



Database and Information Management

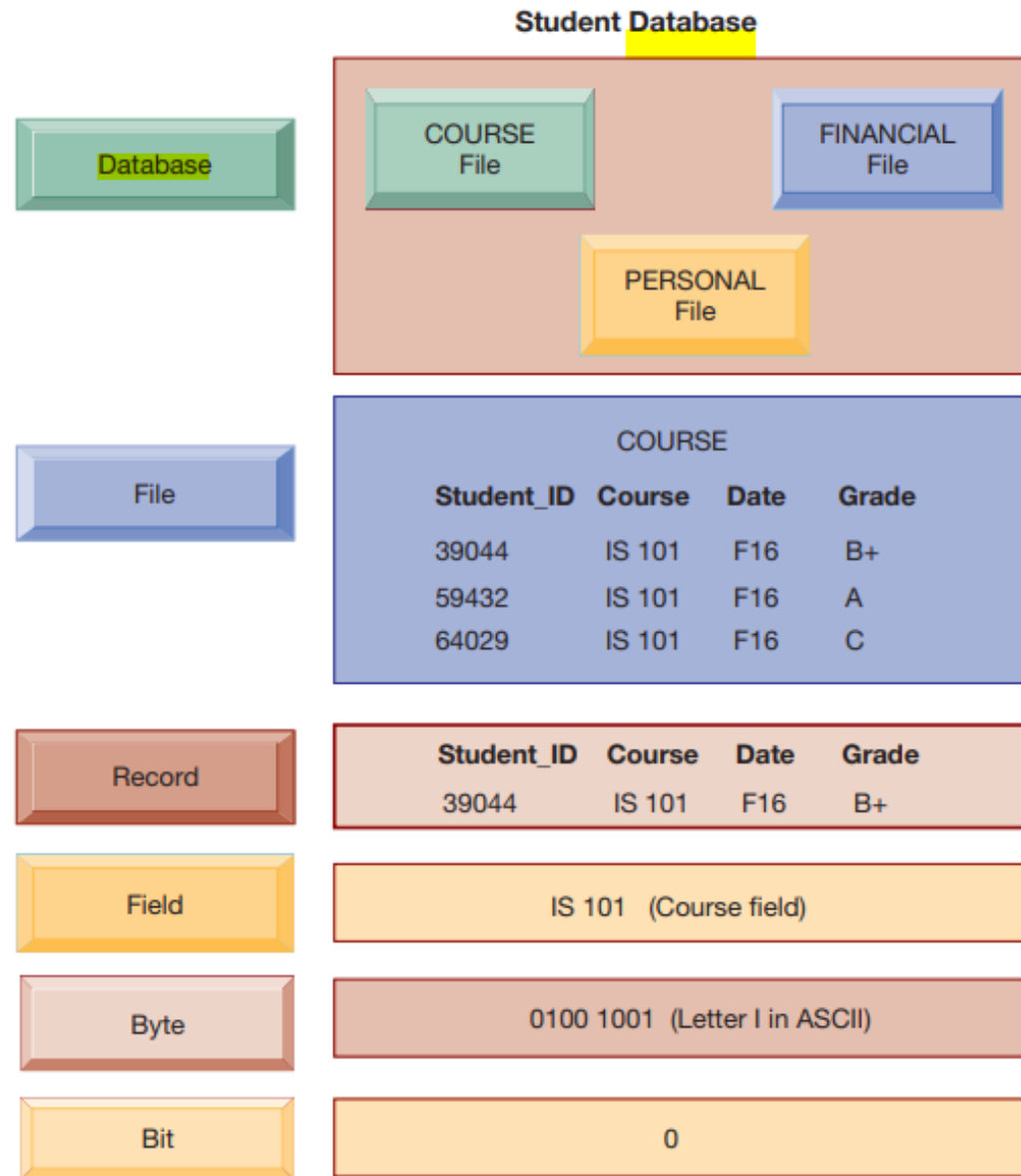
An effective information system

- An effective information system provides users with accurate, timely, and relevant information.
- **Accurate** information is free of errors.
- Information is **timely** when it is available to decision makers when it is needed.
- Information **is relevant** when it is useful and appropriate for the types of work and decisions that require it.

File Organization

- **A bit** represents the smallest unit of data a computer can handle.
- A group of bits, called a **byte**, represents a single character, which can be a letter, a number, or another symbol.
- A grouping of characters into a word, a group of words, or a complete number (such as a person's name or age) is called a **field**.
- A group of related fields comprises a **record**;
- a group of records of the same type is called a **file** .
- A group of related files makes up a **database**.

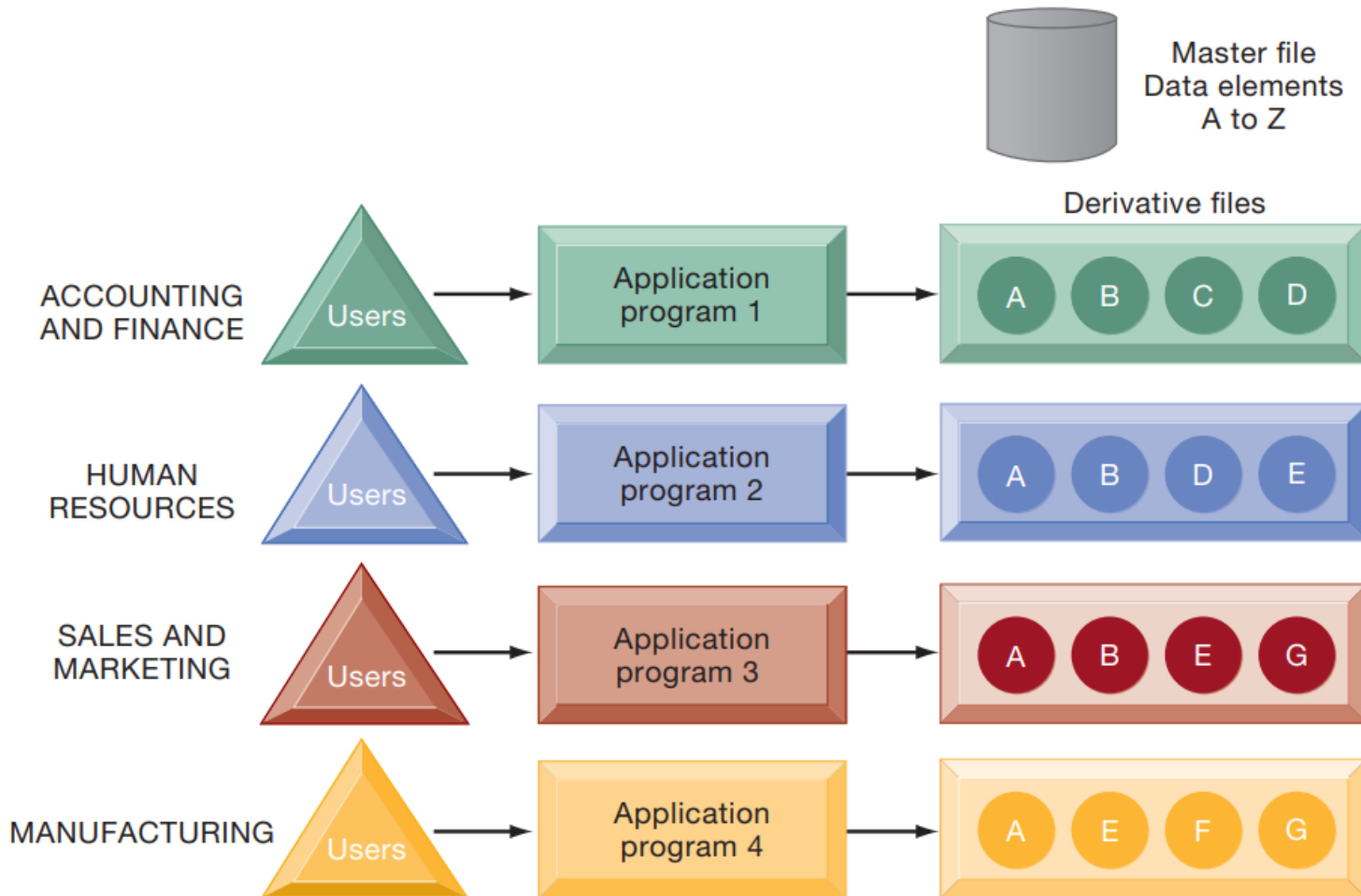
FIGURE 6.1 THE DATA HIERARCHY



Database Concepts

- An **entity** is a person, place, thing, or event on which we store and maintain information.
- Each characteristic or quality describing a particular entity is called an **attribute**.

COURSE				Entity
Student_ID	Course	Date	Grade	attributes
39044	IS 101	F16	B+	
59432	IS 101	F16	A	
64029	IS 101	F16	C	

FIGURE 6.2 TRADITIONAL FILE PROCESSING

The use of a traditional approach to file processing encourages each functional area in a corporation

Problems with the Traditional File

1. data redundancy and inconsistency
2. program-data dependence
3. inflexibility
4. poor data security
5. an inability to share data among applications.

1. Data Redundancy and Inconsistency Data

- Data redundancy presence of duplicate data in multiple data files so that the same data are stored in more than one place or location.
- Data redundancy wastes storage resources and also leads to data inconsistency, where the same attribute may have different values.

example

Emp_code	Emp_name	Emp_experience	Dept_id	Dep_manager
emp001	K.Saman	2	Dep22	Ranil
emp002	R.Sunil	5	Dep18	Sunil
emp003	N. Shafeen	3	Dep17	Anil
emp004	P.Kumara	1	Dep_22	Ranil
emp005	R.Perera	4	Dep18	Sunil
emp006	S.Gamage	2	Dep19	Malki

same data are stored in more than one place
or location

same attribute may have different values

Inventory

Item_ID	Item	Stock
39044	IS101	Extra Large
59435	IS102	Small

FactoryStock

ItemID	Name	Size
39044	Shirt	XL
59435	Short	S

the sales, inventory, and manufacturing systems of a clothing retailer might use different codes to represent clothing size. One system might represent clothing size as “extra large,” whereas another might use the code “XL” for the same purpose.

2. Program-Data Dependence

- Program-data dependence refers to the coupling of data stored in files and the specific programs required to update and maintain those files such that changes in programs require changes to the data.

3. Lack of Flexibility

- A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad hoc reports or respond to unanticipated information requirements in a timely fashion.

4. Poor Security

- Because there is little control or management of data, access to and dissemination of information may be out of control.
- Management may have no way of knowing who is accessing or even making changes to the organization's data.

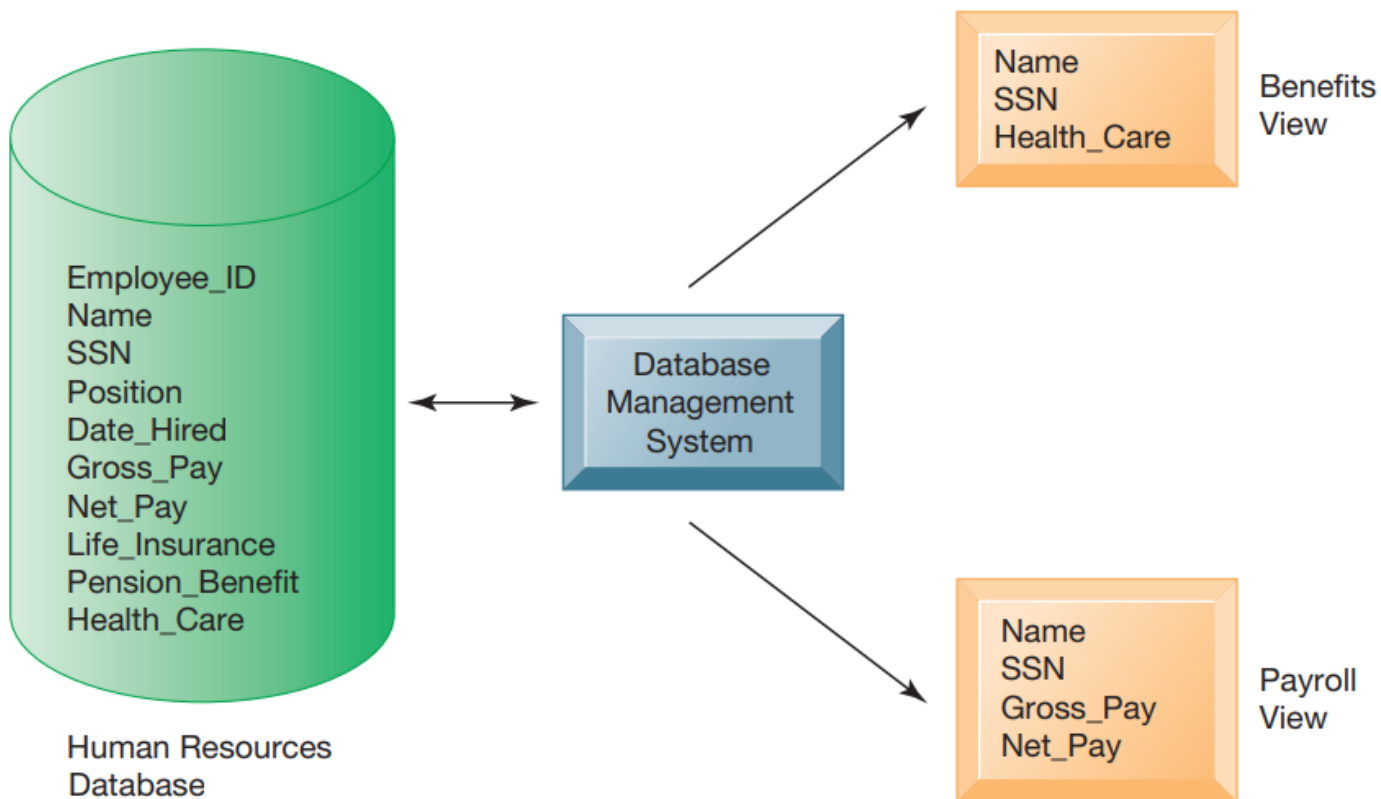
5. Lack of Data Sharing and Availability

- Because pieces of information in different files and different parts of the organization cannot be related to one another, it is virtually impossible for information to be shared or accessed in a timely manner.
- Information cannot flow freely across different functional areas or different parts of the organization.

A database management system (DBMS)

- A database management system (DBMS) is a software designed to define, manipulate, retrieve and manage data in a centralized database
 - The DBMS acts as an interface between application programs and the physical data files.
 - When the application program calls for a data item, such as gross pay, the DBMS finds this item in the database and presents it to the application program.

FIGURE 6.3 HUMAN RESOURCES DATABASE WITH MULTIPLE VIEWS



A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

How a DBMS Solves the Problems of the Traditional File Environment

- A DBMS reduces data redundancy and inconsistency by minimizing isolated files in which the same data are repeated.
- Even if the organization maintains some redundant data, using a DBMS eliminates data inconsistency because the DBMS can help the organization ensure that every occurrence of redundant data has the same values.

Employee dept

example

Emp_code	Emp_name	Emp_experience	Dept_id	Dep_manager
emp001	K.Saman	2	Dep22	Ranil
emp002	R.Sunil	5	Dep18	Sunil
emp003	N. Shafeen	3	Dep17	Anil
emp004	P.Kumara	1	Dep_22	Ranil
emp005	R.Perera	4	Dep18	Sunil
emp006	S.Gamage	2	Dep19	Malki

Employee

Dept-_Mgt

Department

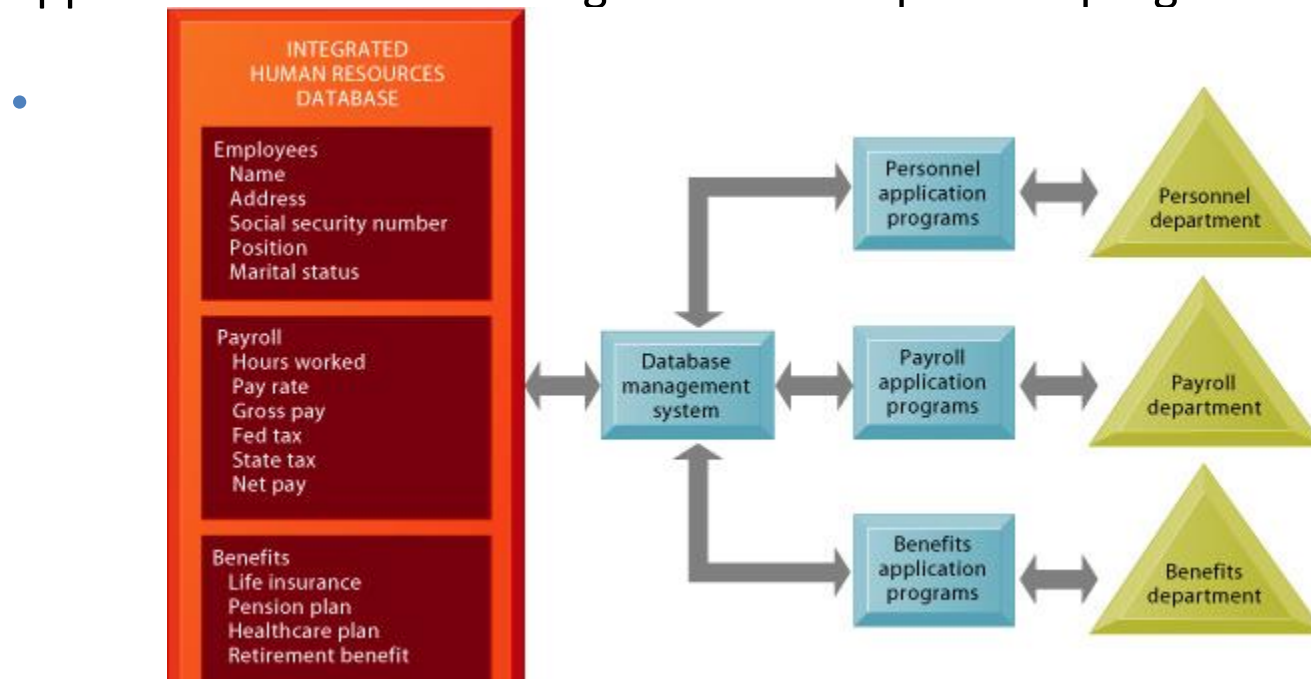
Emp_code	Emp_name	Emp_experience
emp001	K.Saman	2
emp002	R.Sunil	5
emp003	N. Shafeen	3
emp004	P.Kumara	1
emp005	R.Perera	4
emp006	S.Gamage	2

Dept_id	Department
Dep22	Math's
Dep18	Science
Dep17	Economy
Dep19	English

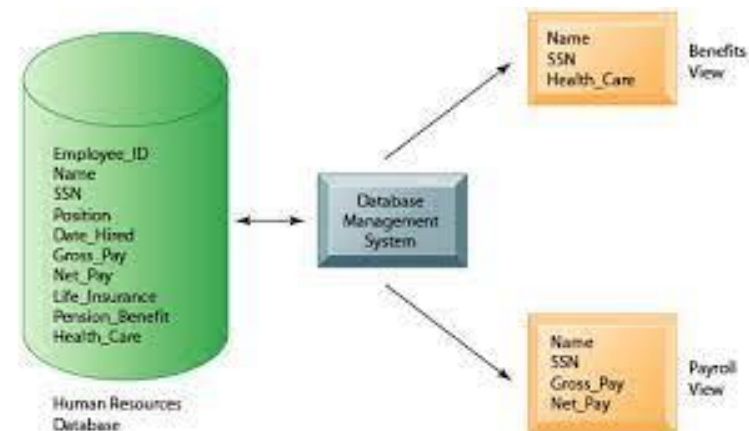
Dept_id	Dep_manager
Dep22	Ranil
Dep18	Sunil
Dep17	Anil
Dep19	Malki

- The DBMS uncouples programs and data, enabling data to stand on their own.

Access and availability of information will be increased and program development and maintenance costs reduced because users and programmers can perform ad hoc queries of the database for many simple applications without having to write complicated programs.



- The DBMS enables the organization to centrally manage data, their use, and security.
- Data sharing throughout the organization is easier because the data are presented to users as being in a single location rather than fragmented in many different systems and files.



Relational DBMS

- Relational databases represent data as two-dimensional tables (called relations).
- Tables may be referred to as files.
- Each table contains data on an entity and its attributes.
- Eg:
 - Microsoft Access- for desktop
 - DB2, Oracle Database, and Microsoft SQL Server are relational DBMS for large mainframes and midrange computers.
 - MySQL is a popular open source DBMS.

FIGURE 6.4 RELATIONAL DATABASE TABLES**SUPPLIER**

Columns (Attributes, Fields)

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Rows
(Records,
Tuples)Key Field
(Primary Key)**PART**

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

SUPPLIER(supplier_number, supplier_name,
supplier_street, supplier_city, state, ZIP_code)

- Each table consists of a grid of columns and rows of data.
- Each individual element of data for each entity is stored as a separate field, and each field represents an **attribute** for that entity.

Related terms to Relational database

Tuples

- The actual information about a single supplier that resides in a table is called a row.
- Rows are commonly referred to as records, or in very technical terms, as **tuples**.

key field

- The field for Supplier_Number in the SUPPLIER table uniquely identifies each record so that the record can be retrieved, updated, or sorted. It is called a **key field**.

Primary key

- Each table in a relational database has one field that is designated as its primary key. This key field is the unique identifier for all the information in any row of the table and this **primary key** cannot be duplicated.

foreign key

- An attribute in a relation of a database that serves as the primary key of another relation in the same database

Domain

- **Domain** refers to all the values which a data element may contain.

Capabilities of Database Management Systems

- A DBMS includes capabilities and tools for organizing, managing, and accessing the data in the database.
- The most important are its data definition language, data dictionary, and data manipulation language.

data definition capability

- DBMS have a data definition capability to specify the structure of the content of the database.
- It would be used to create database tables and to define the characteristics of the fields in each table.
- This information about the database would be documented in a data dictionary.
- A **data dictionary** is an automated or manual file that stores definitions of data elements and their characteristics

Sample data dictionary report

- The sample data dictionary report for a human resources database provides helpful information, such as the size of the data element, which programs and reports use it, and which group in the organization is the owner responsible for maintaining it. The report also shows some of the other names that the organization uses for this piece of data.

NAME: AMT-PAY-BASE
FOCUS NAME: BASEPAY
PC NAME: SALARY

DESCRIPTION: EMPLOYEE'S ANNUAL SALARY

SIZE: 9 BYTES

TYPE: N (NUMERIC)

DATE CHANGED: 01/01/04

OWNERSHIP: COMPENSATION

UPDATE SECURITY: SITE PERSONNEL

ACCESS SECURITY: MANAGER, COMPENSATION PLANNING AND RESEARCH

MANAGER, JOB EVALUATION SYSTEMS

MANAGER, HUMAN RESOURCES PLANNING

MANAGER, SITE EQUAL OPPORTUNITY AFFAIRS

MANAGER, SITE BENEFITS

MANAGER, CLAIMS PAYING SYSTEMS

MANAGER, QUALIFIED PLANS

MANAGER, SITE EMPLOYMENT/EEO

BUSINESS FUNCTIONS USED BY: COMPENSATION

HR PLANNING

EMPLOYMENT

INSURANCE

PENSION

401K

PROGRAMS USING: PI01000

PI02000

PI03000

PI04000

PI05000

REPORTS USING: REPORT 124 (SALARY INCREASE TRACKING REPORT)

REPORT 448 (GROUP INSURANCE AUDIT REPORT)

REPORT 452 (SALARY REVIEW LISTING)

PENSION REFERENCE LISTING

ACCESS DATA DICTIONARY FEATURES

hospital : Database- C:\Users\USER\Documen... Table Tools Sign in

File Home Create External Data Database Tools Help Design Tell me what you want to do

View Primary Key Builder Test Validation Rules Tools Insert Rows Delete Rows Modify Lookups Property Indexes Sheet Show/Hide Create Data Macros Rename/ Delete Macro Field, Record & Table Events Relationships Object Dependencies Relationships

UPDATES AVAILABLE Updates for Office are ready to be installed, but first we need to close some apps. Update now

All Access Objects

Search...

Tables

- course
- Department
- Paste Errors
- Treatment

Forms

- course
- Treatment

Reports

- Report1
- Treatment
- Treatment1

course

Field Name	Data Type	Description (Optional)
Course_name	Short Text	
Course_number	Short Text	
Credit_hours	Number	
Dept_ID	Short Text	

Field Properties

General Lookup

Field Size	Integer
Format	
Decimal Places	Auto
Input Mask	
Caption	
Default Value	0
Validation Rule	
Validation Text	
Required	No
Indexed	No
Text Align	General

A field name can be up to 64 characters long, including spaces. Press F1 for help on field names.

Design view. F6 = Switch panes. F1 = Help.

Num Lock

Data Manipulation

- DBMS includes tools for accessing and manipulating information in databases.
- Most DBMS have a specialized language called a **data manipulation language** that is used to add, change, delete, and retrieve the data in the database.
- The most prominent data manipulation language today is **Structured Query Language**, or SQL.

PRACTICAL WITH ACCESS

Normalization

- Normalization is the process of organizing data in a database.
- This includes creating tables and establishing relationships between those tables according to rules designed both to protect the data and to make the database more flexible by eliminating redundancy and inconsistent dependency.

- The conceptual database design describes how the data elements in the database are to be grouped.
- The design process identifies relationships among data elements and the most efficient way of grouping data elements together to meet business information requirements.
- Groups of data are organized, refined, and streamlined until an overall logical view of the relationships among all the data in the database emerges.
- To use a relational database model effectively, complex groupings of data must be streamlined to minimize redundant data elements and difficult many-to -many relationships.
- The process of creating small, stable, yet flexible and adaptive data structures from complex groups of data is called normalization .

FIGURE 6.9 AN UNNORMALIZED RELATION FOR ORDER

ORDER (Before Normalization)

Order_ Number	Order_ Date	Part_ Number	Part_ Name	Unit_ Price	Part_ Quantity	Supplier_ Number	Supplier_ Name	Supplier_ Street	Supplier_ City	Supplier_ State	Supplier_ Zip
------------------	----------------	-----------------	---------------	----------------	-------------------	---------------------	-------------------	---------------------	-------------------	--------------------	------------------

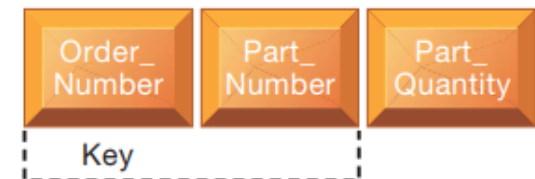
An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

FIGURE 6.10 NORMALIZED TABLES CREATED FROM ORDER

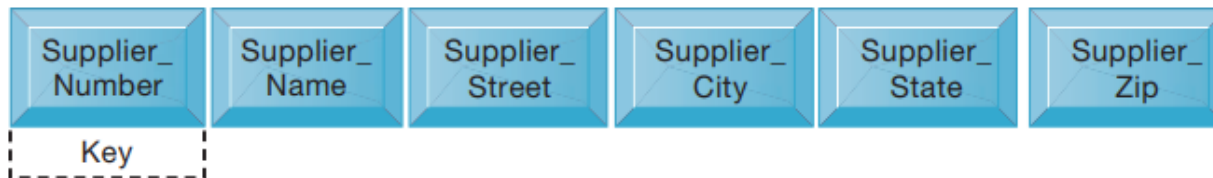
PART



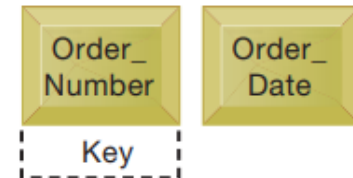
LINE ITEM



SUPPLIER



ORDER

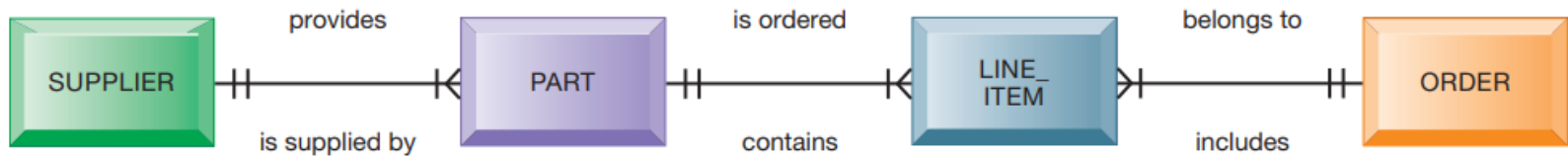


After normalization, the original relation ORDER has been broken down into four smaller relations. The relation ORDER is left with only two attributes, and the relation LINE_ITEM has a combined, or concatenated, key consisting of Order_Number and Part_Number.

Entity relationship Diagram

- This diagram illustrates the relationship between the entities SUPPLIER, PART, LINE_ITEM, and ORDER.
- The boxes represent entities.
- The lines connecting the boxes represent relationships.
- A line connecting two entities that ends in two short marks designates a one-to-one relationship.
- A line connecting two entities that ends with a crow's foot topped by a short mark indicates a one-to-many relationship.

FIGURE 6.11 AN ENTITY-RELATIONSHIP DIAGRAM



This diagram shows the relationships between the entities SUPPLIER, PART, LINE_ITEM, and ORDER that might be used to model the database in Figure 6.10.

- one ORDER can contain many LINE_ITEMS

- If the business doesn't get its **data model right**, the system won't be able to serve the business well.
- The **company's systems** will not be as effective as they could be because they'll have to work with data that may be inaccurate, incomplete, or difficult to retrieve.

example

- Famous Footwear, a shoe store chain with more than 800 locations in 49 states, could not achieve its goal of having “the right style of shoe in the right store for sale at the right price” because its database was not properly designed for rapidly adjusting store inventory.
- The company had an Oracle relational database running on a midrange computer, but the database was designed primarily for producing standard reports for management rather than for reacting to marketplace changes.
- Management could not obtain precise data on specific items in inventory in each of its stores.
- The company had to work around this problem by building a new database where the sales and inventory data could be better organized for analysis and inventory management.