
Advantages of the Database Approach: Program-data independence

- **Separate data descriptions** from the **application programs** that use the data.
  - **RESULT:** Allow changing and evolving (within limits) of an organization's data **without changing** the application programs that process the data (no need to change the program's code).





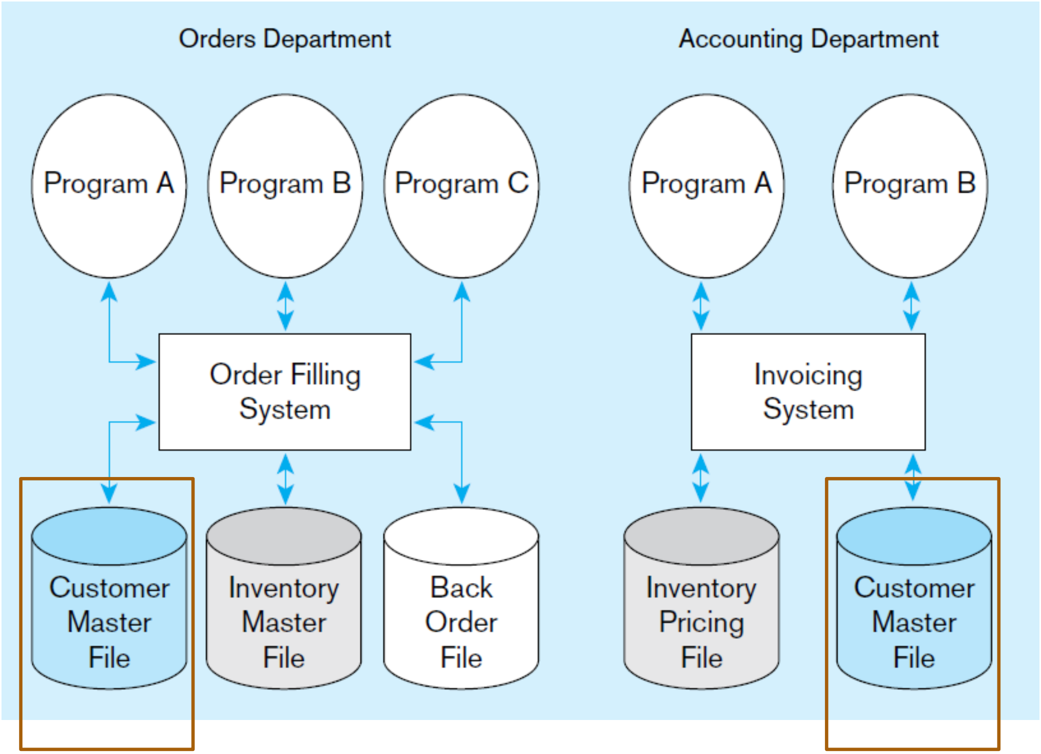
Advantages of the Database Approach: Handling data redundancy



- ❑ Good database design attempts to **integrate** previously separate (and redundant) data files **into a single, logical structure.**
  - Ideally, each primary fact is **recorded in only one place** in the database.

- ❑ Example,
  - For a file processing system, changing a Customer's address (John's address) must be done in **all** Customer Master's **files**.

Advantages of the Database Approach: Improved data consistency

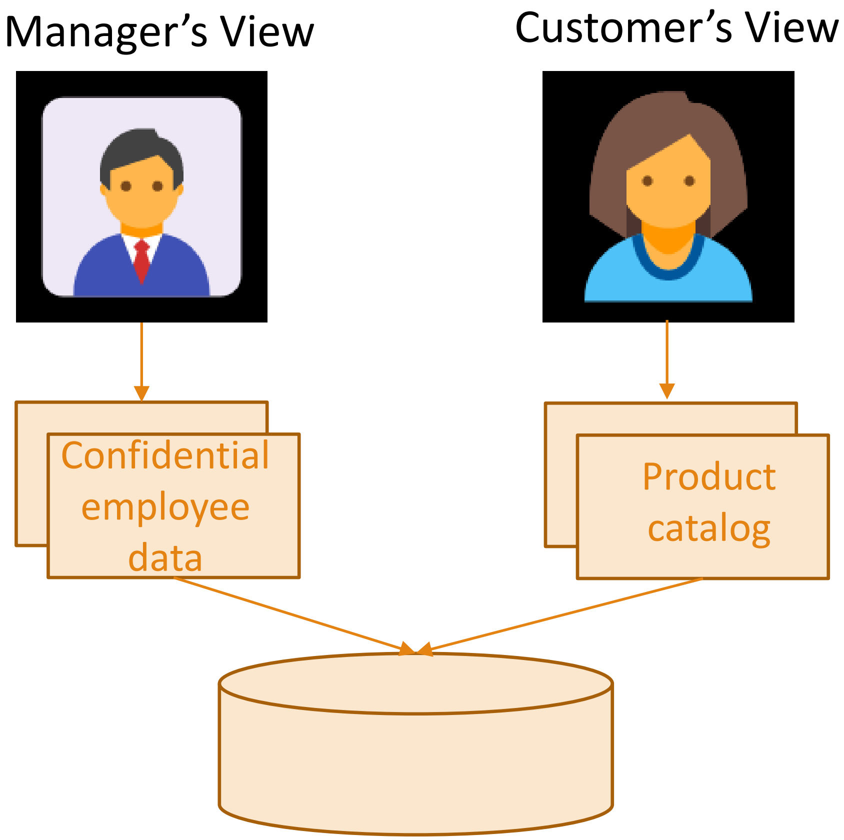


- ❑ By eliminating data redundancy, it greatly **reduces** the opportunities for **data inconsistency.**

- ❑Example,
  - For the database approach, John's address will be **stored in a single place**, therefore, **changing** his address is **easy** and **consistent**.

- ❑ Can also **avoid waste of storage space** that results from redundant data storage.

Advantages of the Database Approach: Improved data sharing

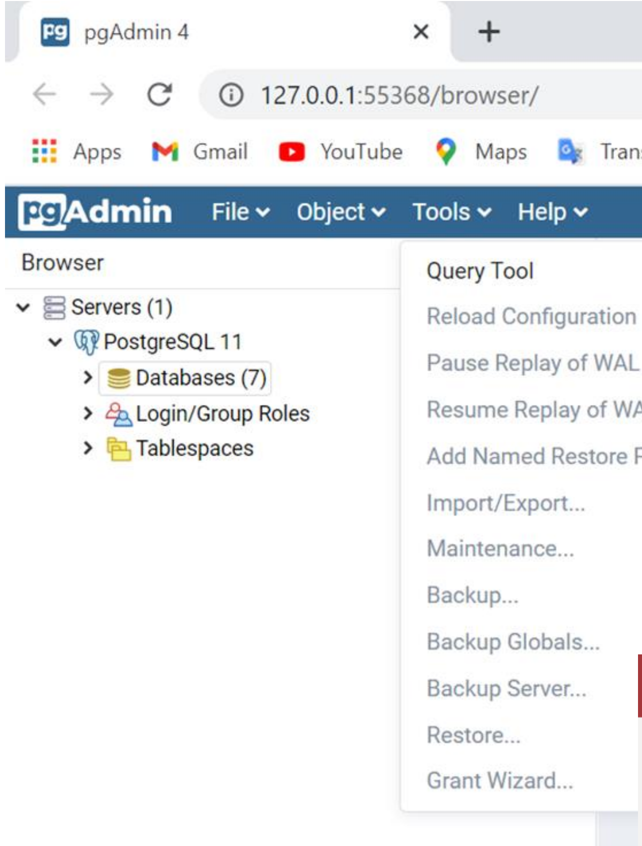


- ❑ Allow for granting permission to internal/external users to use the database
  - ❑ Each user (or group of users) is provided one or more **user views** into the database to facilitate this use.

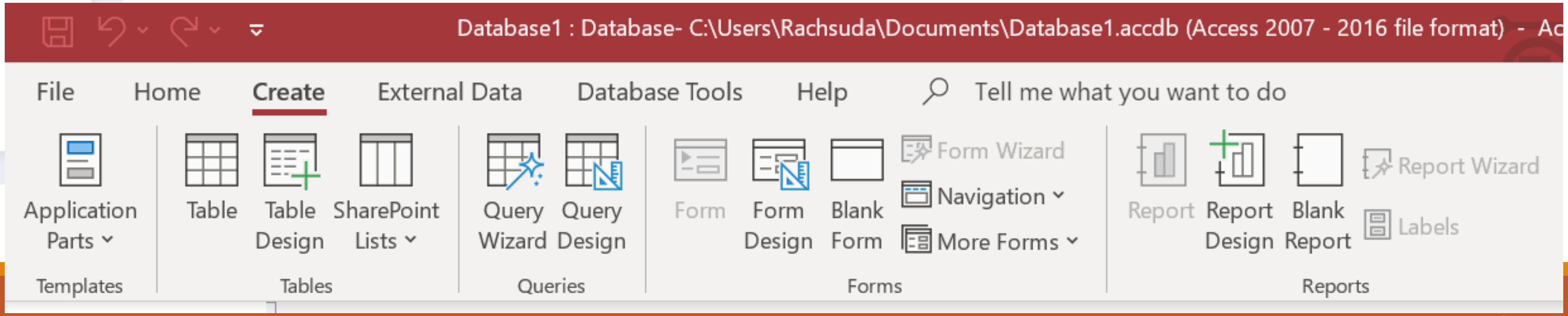
- ❑ **A user view**
  - A **logical description** of **some portion of the database** that is required by a user to perform some task (identified in form of a **form** or **report**)

Advantages of the Database Approach:

Increased application development productivity

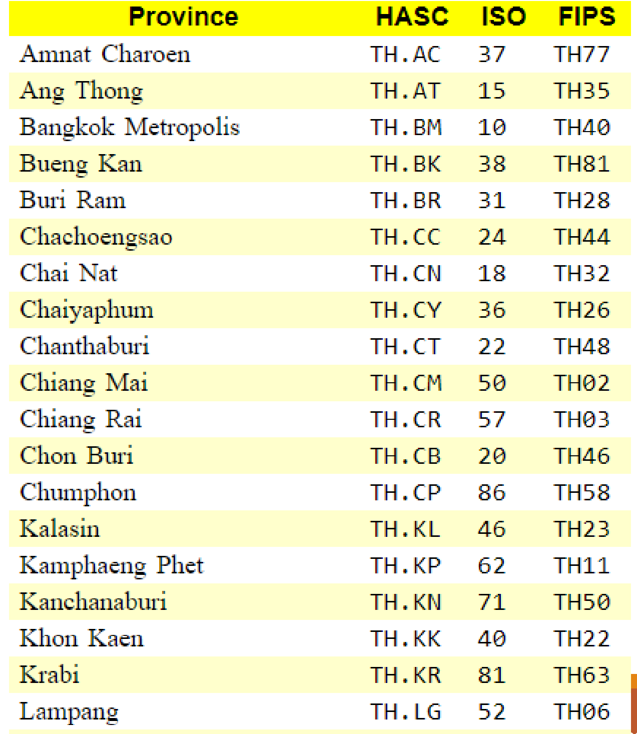


- ❑ Can greatly reduce the cost and time for developing new business applications;
  - The app developer can **focus** on the specific functions of the new application
    - ○ No worry about file design (or low-level implementation details)
  - DBMS provides a number of high-level productivity tools,
    - ○ E.g., forms and report generators, SQL query tool, backup and recovery tool



Advantages of the Database Approach:
Enforcement of standards

- **Example of Province's Code:**
  - **HASC:** Hierarchical administrative subdivision codes.
  - **ISO:** Codes from ISO 3166-2. For full identification in a global context, prefix "TH-" to the
    code (ex: TH-63 represents Tak).
  - **FIPS:** Codes from FIPS PUB 10-4.



Province	HASC	ISO	FIPS
Amnat Charoen	TH.AC	37	TH77
Ang Thong	TH.AT	15	TH35
Bangkok Metropolis	TH.BM	10	TH40
Bueng Kan	TH.BK	38	TH81
Buri Ram	TH.BR	31	TH28
Chachoengsao	TH.CC	24	TH44
Chai Nat	TH.CN	18	TH32
Chaiyaphum	TH.CY	36	TH26
Chanthaburi	TH.CT	22	TH48
Chiang Mai	TH.CM	50	TH02
Chiang Rai	TH.CR	57	TH03
Chon Buri	TH.CB	20	TH46
Chumphon	TH.CP	86	TH58
Kalasin	TH.KL	46	TH23
Kamphaeng Phet	TH.KP	62	TH11
Kanchanaburi	TH.KN	71	TH50
Khon Kaen	TH.KK	40	TH22
Krabi	TH.KR	81	TH63
Lampang	TH.LG	52	TH06

❑ The data repository provides database administrators with a powerful set of tools for developing and enforcing data standards.

❑ **Examples of data standards:**

- Naming conventions
- Data quality standards (an agreement on the representation, format, and definition for common data)
- Uniform procedures for accessing, updating, and protecting data.

Advantages of the Database Approach: Improved data quality

□ The database approach provides a number of tools and processes to improve data quality.

□ Database designers can specify integrity constraints (rules) that are enforced by the DBMS.

- E.g., An online order cannot be placed if a customer hasn’t registered for and account.

CENTRAL RETAIL

JD CENTRAL
ช้อปของดี การันตีของแท้

The1

ROBINSON

Advantages of the Database Approach:
Improved data accessibility and responsiveness

- The basic structure of the SQL query is easy to use.

- An example of SQL command to display information about *computer desks* at Pine Valley Furniture Company:

```
SELECT *
FROM Product_T
WHERE ProductDescription = “Computer Desk”;
```

Advantages of the Database Approach: Reduced program maintenance

- ❑ Can change either the **data** or the **application programs** (within limits) **without** necessitating a change in the other factor. (Independence of Data and Program)
- ❑ An example, **Changing** the **length of year from 2 digits to 4 digits** due to the “Y2K” problem (logical error(s) arising upon "rollover" from xx99 to xx00).

Ref: https://en.wikipedia.org/wiki/Year_2000_problem

Advantages of the Database Approach: Improved decision support

❑ Instantly acquire the summary data based on user’s demands.

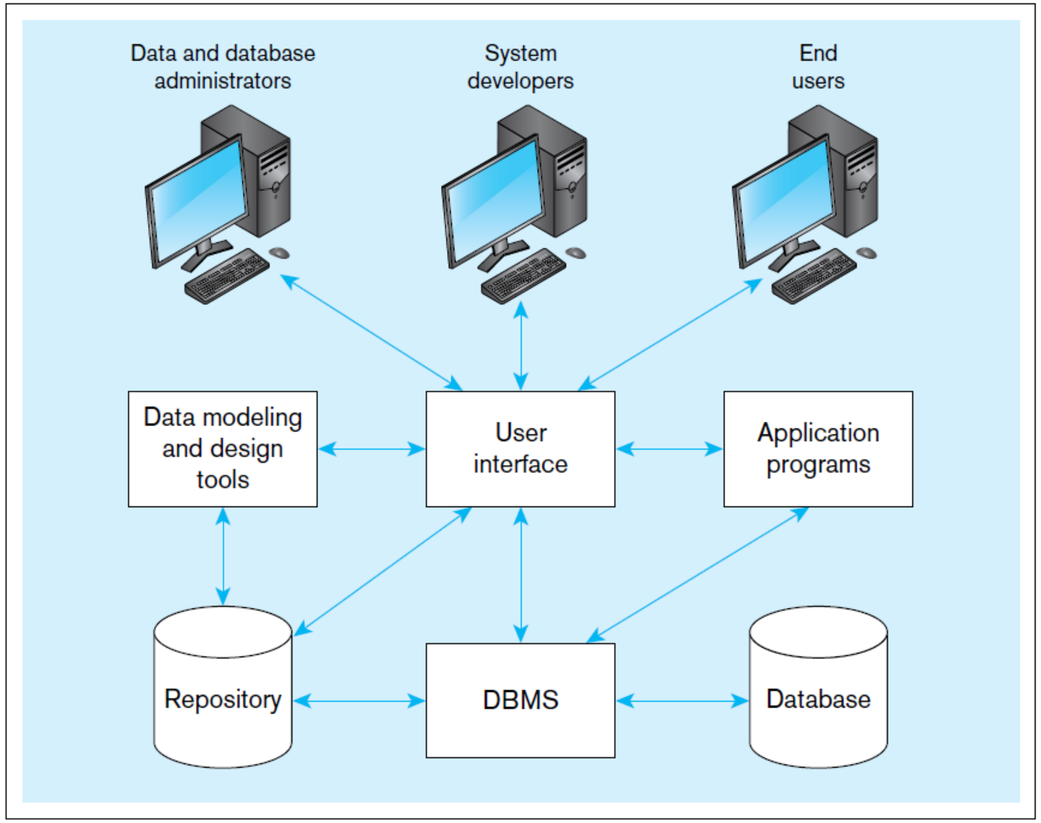
❑ Examples,

- Acquire the top-10 best selling products of the company and the quantities sold.
- Acquire the sale’s monthly report of all products in the branch X.

Costs and Risks of the Database Approach

- ❑ New, specialized personnel
  - Need to train/hire individuals to design and implement, maintain databases and manage people.
  - Also need to retrain them to update technology
- ❑ Installation and management cost and complexity
  - A multiuser DBMS is a large and complex suite of software that has **a high initial cost**,
  - It requires a staff of trained personnel to install and operate and has substantial annual maintenance costs.
- ❑ Conversion costs (changing from file processing systems to database applications)
- ❑ Need for explicit backup and recovery
- ❑ Organizational conflicts
  - A shared database requires a consensus on data definitions and ownership, as well as responsibilities for accurate data maintenance.
    - E.g., conflicts on data definitions, data formats and coding, rights to update shared data, and associated issues are frequent
  - To handle the conflicts, organizational commitment to the database approach is required.

Components of the Database Environment



- ❑ **Data modeling and design tools**
- ❑ **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

About the Term Project

❑ Design and Develop a Database Application for a chosen domain

- Possible Application Domains can be suggested and discussed later

❑ Design Documents in terms of E-R Diagram, Relation Schema Definition and Assumptions about Business Logic and Constraints made in pure English Description

❑ Implementation (using PostgreSQL or any other DBMS)

- SQL Schema Definition, Application Logic in SQL DML + Procedural Programming,
- Interface of your choice: Most probably Web Application Interface
- Need to demonstrate a complete WORKING APPLICATION

❑ Formulate a group of 3 or 4 members

Revised Questions

- What are the problems of managing large data directly on top of OS file systems?
- What are advantages of databases?
- What are costs and risks of databases?
- What are components of database environment?

Lecture 1: Part II:
The Database Development Process

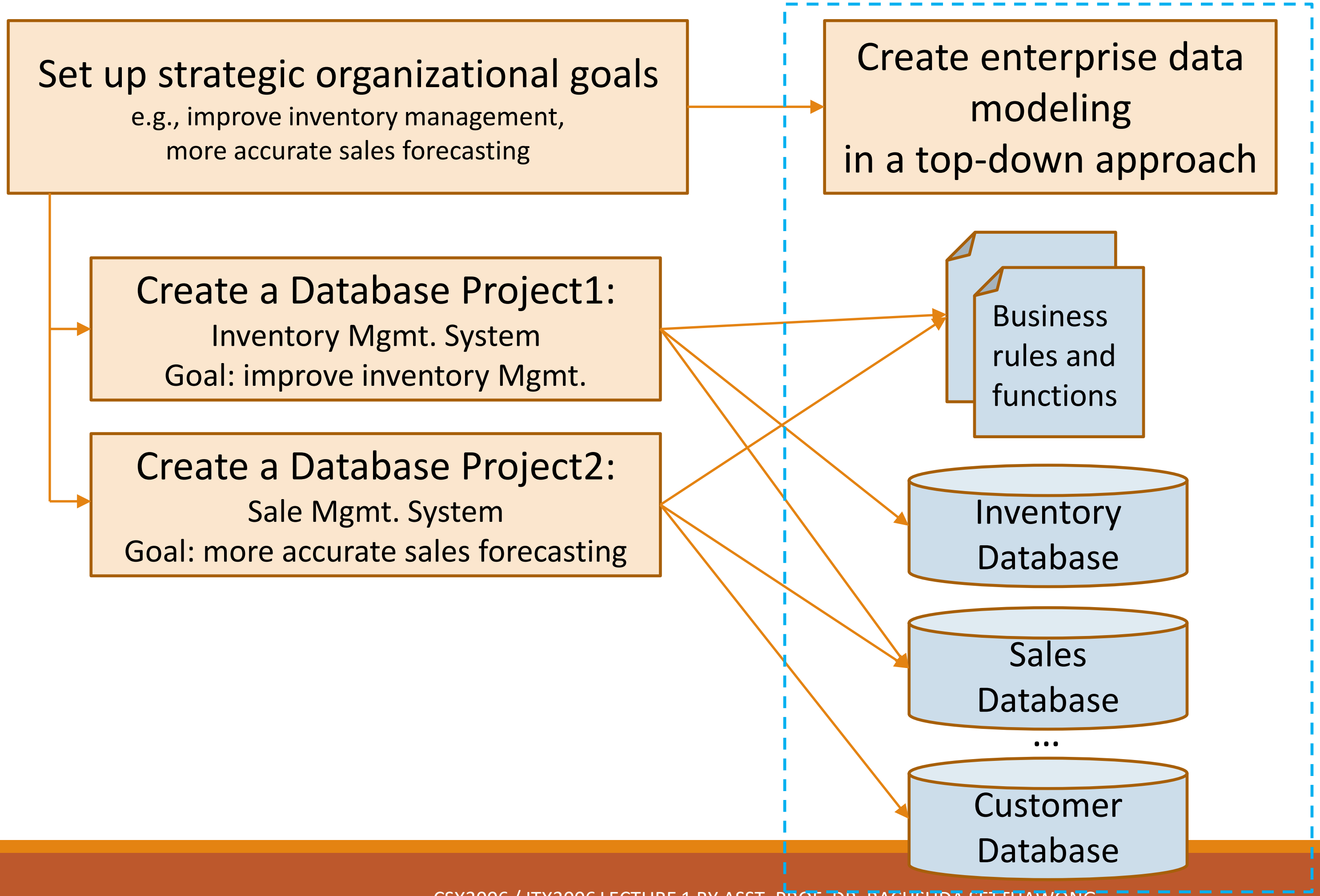
Outline

- ❑ The Database Development Process
- ❑ Systems Development Life Cycle
- ❑ Alternative Information Systems (IS) Development Approaches
- ❑ Three-Schema Architecture for Database Development
- ❑ Roles of People Involved in Database Development
- ❑ The Range of Database Applications
- ❑ Database Evolution at Pine Valley Furniture Company

Objectives

- ❑ Describe the life cycle of a systems development project, with an emphasis on the purpose of database analysis, design, and implementation activities
- ❑ Explain the prototyping and agile-development approaches to database and application development
- ❑ Explain the roles of individuals who design, implement, use, and administer databases
- ❑ Identify four categories of applications that use databases and their key characteristics
- ❑ Explain the differences among external, conceptual, and internal schemas

Database Projects as Parts of an Information Systems Planning and Development Project



Set up strategic organizational goals
e.g., improve inventory management,
more accurate sales forecasting

Create enterprise data modeling
in a top-down approach

Create a Database Project1:
Inventory Mgmt. System
Goal: improve inventory Mgmt.

Create a Database Project2:
Sale Mgmt. System
Goal: more accurate sales forecasting

Database Projects as Parts of an Information Systems Planning and Development Project – *Cont.*

□ **Enterprise data modeling** is a component of **a top-down approach** to information systems planning and development.

- It is one source of database projects.
  - Each projects often develops new databases to meet strategic organizational goals.

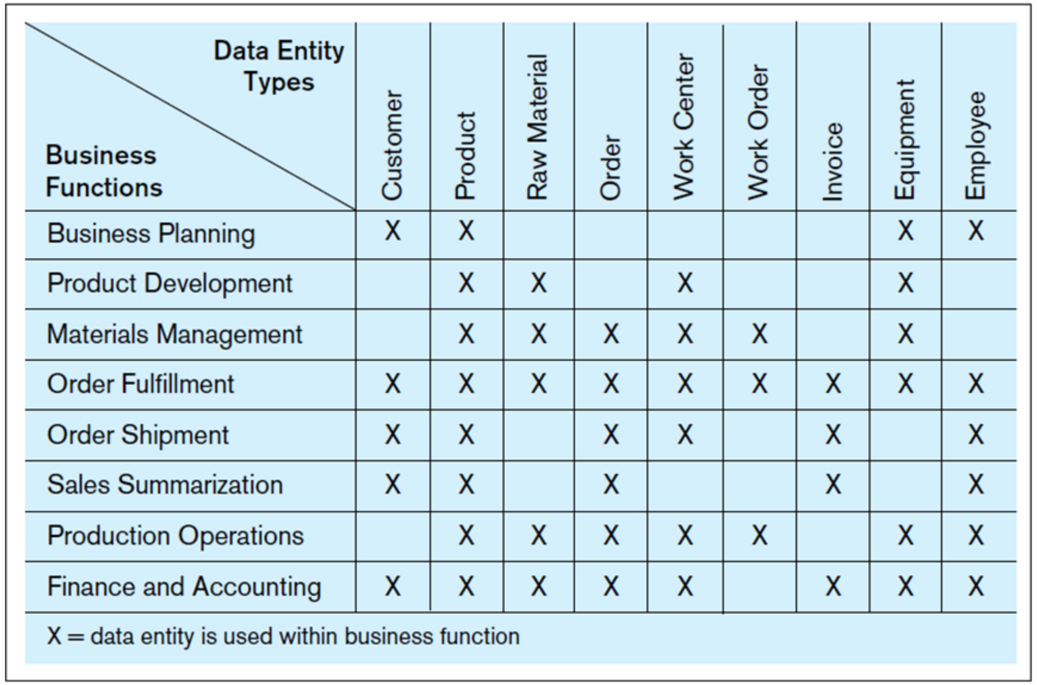
□ **Database projects** usually arise in a bottom-up fashion.

- □ Usually focuses on the creation of one database
- ○ For instance,
  - ○ Information systems users need certain information to do their jobs.
  - ○ Systems professionals see a need to improve data management in the organization.

□ A database and the associated information processing functions **are developed together** as part of **a comprehensive information systems development project**

The Database Development Process – 1

An Example of Business Function-to-Data Entity Matrix



Data Entity Types / Business Functions	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

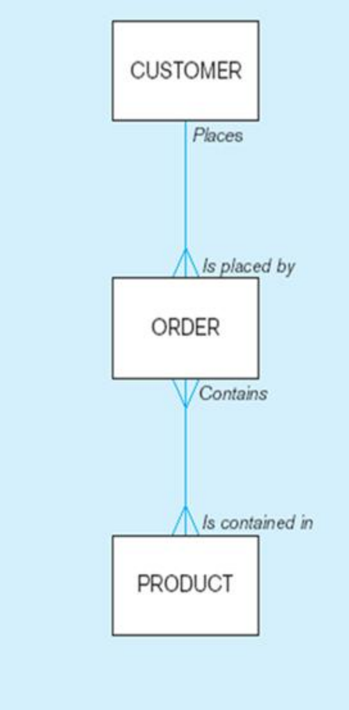
X = data entity is used within business function

❑ **Enterprise data modeling*** is the first step to specify the scope and general contents of organization databases.

1. Review and analyze current systems to obtain
   - Business rules (summary of how the business operates)
   - Business functions
   - Data (often captured using matrixes)

**: The model describes the scope of data maintained by the organization (may encompass many databases)*

The Database Development Process – 2



Segment of an enterprise data model

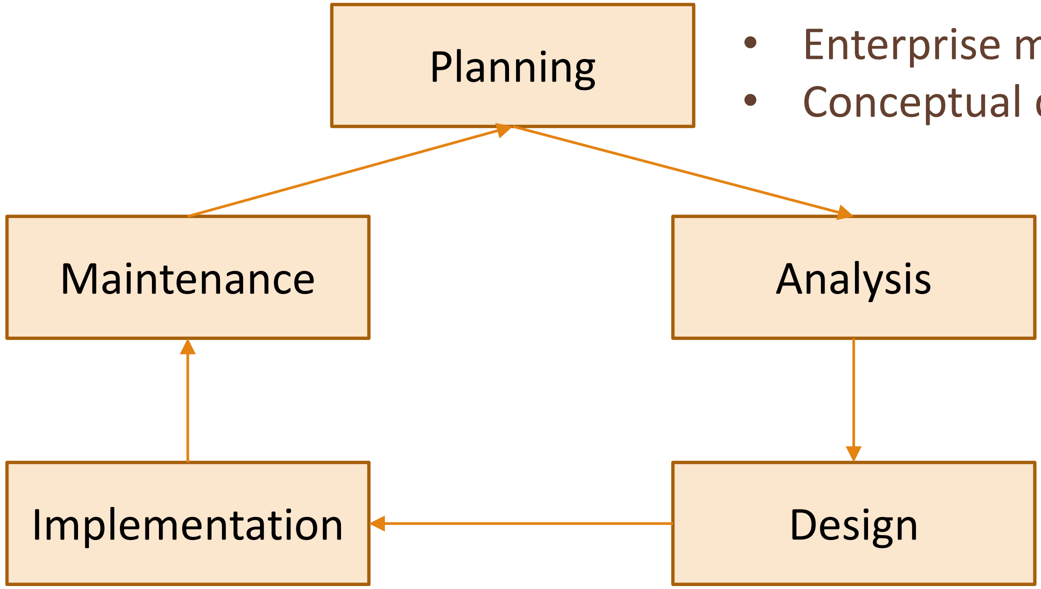
❑ **Enterprise data modeling*** is the first step to specify the scope and general contents of organization databases.

1. Review and analyze current systems
2. Describe the data needed at a very high level of abstraction
   - Business-oriented graphical representations (entity types) and descriptions
   - Business rules (Constraints)
   - Business functions (Relationships)
3. Plan one or more database development projects

**: The model describes the scope of data maintained by the organization (may encompass many databases)*

Systems Development Life Cycle (SDLC)

- ❑ A traditional process for conducting an information systems development project



- Enterprise modeling
- Conceptual data modeling

- Conceptual data modeling, *cont.*

- Logical database design
- Physical database design and definition

- Database implementation

- Database maintenance

Systems Development Life Cycle (SDLC): Planning

Purpose: to develop a preliminary understanding of

- A business situation
- How information systems might help solve a problem

Outcome: Enterprise and conceptual schemas (*Enterprise level*)

Enterprise modeling and conceptual data modeling (*Enterprise level*)

- Analyze current **data processing** and the general **business functions** and the needs of database
- Justify **need for new data and databases** in support of business
- Identify **scope of database requirements** for a *proposed* information system
- Analyze **overall data requirements** for business function(s) supported by database

Systems Development Life Cycle (SDLC): Analysis

Purpose:

- To analyze the business situation thoroughly
- To determine and structure requirements
- To select among competing system features

Outcome: a Conceptual Schema (*Detail level*)

Conceptual data modeling (*Detail level*)

- Produce a detailed data model that identifies all the organizational data that must be managed for this information system
  - The data model includes all entities, relationships, attributes, and business rules.

Systems Development Life Cycle (SDLC): Design

Purpose:

- To elicit and structure all information requirements;
- To develop all technology and organizational specifications

Outcome: a Logical Schema

Logical database design

- Analyze in detail the transactions, forms, displays, and inquiries (database views) required by the business functions supported by the database
- Integrate database views into conceptual data model
- Identify data integrity and security requirements, and populate repository
- Transform the conceptual schema into a logical schema
  - A relational model: tables, columns, rows, primary keys, foreign keys, and constraints
- Normalization

Systems Development Life Cycle (SDLC):
Design – *Cont.*

Purpose:

- To elicit and structure all information requirements;
- To develop all technology and organizational specifications

Outcome: a Physical Schema

Physical database design and definition

- Define database to DBMS (often generated from repository)
- Decide on physical organization of data
- Design an outline of programs to process and manage data

Systems Development Life Cycle (SDLC): Implementation

Purpose:

- To write programs
- To build databases
- To test and install the new system
- To train users
- To finalize documentation

Outcome: a Database System

Database implementation

- Code and test database processing programs
- Complete database documentation and training materials
- Install database and convert data from prior systems

Systems Development Life Cycle (SDLC): Maintenance

Purpose:

- To monitor the operation and usefulness of the system
- To repair and enhance the system

Outcome: an Updated Database System

Database maintenance

- Analyze database and database applications to ensure that evolving information requirements are met
- Tune database for improved performance
- Fix errors in database and database applications and recover database when it is contaminated

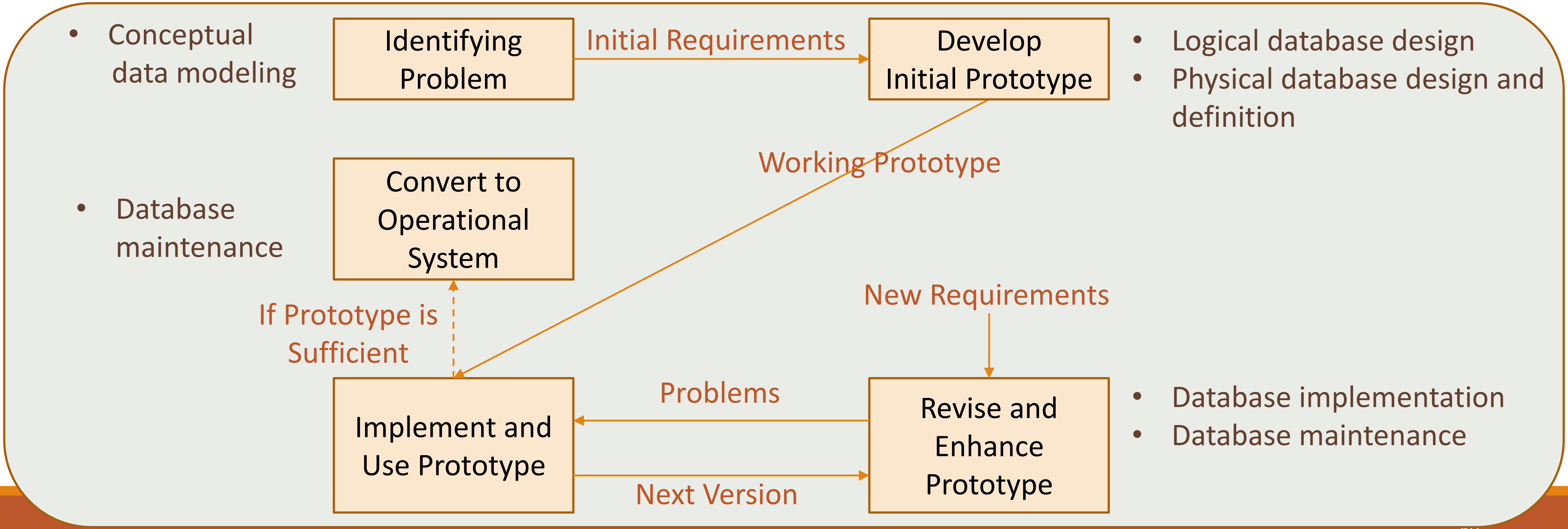
Alternative Information Systems (IS) Development Approaches:
Rapid Application Development (RAD) methods

- ❑ Is an iterative process of ***rapidly repeating*** analysis, design, and implementation steps ***until*** they converge on the system the user wants.

- ❑ Work best when most of the necessary database structures already exist, and hence for systems that primarily retrieve data

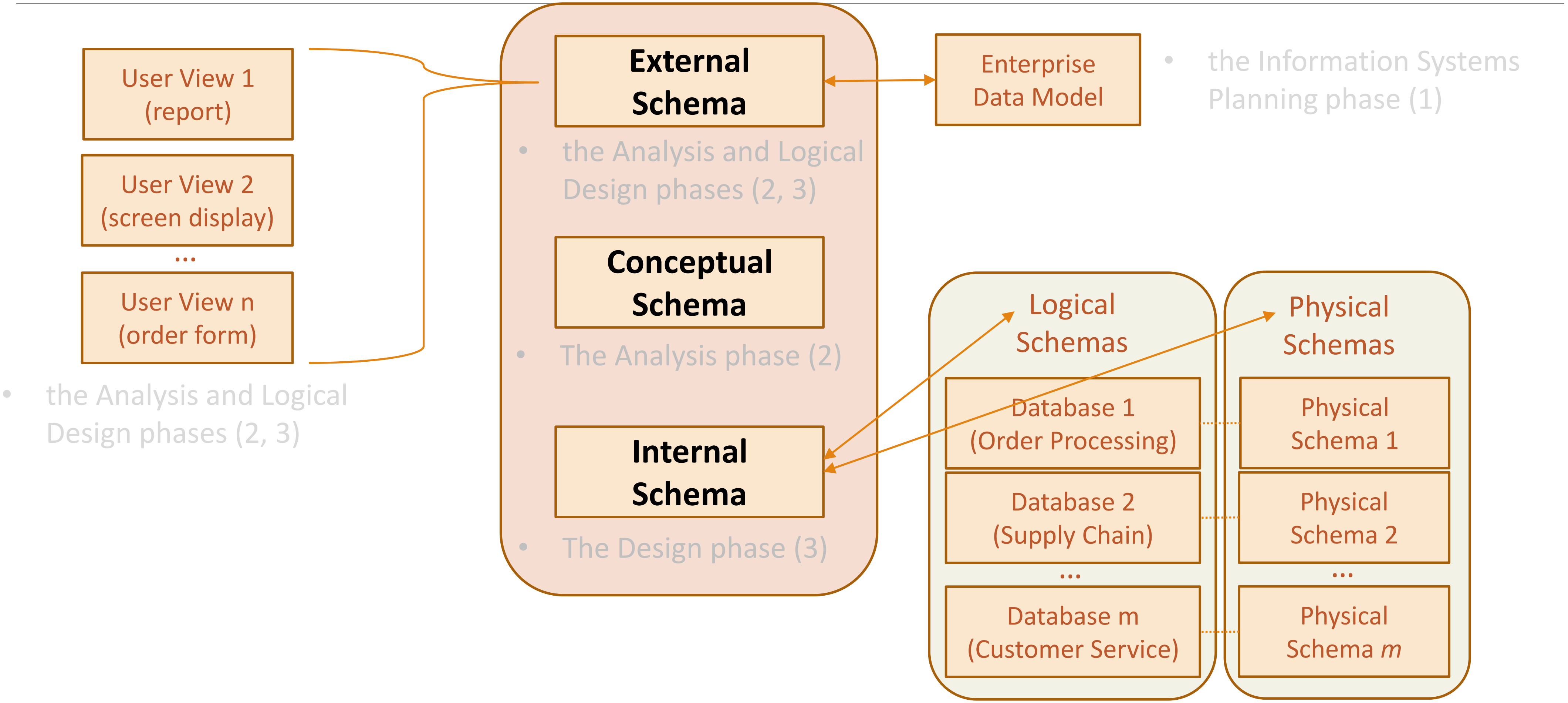
One of the most popular RAD methods:
Prototyping

- An iterative process of systems development
  - ***Requirements are converted to a working system*** that is ***continually revised*** through close work between analysts and users.



Three-Schema Architecture for Database Development

(for describing the structure of data)



User View 1 (report)

User View 2 (screen display)

…

User View n (order form)

- the Analysis and Logical Design phases (2, 3)

External Schema

- the Analysis and Logical Design phases (2, 3)

Conceptual Schema

- The Analysis phase (2)

Internal Schema

- The Design phase (3)

Enterprise Data Model

- the Information Systems Planning phase (1)

Logical Schemas

Database 1 (Order Processing)

Database 2 (Supply Chain)

…

Database m (Customer Service)

Physical Schemas

Physical Schema 1

Physical Schema 2

…

Physical Schema *m*

Three-Schema Architecture for Database Development – Cont.

(for describing the structure of data)

❑ External schema

- Is the **view(s)** of the **database users** (e.g., managers and employees.)
- Be represented as a combination of **the enterprise data model** (a top-down view) and a collection of detailed (or bottom-up) **user views**.

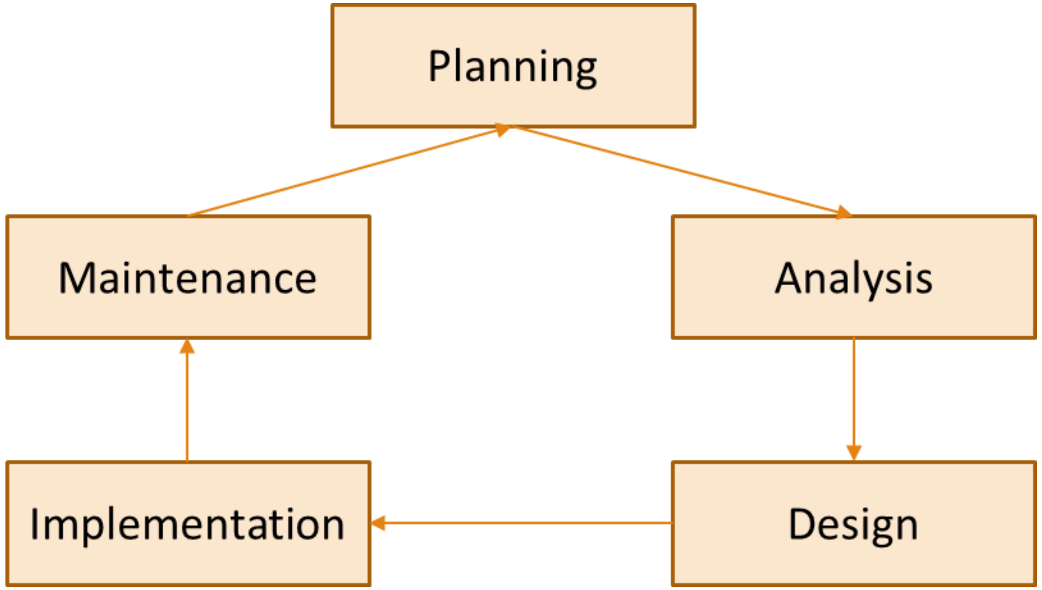
❑ Conceptual schema

- Represents the **view** of the **data architect** or **data administrator**.
- Combines the different external views into a single, coherent, and comprehensive definition of the **enterprise's data**.

❑ Internal schema

- **The logical schema** is the **representation of data** for a *type of data management technology* (e.g., **relational**).
- **The physical schema** describes **how data are to be represented and stored** in secondary storage using a particular DBMS (e.g., Oracle).

A Project Team for System/Database Development



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

The project leader is responsible for selecting and managing all of these people as an effective team

Roles in a Project Team

❑ **Project managers**

- Oversee assigned projects, including team composition, analysis, design, implementation, and support of projects.

❑ **Business analysts**

- Work with both management and users to analyze the business situation and develop detailed system and program specifications for projects

❑ **Systems analysts**

- Perform business analyst activities
- Specify computer systems requirements
- Typically have a stronger systems development background than business analysts

Roles in a Project Team – Cont.

❑ **Database analysts and data modelers**

- Determine the requirements and design for the database component of the information system

❑ **Users**

- Provide assessments of their information needs
- Monitor that the developed system meets their needs.

❑ **Database architects**

- Establish standards for data in business units
- Attain optimum data location, currency, and quality

Roles in a Project Team – Cont.

❑ Data administrators

- Be responsible for existing and future databases
- Ensure consistency and integrity across databases
- Provide consulting and training to other project team members

❑ Programmers

- Design and write computer programs that have commands to maintain and access data in the database embedded in them.

❑ Other technical experts

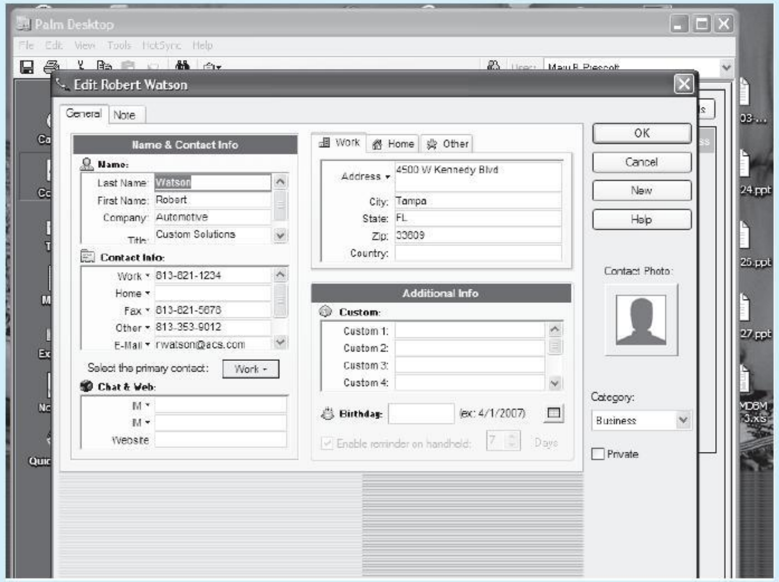
- Such as networking, operating systems, testing, data warehousing, etc.

The Range of Database Applications

- Personal databases
- Multitier Client/Server Databases
- Enterprise Applications

The Range of Database Applications:
Personal databases

Customer



Edit Robert Watson

Name & Contact Info

Name:
Last Name: Watson
First Name: Robert
Company: Automotive
Title: Custom Solutions

Contact Info:
Work: 813-821-1234
Home:
Fax: 813-821-5678
Other: 813-353-9012
E-Mail: rwatson@acs.com

Select the primary contact: Work

Chat & Web:
IM:
IM:
Website:

Address: 4500 W Kennedy Blvd
City: Tampa
State: FL
Zip: 33609
Country:

Additional Info

Custom:
Custom 1:
Custom 2:
Custom 3:
Custom 4:

Birthday: (ex: 4/1/2007)

☑ Enable reminder on handheld: 7 Days

Category: Business

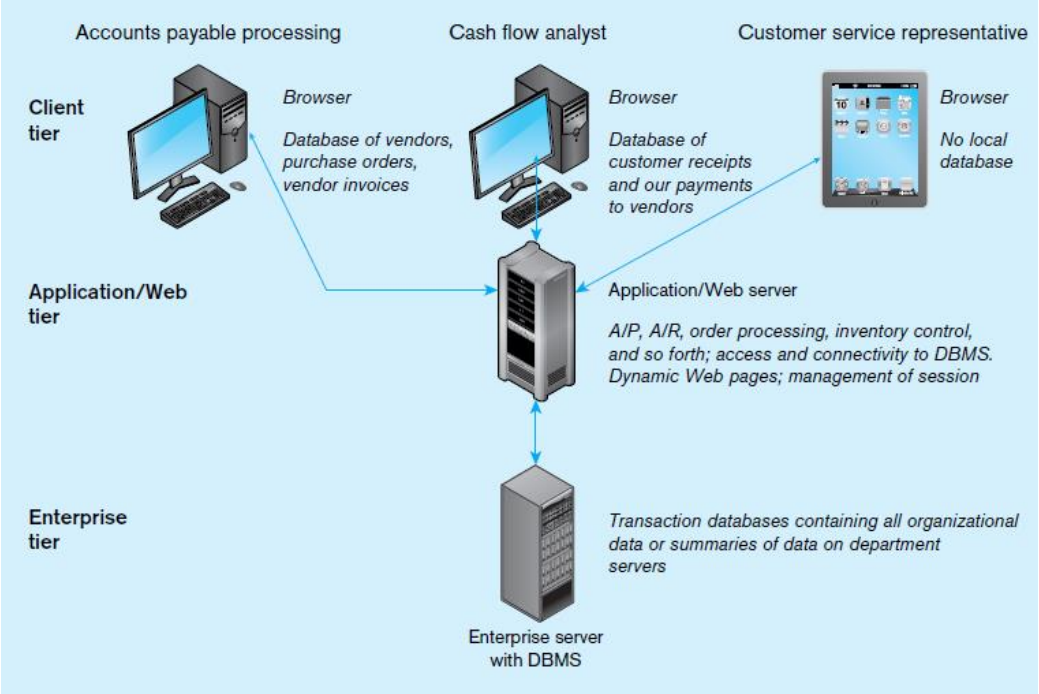
□ Private

Contact History for Customer

Date	Time	Contact	Comments
08/04/2008	10:00 AM	Roberts	Review proposal
08/19/2008	08:00 AM	Roberts	Revise schedule
09/10/2008	09:00 AM	Pearson	Sign contract
09/21/2008	02:00 PM	Roberts	Follow up

- **Purpose:** to provide the user with the ability to manage small amounts of data in an efficient manner
- **Cons:** not easy to share data

The Range of Database Applications:
Multitier Client/Server Databases


Accounts payable processing
Cash flow analyst
Customer service representative
Client tier
Browser
Database of vendors, purchase orders, vendor invoices
Browser
Database of customer receipts and our payments to vendors
Browser
No local database
Application/Web tier
Application/Web server
A/P, A/R, order processing, inventory control, and so forth; access and connectivity to DBMS. Dynamic Web pages; management of session
Enterprise tier
Transaction databases containing all organizational data or summaries of data on department servers
Enterprise server with DBMS

Multitiered client/server database architecture

❑ **Pros**

- Capable of sharing data among large number of users
- The ease of separating the development/ maintenance of the database from the information systems modules that focus on business logic and/or presentation logic
- Improve performance and maintainability of the application and database

The Range of Database Applications: Enterprise Applications (Backbone)

Purpose: support organization-wide operations and decision making.

Scope: the entire organization or enterprise (or, many different departments)

- ❑ The systems that keep an organization running
  - Consist of the processes that control and execute basic business tasks
  - Focus on capturing the data surrounding the "**transactions**," which define how a business is conducted
    - The hundreds or millions of events taken place in an organization every day
    - E.g., purchase order (E-commerce), registration records (University), payrolls, stock replenishment
- ❑ An organization may have several enterprise databases;
  - A single operational enterprise database is impractical for many medium to large organizations;
    - o Difficulties in performance for very large databases
    - o Diverse needs of different users
    - o The complexity of achieving a single definition of data (metadata) for all database users

Two Major Developments of Enterprise Databases

❑ Enterprise resource planning (ERP) systems

❑ Data warehousing implementations

Enterprise Resource Planning (ERP)

- ERP is a business management system that integrates **all functions** of the enterprise.
  - Commonly offered modules:
    - Human resources
    - CRM
    - Finance/Accounting
    - IT Helpdesk
    - eCommerce
    - Supply Chain Management
    - Order Processing
    - Inventory and Procurement
- ERP systems are software applications that provide the data necessary for the enterprise to examine and manage its activities.
  - Work with the current operational data of the enterprise

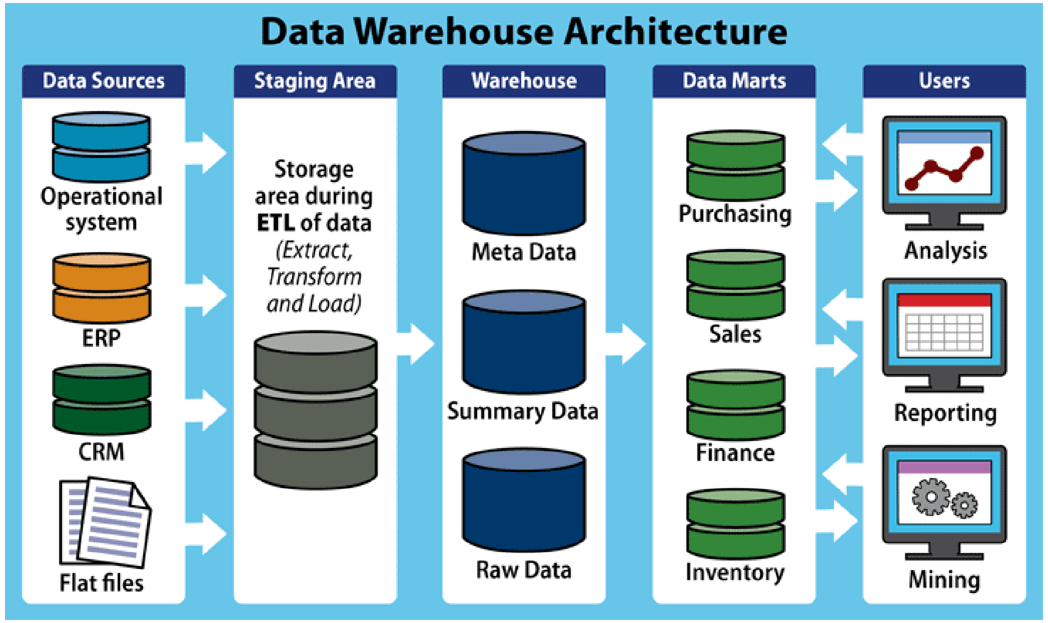
Ref: https://technologyadvice.com/erp/

Top ERP Vendors by Company Size

Enterprise	Medium-Sized	Small Business
SAP	Netsuite	Deltek
Oracle	Sage	Work(etc)
Microsoft Dynamics	Infor	Syspro
IFS Applications	Macola	Intacct

Source: https://technologyadvice.com/erp/

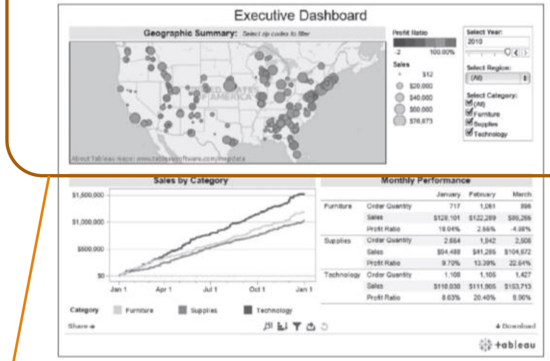
Data Warehouse



- ❑ An integrated decision support database, whose content is derived from the various *operational databases*, e.g., personal, workgroup, department, and ERP databases.
- ❑ Data warehouses provide users with the opportunity to work with historical data *to identify patterns and trends and answers to strategic business questions*.

Source: https://www.datamation.com/big-data/top-15-data-warehouse-tools/

An example of an executive dashboard



Executive Dashboard

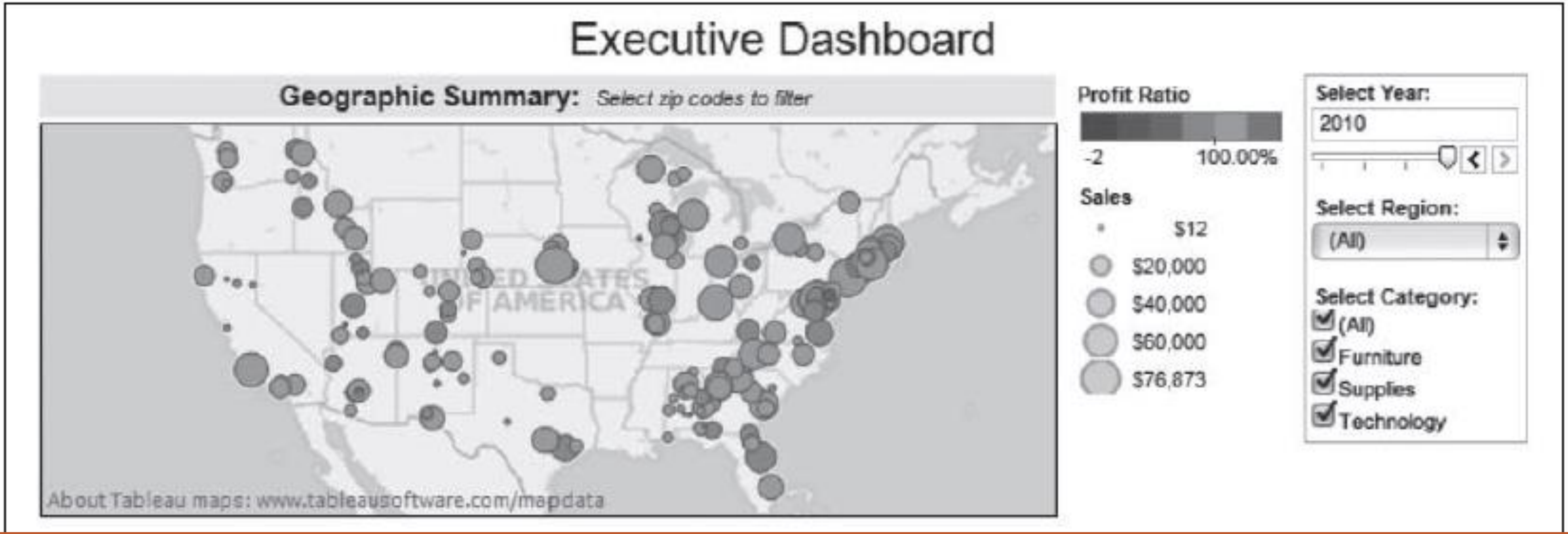
Select Year: 2010

Select Region: (All)

Select Category:
☑ (All)
☑ Furniture
☑ Supplies
☑ Technology

Monthly Performance

		January	February	March
Furniture	Order Quantity	717	1,061	898
	Sales	$128,101	$122,289	$86,266
	Profit Ratio	18.04%	2.56%	-4.98%
Supplies	Order Quantity	2,664	1,942	2,506
	Sales	$94,488	$41,285	$104,672
	Profit Ratio	9.70%	13.39%	22.64%
Technology	Order Quantity	1,108	1,105	1,427
	Sales	$118,030	$111,905	$153,713
	Profit Ratio	8.63%	20.40%	5.96%



Executive Dashboard

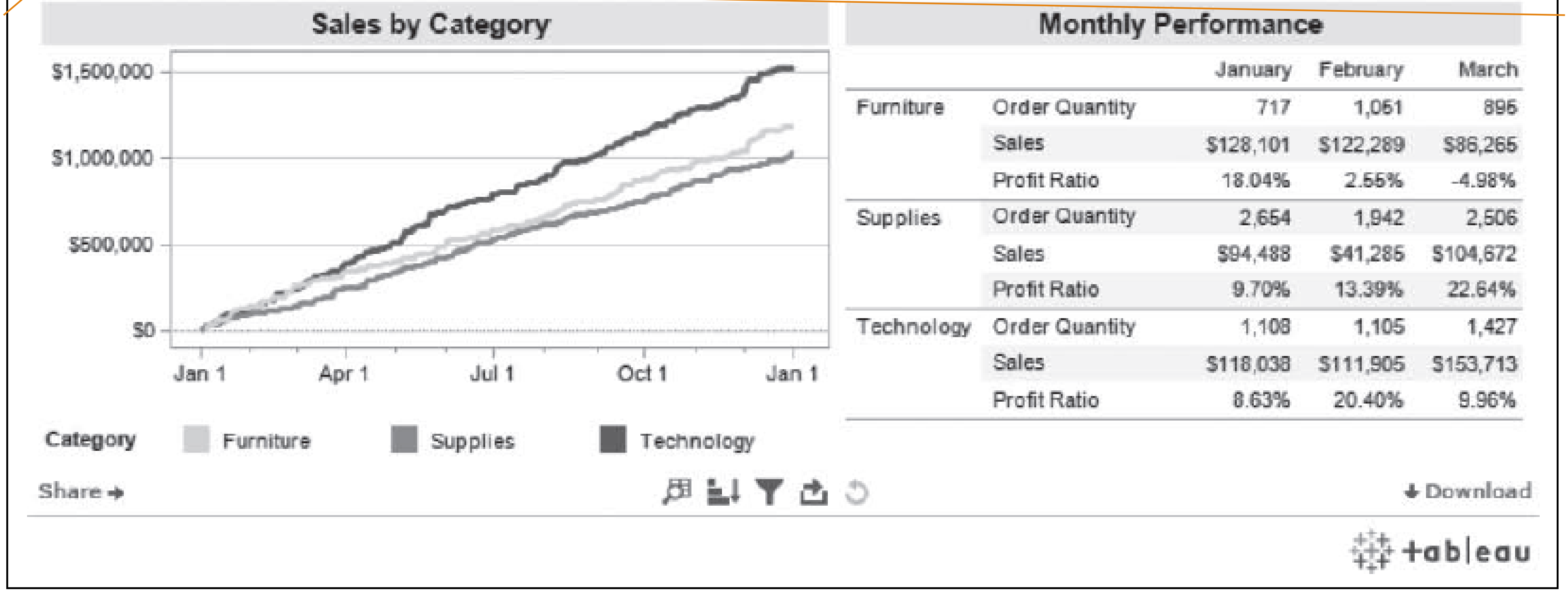
Geographic Summary: Select zip codes to filter
UNITED STATES OF AMERICA
About Tableau maps: www.tableausoftware.com/mapdata
Profit Ratio
-2
100.00%
Sales
$12
$20,000
$40,000
$60,000
$76,873

Select Year: 2010

Select Region: (All)

Select Category:
☑ (All)
☑ Furniture
☑ Supplies
☑ Technology

An example of an executive dashboard (Cont.)



Monthly Performance

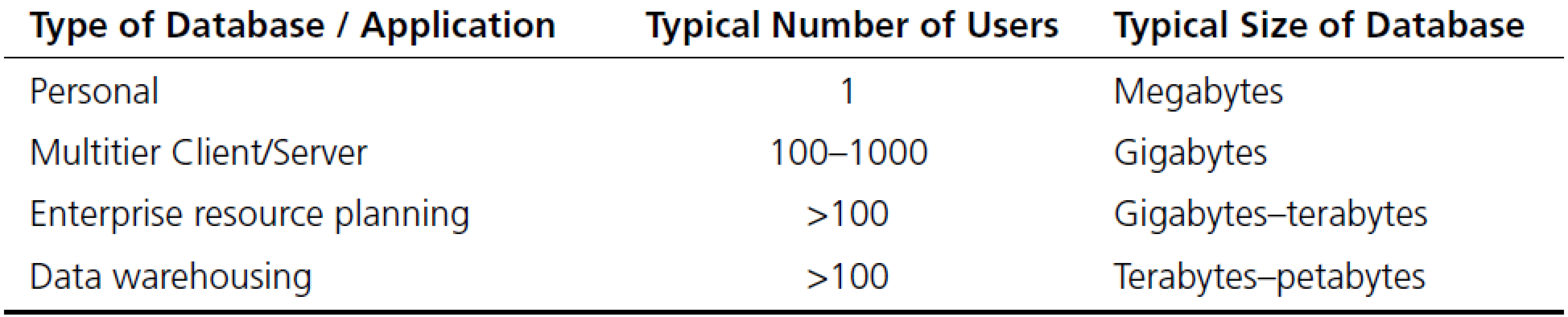
		January	February	March
Furniture	Order Quantity	717	1,061	896
	Sales	$128,101	$122,289	$86,265
	Profit Ratio	18.04%	2.55%	-4.98%
Supplies	Order Quantity	2,654	1,942	2,506
	Sales	$94,488	$41,285	$104,672
	Profit Ratio	9.70%	13.39%	22.64%
Technology	Order Quantity	1,108	1,105	1,427
	Sales	$118,038	$111,905	$153,713
	Profit Ratio	8.63%	20.40%	9.96%

Best Data Warehouse Software & Tools for 2021

- □ SAP Data Warehouse Cloud
- □ VMware Tanzu Greenplum
- □ Snowflake
- □ Panoply
- □ Teradata Vantage
- □ IBM
- □ Oracle Autonomous Data Warehouse
- □ Azure Synapse Analytics
- □ Amazon Redshift
- □ Google BigQuery

Ref: https://project-management.com/best-data-warehouse-software/

Summary of Database Applications



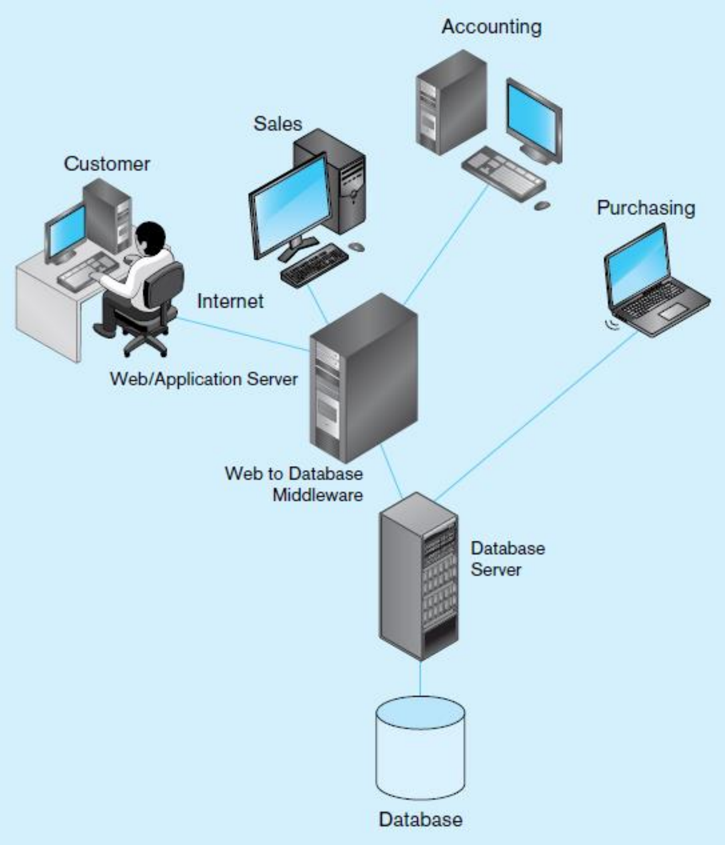
Type of Database / Application	Typical Number of Users	Typical Size of Database
Personal	1	Megabytes
Multitier Client/Server	100–1000	Gigabytes
Enterprise resource planning	>100	Gigabytes–terabytes
Data warehousing	>100	Terabytes–petabytes

Case Study: Developing A Database Application for Pine Valley Furniture Company – 1

The late 1990s: adopting a database approach for the company

- Integrating the data (previously stored in separate files) into a single database structure
- Compiling the metadata (to be the same structure)
- Using the **DBMS** to manage the data:
  - Provides the interface between the various database applications for organizational users and the database(s).
  - Allows users to share the data and to query
- Using a local area network (LAN) (to link employee workstations in the various departments to a database server)

Case Study: Developing A Database Application for Pine Valley Furniture Company – 2



Computer System for Pine Valley Furniture Company

The early 2000s, Mounting a two-phase effort to introduce Internet technology

1. **Installing an intranet** (to allow employees fast Web-based access to company information)
   - **Examples** of relating data accessed: phone directories, furniture design specifications, e-mail
2. **Adding a Web interface** to some of its business applications (e.g., order entry)
   - More internal business activities that require access to data in the database server could also be conducted by employees through its intranet

Remark:

- Most applications that use the database server still do not have a Web interface
- The applications are required to be stored on employees' workstations

Case Study: Developing A Database Application for Pine Valley Furniture Company – Database Evolution

Helen, a product manager

Goal: needed an efficient way to analyze sales of her products.

  - These analyses are ad hoc, driven by rapidly changing and unanticipated business conditions, comments from furniture store managers, trade industry gossip, or personal experience.

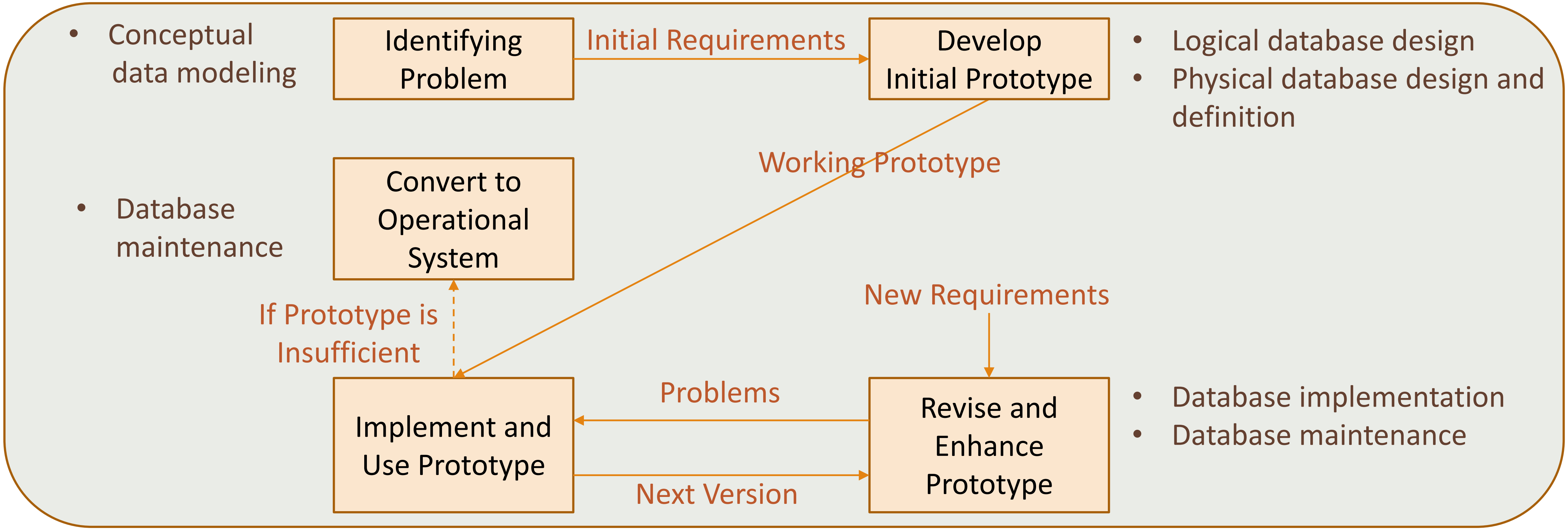
- **Scope**: a direct access to sales data with an easy-to-use interface (able to search for answers to the various marketing questions)

Chris, a systems analyst

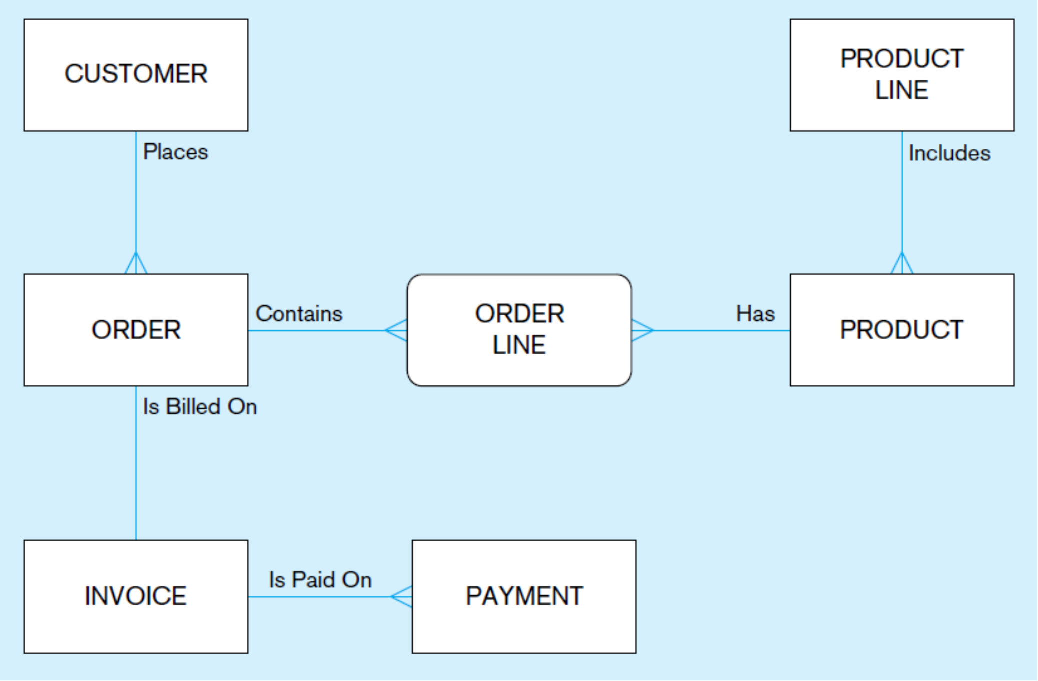
Analyzing the requirements, the current systems, and selecting the IS development approach:

- Have **existing databases** that support its operational business functions.
  - It can extract the data needs from existing databases.
- Decide to building **a new stand-alone databases** to serve Helen's needs
  - Therefore, the unstructured and unpredictable use of data will not interfere with the access to the operational databases
- Due to the needs for data analysis, a combination of **prototyping** and life-cycle approaches is applied in developing the requested system.
  - Choose to develop the system using **Microsoft Access**, Pine Valley's preferred technology for personal databases.
  - **Outcome:** The prototype will end up being the actual system.

A Prototyping Approach



Case Study: Developing A Database Application for Pine Valley Furniture Company – Project Planning



***Preliminary* data model** for
Home Office product line marketing support system

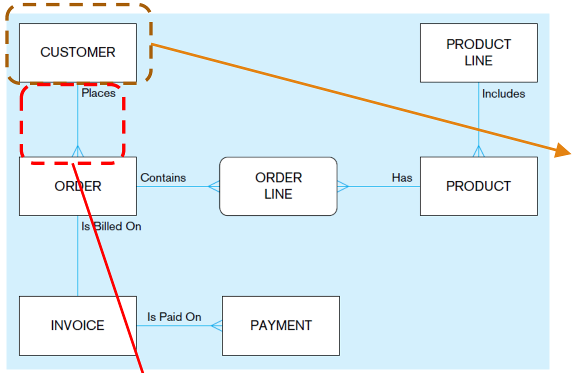
Interviewing and taking notes to have an understanding on Helen's business area

- **Details:** business area and its objectives, business functions, data entity types, other dealing business objects, dealing issues, requirements, stakeholders
- Ask general questions using business and marketing terminology as much as possible

Generate two quick analyses

- Identifies all of the databases that contain data associated with the data entities obtained, and relating data attributes
  - For example, the objective to exceed sales goals for each product finish category of office furniture suggests that Helen wants product annual sales goals in her system
- Draws a conceptual data model that represents the data entities with the associated data attributes, as well as the major relationships among these data entities.

Case Study: Developing A Database Application for Pine Valley Furniture Company – Analyzing Database Requirements



A Business Rule:
Each CUSTOMER Places any number of ORDERs. Conversely, each ORDER Is Placed By exactly one CUSTOMER.

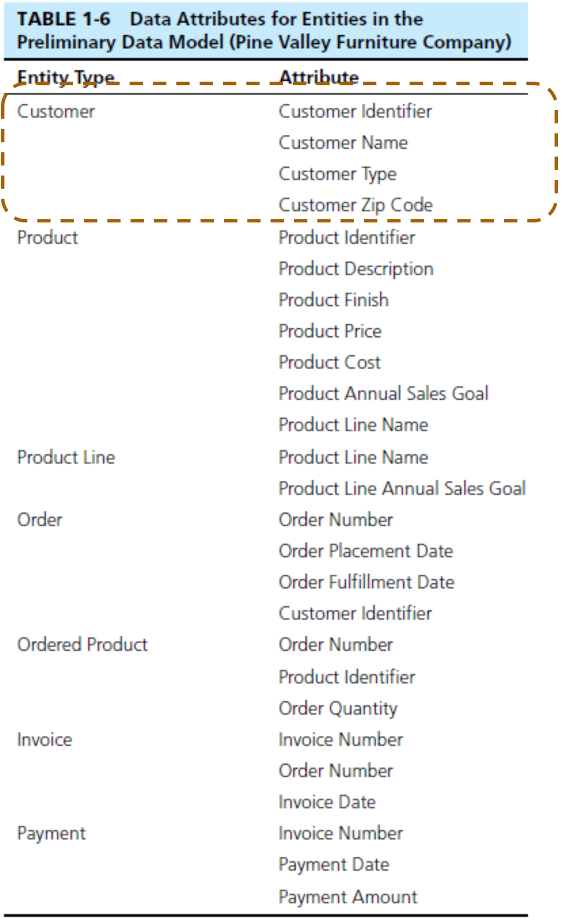


TABLE 1-6 Data Attributes for Entities in the Preliminary Data Model (Pine Valley Furniture Company)

Entity Type	Attribute
Customer	Customer Identifier
	Customer Name
	Customer Type
	Customer Zip Code
Product	Product Identifier
	Product Description
	Product Finish
	Product Price
	Product Cost
	Product Annual Sales Goal
	Product Line Name
Product Line	Product Line Name
	Product Line Annual Sales Goal
Order	Order Number
	Order Placement Date
	Order Fulfillment Date
	Customer Identifier
Ordered Product	Order Number
	Product Identifier
	Order Quantity
Invoice	Invoice Number
	Order Number
	Invoice Date
Payment	Invoice Number
	Payment Date
	Payment Amount

Walks through each data entity in his initial ideas for the database application

- Explaining what it means and what business policies and procedures are represented by each line between entities
- Paying close attention to Helen’s reactions
- Altering the models wrt the feedback.

BEFORE

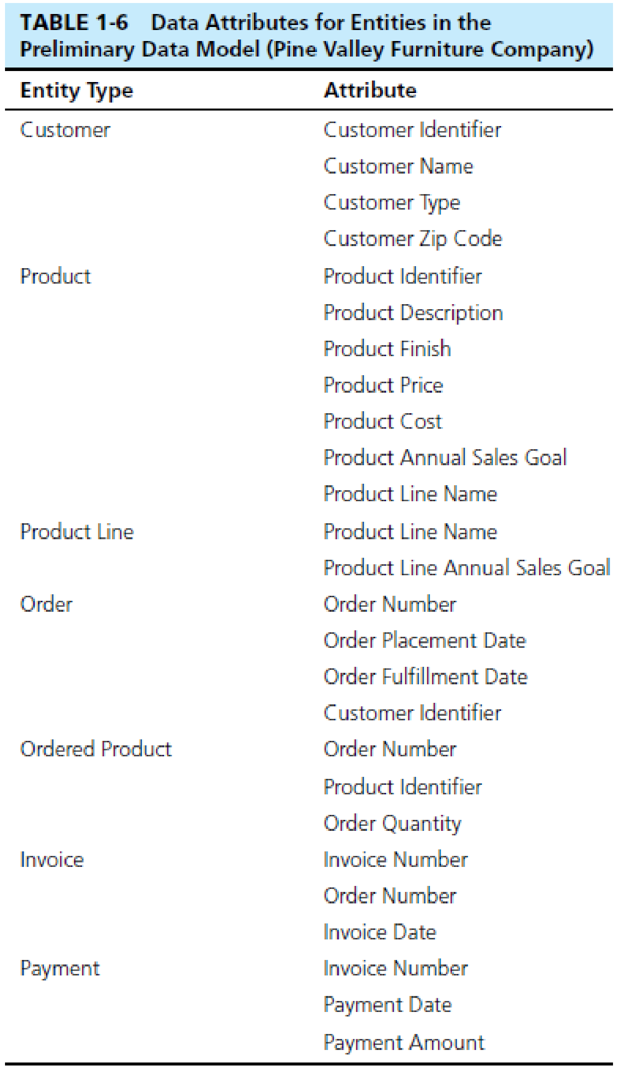


TABLE 1-6 Data Attributes for Entities in the Preliminary Data Model (Pine Valley Furniture Company)

Entity Type	Attribute
Customer	Customer Identifier
	Customer Name
	Customer Type
	Customer Zip Code
Product	Product Identifier
	Product Description
	Product Finish
	Product Price
	Product Cost
	Product Annual Sales Goal
	Product Line Name
Product Line	Product Line Name
	Product Line Annual Sales Goal
Order	Order Number
	Order Placement Date
	Order Fulfillment Date
	Customer Identifier
Ordered Product	Order Number
	Product Identifier
	Order Quantity
Invoice	Invoice Number
	Order Number
	Invoice Date
Payment	Invoice Number
	Payment Date
	Payment Amount

AFTER

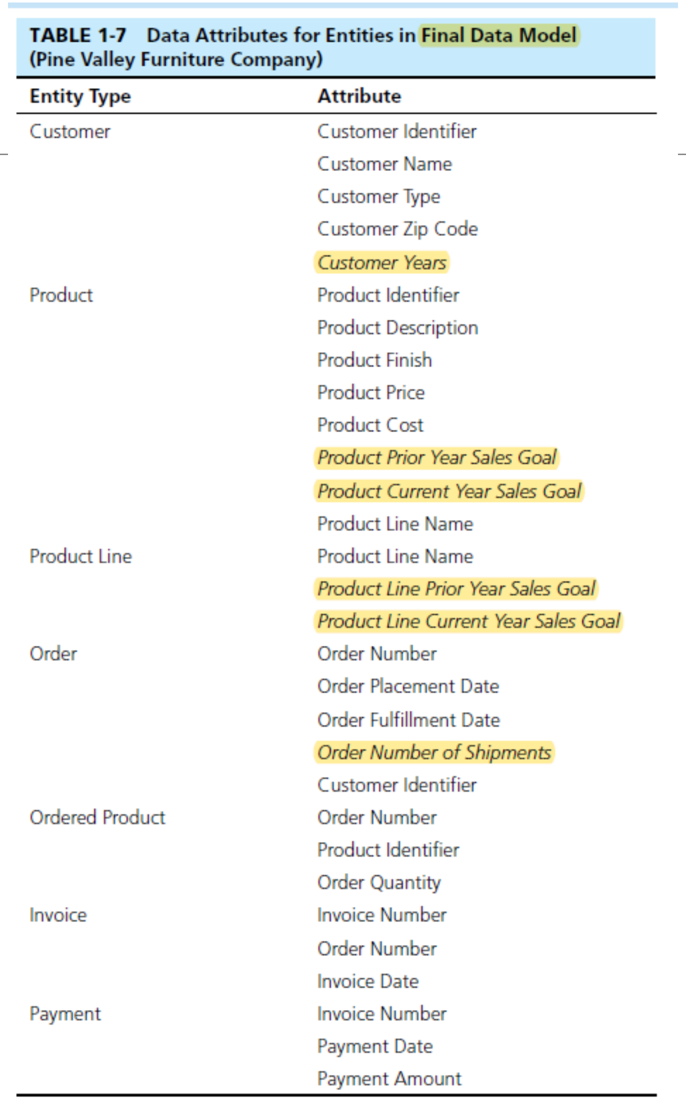
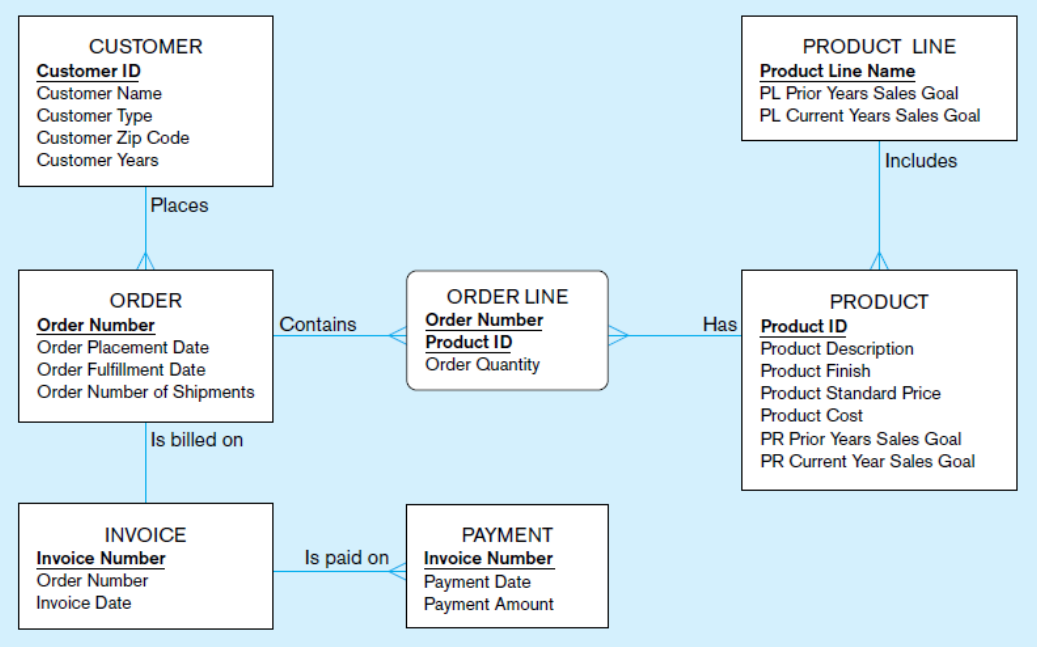


TABLE 1-7 Data Attributes for Entities in Final Data Model (Pine Valley Furniture Company)

Entity Type	Attribute
Customer	Customer Identifier
	Customer Name
	Customer Type
	Customer Zip Code
	Customer Years
Product	Product Identifier
	Product Description
	Product Finish
	Product Price
	Product Cost
	Product Prior Year Sales Goal
	Product Current Year Sales Goal
	Product Line Name
Product Line	Product Line Name
	Product Line Prior Year Sales Goal
	Product Line Current Year Sales Goal
Order	Order Number
	Order Placement Date
	Order Fulfillment Date
	Order Number of Shipments
	Customer Identifier
Ordered Product	Order Number
	Product Identifier
	Order Quantity
Invoice	Invoice Number
	Order Number
	Invoice Date
Payment	Invoice Number
	Payment Date
	Payment Amount

Case Study: Developing A Database Application for Pine Valley Furniture Company – Designing the Database

To build a **prototype**, 1) create a **project data model** and 2) translates the data model into a set of **tables (relations)**


CUSTOMER
Customer ID
Customer Name
Customer Type
Customer Zip Code
Customer Years

PRODUCT LINE
Product Line Name
PL Prior Years Sales Goal
PL Current Years Sales Goal

ORDER
Order Number
Order Placement Date
Order Fulfillment Date
Order Number of Shipments

ORDER LINE
Order Number
Product ID
Order Quantity

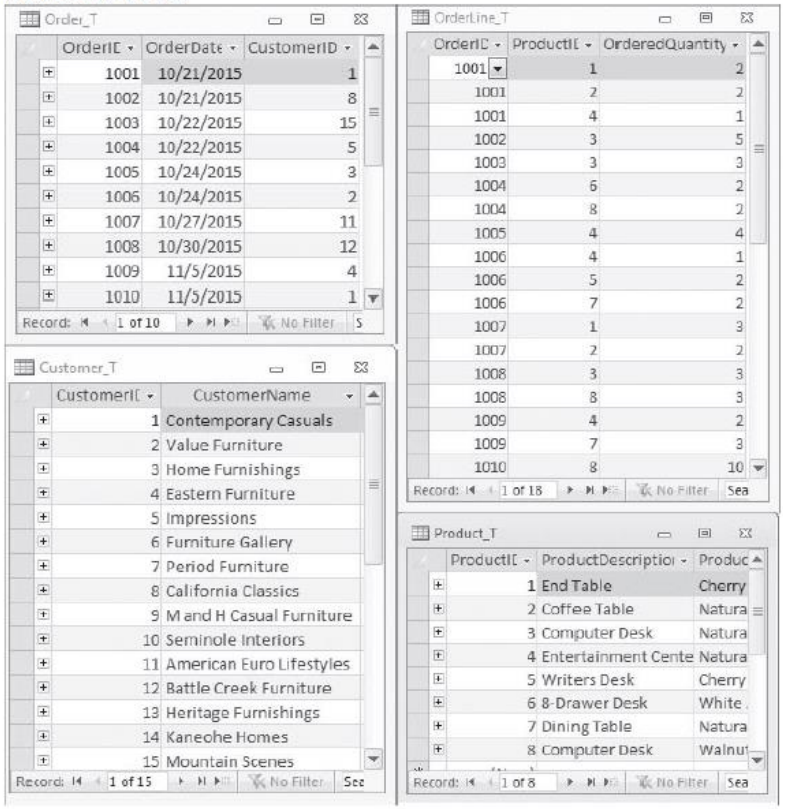
PRODUCT
Product ID
Product Description
Product Finish
Product Standard Price
Product Cost
PR Prior Years Sales Goal
PR Current Year Sales Goal

INVOICE
Invoice Number
Order Number
Invoice Date

PAYMENT
Invoice Number
Payment Date
Payment Amount

***Project* data model** for
Home Office product line marketing support system

(a) Order and Order Line Tables


Order_T

OrderID	OrderDate	CustomerID
1001	10/21/2015	1
1002	10/21/2015	8
1003	10/22/2015	15
1004	10/22/2015	5
1005	10/24/2015	3
1006	10/24/2015	2
1007	10/27/2015	11
1008	10/30/2015	12
1009	11/5/2015	4
1010	11/5/2015	1

Record: 1 of 10 No Filter

OrderLine_T

OrderID	ProductID	OrderedQuantity
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

Record: 1 of 18 No Filter Sea

Customer_T

CustomerID	CustomerName
1	Contemporary Casuals
2	Value Furniture
3	Home Furnishings
4	Eastern Furniture
5	Impressions
6	Furniture Gallery
7	Period Furniture
8	California Classics
9	M and H Casual Furniture
10	Seminole Interiors
11	American Euro Lifestyles
12	Battle Creek Furniture
13	Heritage Furnishings
14	Kaneohe Homes
15	Mountain Scenes

Record: 1 of 15 No Filter Sea

Product_T

ProductID	ProductDescriptio	Produc
1	End Table	Cherry
2	Coffee Table	Natura
3	Computer Desk	Natura
4	Entertainment Cente	Natura
5	Writers Desk	Cherry
6	8-Drawer Desk	White .
7	Dining Table	Natura
8	Computer Desk	Walnut

Record: 1 of 8 No Filter Sea

(b) Customer table

(c) Product table

Examples of four **relations**
(Pine Valley Furniture Company)

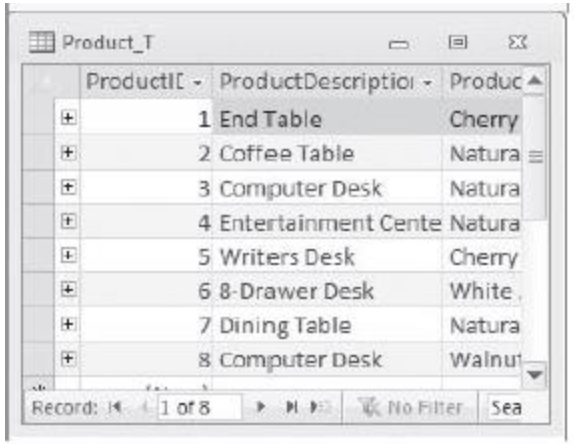
Attributes' data type

Identifier:
there can be only one value for this attributes.

```
CREATE TABLE Product_T
(ProductID                NUMBER(11,0) NOT NULL PRIMARY KEY
ProductDescription        VARCHAR (50),
ProductFinish             VARCHAR (20),
ProductStandardPrice      DECIMAL(6,2),
ProductCost               DECIMAL,
ProductPriorYearGoal      DECIMAL,
ProductCurrentYearGoal    DECIMAL,
ProductLineID             VARCHAR (40),
FOREIGN KEY               (ProductLineID)  REFERENCES ProductLine_T (ProductLineID));
```

SQL definition of Product table

Referential integrity constraint:
a foreign key must have a matching primary key or it must be null.



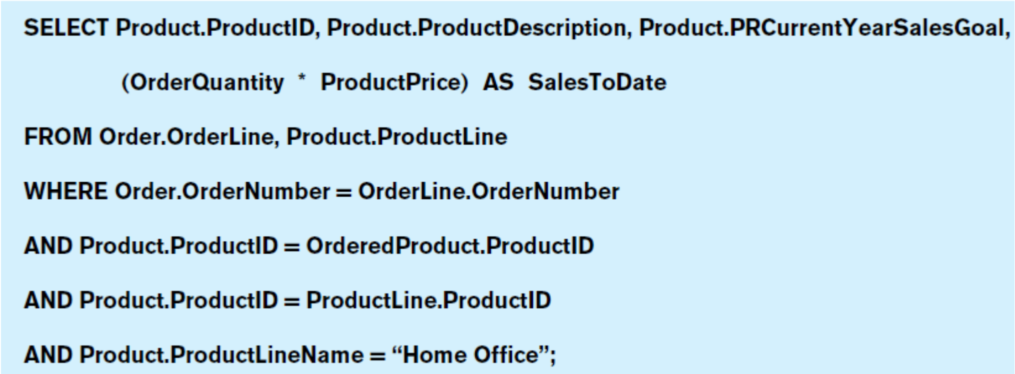
ProductI[	ProductDescriptio	Produc
1	End Table	Cherry
2	Coffee Table	Natura
3	Computer Desk	Natura
4	Entertainment Cente	Natura
5	Writers Desk	Cherry
6	8-Drawer Desk	White
7	Dining Table	Natura
8	Computer Desk	Walnut

(c) Product table

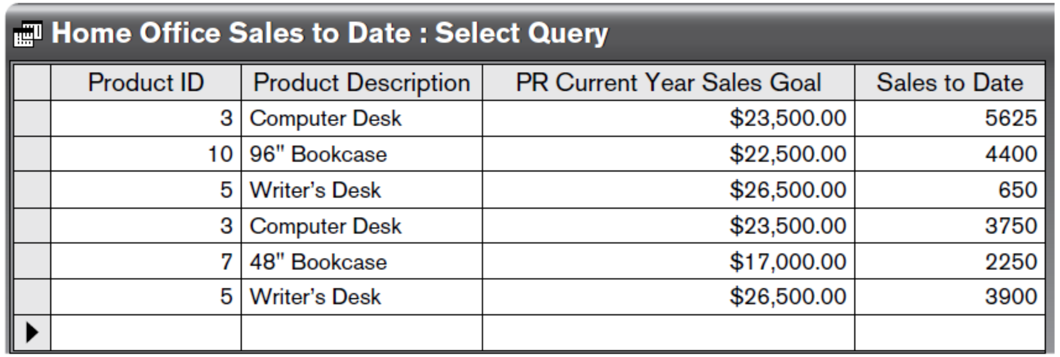
Create indexes to speed up the search

- For example, Helen wants to use the database is to *look at sales by **product finish***.
  - Thus, create an index on the Product_T table using the ***Product Finish** attribute*.

Case Study: Developing A Database Application for Pine Valley Furniture Company – Using the Database



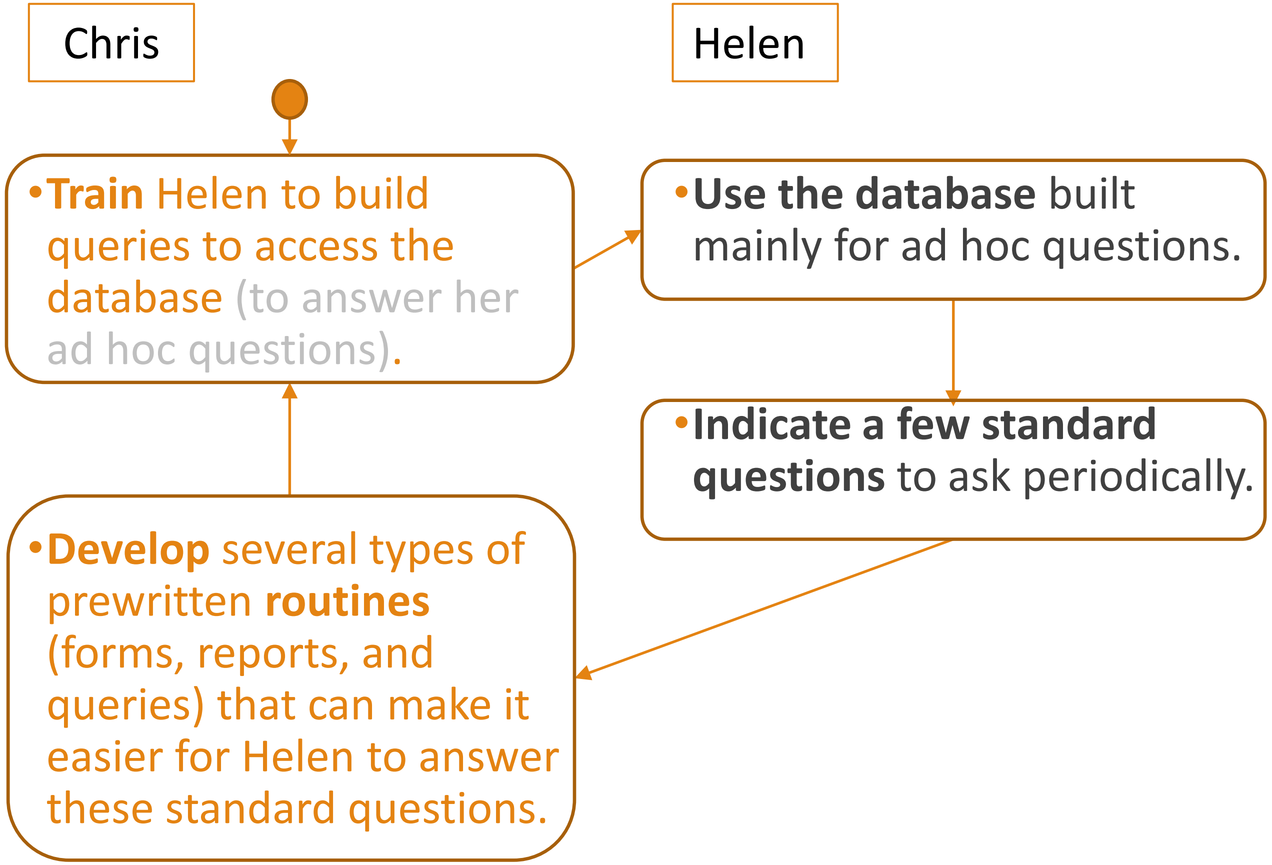
```
SELECT Product.ProductID, Product.ProductDescription, Product.PRCurrentYearSalesGoal,
        (OrderQuantity * ProductPrice) AS SalesToDate
FROM Order.OrderLine, Product.ProductLine
WHERE Order.OrderNumber = OrderLine.OrderNumber
AND Product.ProductID = OrderedProduct.ProductID
AND Product.ProductID = ProductLine.ProductID
AND Product.ProductLineName = “Home Office”;
```



Home Office Sales to Date : Select Query

Product ID	Product Description	PR Current Year Sales Goal	Sales to Date
3	Computer Desk	$23,500.00	5625
10	96" Bookcase	$22,500.00	4400
5	Writer’s Desk	$26,500.00	650
3	Computer Desk	$23,500.00	3750
7	48" Bookcase	$17,000.00	2250
5	Writer’s Desk	$26,500.00	3900

Home Office product line sales comparison

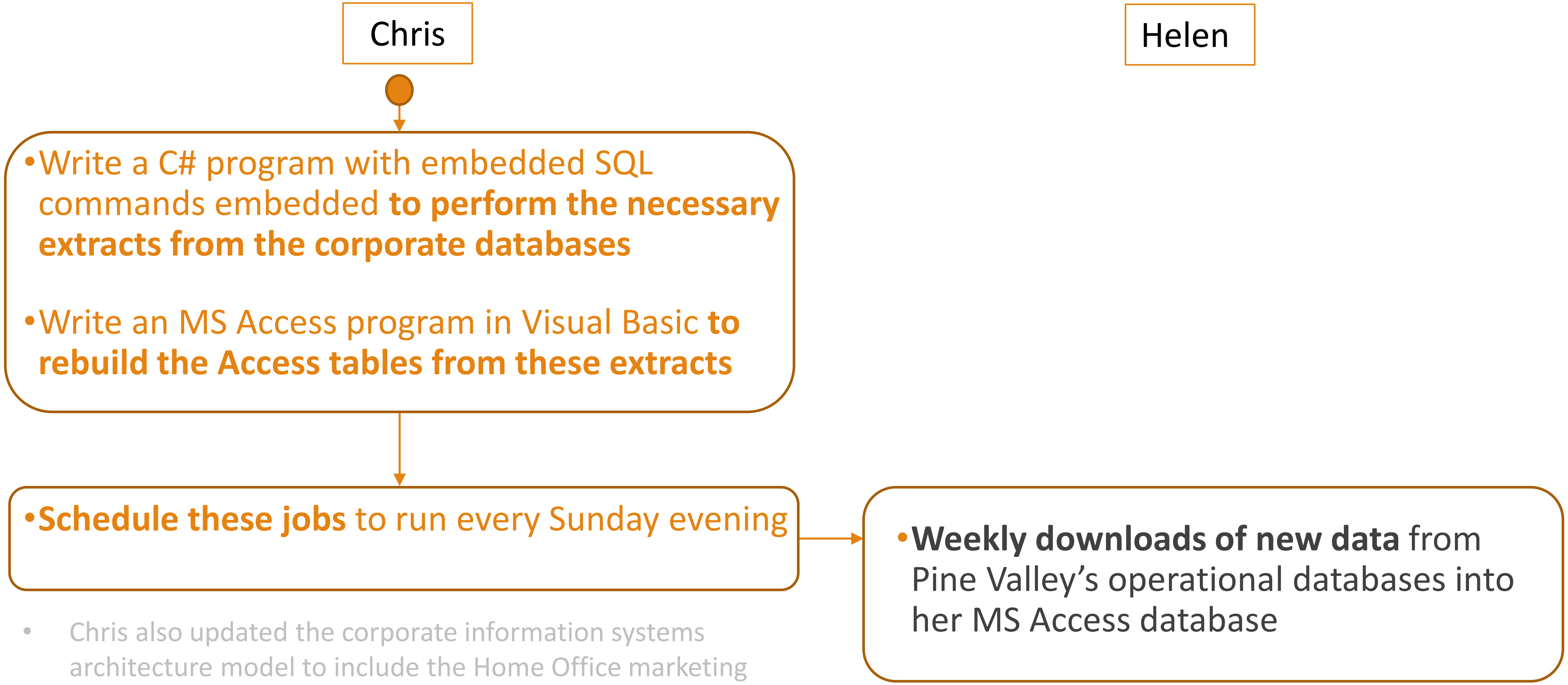


Chris

Helen

- **Train** Helen to build queries to access the database (to answer her ad hoc questions).
- **Use the database** built mainly for ad hoc questions.
- **Indicate a few standard questions** to ask periodically.
- **Develop** several types of prewritten **routines** (forms, reports, and queries) that can make it easier for Helen to answer these standard questions.

Case Study: Developing A Database Application for Pine Valley Furniture Company – Administering the Database



Chris

Helen

- Write a C# program with embedded SQL commands embedded **to perform the necessary extracts from the corporate databases**
- Write an MS Access program in Visual Basic **to rebuild the Access tables from these extracts**

- **Schedule these jobs** to run every Sunday evening

- **Weekly downloads of new data** from Pine Valley's operational databases into her MS Access database

- Chris also updated the corporate information systems architecture model to include the Home Office marketing support system.

References

- ❑ Jeffrey A. Hoffer, V. Ramesh and Helkki Topi, Modern Database Management, 12th Edition, 2016

- ❑ Jeffrey A. Hoffer, Mary B. Prescott and Fred R. McFadden, Modern Database Management, 8th Edition, 2006

- ❑ Figures used in this lecture: Copyright © 2009 Pearson Education, Inc. Publishing as Prentice Hall