Informática y Comunicaciones

Chapter 6
Introduction to Databases
Chapter 6: outline

6.1 Getting started
6.2 Entity and Relations
Purpose of a Database

- The purpose of a database is to keep track of things.
- Unlike a list or spreadsheet, a database may store information that is more complicated than a simple list.
Problems with Lists: Redundancy

- In a list, each row is intended to stand on its own. As a result, the same information may be entered several times.
  - For Example: A list of Projects may include the Project Manager’s Name, ID, and Phone Extension. If a particular person is managing 10 projects, his/her information would have to be entered 10 times.
Problems with Lists: Multiple Themes

- In a list, each row may contain information on more than one theme. As a result, needed information may appear in the lists only if information on other themes is also present.
  - For Example: A list of Projects may include Project Manager information (Name, ID, and Phone Extension) and Project information (Name, ID, StartDate, Budget) in the same row.
List Modification Issues

- Redundancy and multiple themes create modification problems:
  - Deletion problems
  - Update problems
  - Insertion problems
List Modification Issues (Cont’d)

If Adviser Baker is changed to Taing, we need to change AdviserEmail as well. If changed to Valdez, we need to change AdviserEmail, Department, and AdminLastName.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LastName</td>
<td>FirstName</td>
<td>Email</td>
<td>AdviserLastName</td>
<td>AdviserEmail</td>
<td>Department</td>
</tr>
<tr>
<td>2</td>
<td>Andrews</td>
<td>Matthew</td>
<td><a href="mailto:Matthew.Andrews@ourcampus.edu">Matthew.Andrews@ourcampus.edu</a></td>
<td>Baker</td>
<td><a href="mailto:Linda.Baker@ourcampus.edu">Linda.Baker@ourcampus.edu</a></td>
<td>Accounting</td>
</tr>
<tr>
<td>3</td>
<td>Brisbon</td>
<td>Lisa</td>
<td><a href="mailto:Lisa.Brisbon@ourcampus.edu">Lisa.Brisbon@ourcampus.edu</a></td>
<td>Valdez</td>
<td><a href="mailto:Richard.Valdez@ourcampus.edu">Richard.Valdez@ourcampus.edu</a></td>
<td>Chemistry</td>
</tr>
<tr>
<td>4</td>
<td>Fischer</td>
<td>Douglas</td>
<td><a href="mailto:Douglas.Fischer@ourcampus.edu">Douglas.Fischer@ourcampus.edu</a></td>
<td>Baker</td>
<td><a href="mailto:Linda.Baker@ourcampus.edu">Linda.Baker@ourcampus.edu</a></td>
<td>Accounting</td>
</tr>
<tr>
<td>5</td>
<td>Hwang</td>
<td>Terry</td>
<td><a href="mailto:Terry.Hwang@ourcampus.edu">Terry.Hwang@ourcampus.edu</a></td>
<td>Taing</td>
<td><a href="mailto:Susan.Taing@ourcampus.edu">Susan.Taing@ourcampus.edu</a></td>
<td>Accounting</td>
</tr>
<tr>
<td>6</td>
<td>Lai</td>
<td>Tzu</td>
<td><a href="mailto:Tzu.Lai@ourcampus.edu">Tzu.Lai@ourcampus.edu</a></td>
<td>Valdez</td>
<td><a href="mailto:Richard.Valdez@ourcampus.edu">Richard.Valdez@ourcampus.edu</a></td>
<td>Chemistry</td>
</tr>
<tr>
<td>7</td>
<td>Marino</td>
<td>Chip</td>
<td><a href="mailto:Chip.Manno@ourcampus.edu">Chip.Manno@ourcampus.edu</a></td>
<td>Tran</td>
<td><a href="mailto:Ken.Tran@ourcampus.edu">Ken.Tran@ourcampus.edu</a></td>
<td>InfoSystems</td>
</tr>
<tr>
<td>8</td>
<td>Thompson</td>
<td>James</td>
<td><a href="mailto:James.Thompson@ourcampus.edu">James.Thompson@ourcampus.edu</a></td>
<td>Taing</td>
<td><a href="mailto:Susan.Taing@ourcampus.edu">Susan.Taing@ourcampus.edu</a></td>
<td>Accounting</td>
</tr>
<tr>
<td>9</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>Biology</td>
</tr>
</tbody>
</table>

Deleted row—Student, Adviser, and Department data lost

Inserted row—both Student and Adviser data missing
Relational databases are designed to address many of the information complexity issues.
A relational database stores information in tables. Each informational topic is stored in its own table.

In essence, a relational database will break-up a list into several parts—one part for each theme in the list.

A Project List would be divided into a CUSTOMER Table, a PROJECT Table, and a PROJECT_MANAGER Table.
In our relational database, we broke our list into several tables. Somehow the tables must be *joined* back together.

In a relational database, tables are joined together using the value of the data.

If a PROJECT has a CUSTOMER, the Customer_ID is stored as a column in the PROJECT table. The value stored in this column can be used to retrieve specific customer information from the CUSTOMER table.
A relational database is more complicated than a list.

However, a relational database minimizes data redundancy, preserves complex relationships among topics, and allows for partial data.

Furthermore, a relational database provides a solid foundation for user forms and reports.
Relational Database Example

STUDENT data linked to ADVISER data via AdviserLastName

Introduction to Databases 6-12
A Relational Database Solves the Problems of Lists

- Changed data—data remains consistent
- Inserted data—no STUDENT data required
- Deleted data—no ADVISER data lost
The Department, Advisor and Student Tables

Can insert DEPARTMENT data as needed—no ADVISER or STUDENT data required.

Can change STUDENT Adviser name as needed—new value is linked to its own data.

Can delete STUDENT data as needed—no DEPARTMENT or ADVISER data lost.
The Art Course Database Tables

Can change COURSE CourseDate without problems

Can insert new COURSE data as needed

Can delete ENROLLMENT rows as needed—no adverse consequences
The Project Equipment Tables
Structured Query Language (SQL) is an international standard for creating, processing and querying databases and their tables.

Many database applications use SQL to retrieve, format, report, insert, delete, and/or modify data for users.
SQL Example

• We can use SQL to combine the data in the three tables in the Art Course Database to recreate the original list structure of the data.
• We do this by using an SQL SELECT statement.
SELECT CUSTOMER.CustomerLastName,  
          CUSTOMER.CustomerFirstName,  
          CUSTOMER.Phone,  
          COURSE.CourseDate,  
          ENROLLMENT.AmountPaid,  
          COURSE.Course, COURSE.Fee  
FROM CUSTOMER, ENROLLMENT, COURSE  
WHERE CUSTOMER.CustomerNumber  
    = ENROLLMENT.CustomerNumber  
AND COURSE.CourseNumber  
    = ENROLLMENT.CourseNumber;
### SQL Example Results

**Art Course List**

<table>
<thead>
<tr>
<th>CustomerLastName</th>
<th>CustomerFirstName</th>
<th>Phone</th>
<th>CourseDate</th>
<th>AmountPaid</th>
<th>Course</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson</td>
<td>Ariel</td>
<td>206-567-1234</td>
<td>10/1/2013</td>
<td>$250.00</td>
<td>Adv Pastels</td>
<td>$500.00</td>
</tr>
<tr>
<td>Johnson</td>
<td>Ariel</td>
<td>206-567-1234</td>
<td>3/15/2013</td>
<td>$350.00</td>
<td>Int Pastels</td>
<td>$350.00</td>
</tr>
<tr>
<td>Green</td>
<td>Robin</td>
<td>425-678-8765</td>
<td>9/15/2013</td>
<td>$350.00</td>
<td>Beg Oils</td>
<td>$350.00</td>
</tr>
<tr>
<td>Jackson</td>
<td>Charles</td>
<td>360-789-3456</td>
<td>10/1/2013</td>
<td>$500.00</td>
<td>Adv Pastels</td>
<td>$500.00</td>
</tr>
<tr>
<td>Pearson</td>
<td>Jeffery</td>
<td>206-567-2345</td>
<td>10/1/2013</td>
<td>$500.00</td>
<td>Adv Pastels</td>
<td>$500.00</td>
</tr>
<tr>
<td>Sears</td>
<td>Miguel</td>
<td>360-789-4567</td>
<td>9/15/2013</td>
<td>$350.00</td>
<td>Beg Oils</td>
<td>$350.00</td>
</tr>
<tr>
<td>Kyle</td>
<td>Leah</td>
<td>425-678-7654</td>
<td>11/15/2013</td>
<td>$250.00</td>
<td>Adv Pastels</td>
<td>$500.00</td>
</tr>
<tr>
<td>Myers</td>
<td>Lynda</td>
<td>360-789-5678</td>
<td>10/15/2013</td>
<td>$0.00</td>
<td>Beg Oils</td>
<td>$350.00</td>
</tr>
</tbody>
</table>

Introduction to Databases 6-20
The four components of a database system are:

- Users
- Database Application
- Database Management System (DBMS)
- Database
Components of a Database System

- Users
- Database Application
- Database management system (DBMS)
- Database
A **user** of a database system will

- use a database application to track things;
- use forms to enter, read, delete and query data; and
- produce reports.
The Database

- A database is a *self-describing* collection of *related* records.
  - The database itself contains the definition of its structure.
  - Metadata is data describing the structure of the database data.

- Tables within a relational database are related to each other.
Example Database Metadata: A Relationship Diagram
Database Contents

- User data
- Metadata
- Indexes and other overhead data
- Application metadata
Database Management System (DBMS)

- A **database management system (DBMS)** serves as an intermediary between database applications and the database.
- The DBMS manages and controls database activities.
- The DBMS creates, processes and administers the databases it controls.
Functions of a DBMS

- Create databases
- Create tables
- Create supporting structures
- Read database data
- Modify database data (insert, update, and delete)
- Maintain database structures
- Enforce rules
- Control concurrency
- Provide security
- Perform backup and recovery
Referential Integrity Constraints

- The DBMS will enforce many constraints.
- **Referential integrity constraints** ensure that the values of a column in one table are valid based on the values in another table.
  - If a 5 was entered as a CustomerID in the PROJECT table, a Customer having a CustomerID value of 5 must exist in the CUSTOMER table.
A **database application** is a set of one or more computer programs that serves as an intermediary between the user and the DBMS.
Functions of Database Applications

- Create and process forms
- Process user queries
- Create and process reports
- Execute application logic
- Control database applications
Database Applications: Example Data Entry Form

Customer Data Entry Form

Customer Number: [redacted]
Customer Last Name: Johnson
Customer First Name: Ariel
Phone: 206-567-1234

Course Enrollment Data

<table>
<thead>
<tr>
<th>Course</th>
<th>Course Date</th>
<th>Fee</th>
<th>Amount Paid</th>
<th>Amount Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int Pastels</td>
<td>3/15/2013</td>
<td>$350.00</td>
<td>$350.00</td>
<td>$0.00</td>
</tr>
<tr>
<td>Adv Pastels</td>
<td>10/1/2013</td>
<td>$500.00</td>
<td>$250.00</td>
<td>$250.00</td>
</tr>
</tbody>
</table>
Database Applications: Example Query

(a) Query Parameter Form

(b) Query Results
## Course Enrollment Report

<table>
<thead>
<tr>
<th>Course</th>
<th>CourseDate</th>
<th>CustomerLastName</th>
<th>CustomerFirstName</th>
<th>Phone</th>
<th>Fee</th>
<th>AmountPaid</th>
<th>AmountDue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adv Pastels</strong></td>
<td>10/1/2013</td>
<td>Jackson</td>
<td>Charles</td>
<td>360-789-3456</td>
<td>$500.00</td>
<td>$500.00</td>
<td>$0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Johnson</td>
<td>Ariel</td>
<td>206-567-1234</td>
<td>$500.00</td>
<td>$250.00</td>
<td>$250.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pearson</td>
<td>Jeffery</td>
<td>206-567-2345</td>
<td>$500.00</td>
<td>$500.00</td>
<td>$0.00</td>
</tr>
<tr>
<td><strong>Beg Oils</strong></td>
<td>11/15/2013</td>
<td>Kyle</td>
<td>Leah</td>
<td>425-678-7654</td>
<td>$500.00</td>
<td>$250.00</td>
<td>$250.00</td>
</tr>
<tr>
<td><strong>Int Pastels</strong></td>
<td>9/15/2013</td>
<td>Green</td>
<td>Robin</td>
<td>425-678-8765</td>
<td>$350.00</td>
<td>$350.00</td>
<td>$0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sears</td>
<td>Miguel</td>
<td>360-789-4567</td>
<td>$350.00</td>
<td>$350.00</td>
<td>$0.00</td>
</tr>
<tr>
<td></td>
<td>10/15/2013</td>
<td>Myers</td>
<td>Lynda</td>
<td>360-789-5678</td>
<td>$350.00</td>
<td>$0.00</td>
<td>$350.00</td>
</tr>
<tr>
<td><strong>Int Pastels</strong></td>
<td>3/15/2013</td>
<td>Johnson</td>
<td>Ariel</td>
<td>206-567-1234</td>
<td>$350.00</td>
<td>$350.00</td>
<td>$0.00</td>
</tr>
</tbody>
</table>
Personal Database Systems

Personal database systems typically

- have one application.
- have only a few tables.
- are simple in design.
- involve only one computer.
- support one user at a time.
Personal Database Systems

User

Database application

Microsoft Access or other personal DBMS

Database management system (DBMS)

Database
Personal Database Systems: An SQL Query in Microsoft Access

The SQL has been arranged to make it easy to read

```
SELECT CUSTOMER.CustomerLastName, 
       CUSTOMER.CustomerFirstName, 
       CUSTOMER.Phone, 
       COURSE.CourseDate, 
       ENROLLMENT.AmountPaid, 
       COURSE.Course, 
       COURSE.Fee 
FROM CUSTOMER, ENROLLMENT, COURSE 
WHERE ((CUSTOMER.CustomerNumber)=[ENROLLMENT].[CustomerNumber]) 
     AND ((COURSE.CourseNumber)=[ENROLLMENT].[CourseNumber]));
```
Personal Database Systems:
SQL Query Results in Microsoft Access

- The database named **Art-Course-Database**
- The table object **CUSTOMER** is displayed under the All Access Objects
- The query object **Art Course List** stores the query itself
- The query results in table format
Enterprise-Class Database Systems

Enterprise-Class database systems typically:

- support several users simultaneously,
- include more than one application,
- involve multiple computers,
- are complex in design,
- have many tables, and
- have many databases.
Organizational Database Systems

Database application A
Java code

Database application B
C# code

Database application C
HTML and ASP.NET

Database management system (DBMS)

SQL Server (Microsoft)
Oracle Database (Oracle)
MySQL (Oracle)
Others

Database
Commercial DBMS Products

- Example of Desktop DBMS Products
  - Microsoft Access

- Examples of Organizational DBMS Products
  - Microsoft’s SQL Server
  - Oracle’s Oracle
  - Sun Microsystem’s MySQL
  - IBM’s DB2
Chapter 6: outline

6.1 Getting started
6.2 Entity and Relations
An entity is something of importance to a user that needs to be represented in a database.

An entity represents one theme or topic.

In an entity-relationship model, entities are restricted to things that can be represented by a single table.
A relation is a two-dimensional table that has specific characteristics. The table dimensions, like a matrix, consist of rows and columns.
Characteristics of a Relation

1. Rows contain data about an entity.
2. Columns contain data about attributes of the entity.
3. Cells of the table hold a single value.
4. All entries in a column are of the same kind.
5. Each column has a unique name.
6. The order of the columns is unimportant.
7. The order of the rows is unimportant.
8. No two rows may be identical.
# A Sample Relation

<table>
<thead>
<tr>
<th>EmployeeNumber</th>
<th>FirstName</th>
<th>LastName</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Mary</td>
<td>Abernathy</td>
</tr>
<tr>
<td>101</td>
<td>Jerry</td>
<td>Cadley</td>
</tr>
<tr>
<td>104</td>
<td>Alex</td>
<td>Copley</td>
</tr>
<tr>
<td>107</td>
<td>Megan</td>
<td>Jackson</td>
</tr>
</tbody>
</table>
### A Nonrelation Example

**Cells of the table hold multiple values**

<table>
<thead>
<tr>
<th>EmployeeNumber</th>
<th>Phone</th>
<th>LastName</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>335-6421, 454-9744</td>
<td>Abernathy</td>
</tr>
<tr>
<td>101</td>
<td>215-7789</td>
<td>Cadley</td>
</tr>
<tr>
<td>104</td>
<td>610-9850</td>
<td>Copley</td>
</tr>
<tr>
<td>107</td>
<td>299-9090</td>
<td>Jackson</td>
</tr>
</tbody>
</table>
Example of a Nonrelational Table

<table>
<thead>
<tr>
<th>EmployeeNumber</th>
<th>Phone</th>
<th>LastName</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>335-6421</td>
<td>Abernathy</td>
</tr>
<tr>
<td>101</td>
<td>215-7789</td>
<td>Cadley</td>
</tr>
<tr>
<td>104</td>
<td>610-9850</td>
<td>Copley</td>
</tr>
<tr>
<td>100</td>
<td>335-6421</td>
<td>Abernathy</td>
</tr>
<tr>
<td>107</td>
<td>299-9090</td>
<td>Jackson</td>
</tr>
</tbody>
</table>

*No two rows may be identical*
## Terminology

### Synonyms…

<table>
<thead>
<tr>
<th>Table</th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>File</td>
<td>Record</td>
<td>Field</td>
</tr>
<tr>
<td>Relation</td>
<td>Tuple</td>
<td>Attribute</td>
</tr>
</tbody>
</table>
A Key

- A **key** is one (or more) column(s) of a relation that is (are) used to identify a row.
## Uniqueness of Keys

<table>
<thead>
<tr>
<th>Unique Key</th>
<th>Nonunique Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data value is unique for each row. Consequently, the key will uniquely identify a row.</td>
<td>Data value may be shared among several rows. Consequently, the key will identify a set of rows.</td>
</tr>
</tbody>
</table>
A Composite Key

- A **composite key** is a key that contains two or more attributes.
- For a key to be unique, it must often become a composite key.
Composite Key Characteristics

- To identify a family member, you need to know a FamilyID, a FirstName, and a Suffix (e.g., Jr.).
- The composite key is: (FamilyID, FirstName, Suffix).
- One needs to know the value of all three columns to uniquely identify an individual.
A Candidate Key

- A candidate key is called “candidate” because it is a candidate to become the primary key.
- A candidate key is a unique key.
A primary key is a candidate key chosen to be the main key for the relation.

If you know the value of the primary key, you will be able to uniquely identify a single row.
A Surrogate Key

- A surrogate key is a unique, numeric value that is added to a relation to serve as the primary key.
- Surrogate key values have no meaning to users and are usually hidden on forms, queries, and reports.
- A surrogate key is often used in place of a composite primary key.
Surrogate Key Example

- If the Family Member primary key is FamilyID, FirstName, Suffix, it would be easier to append and use a surrogate key of FamMemberID.
- FamilyID, FirstName and Suffix remain in the relation.
Relationships Between Tables

- A table may be related to other tables.
- For example
  - An Employee works in a Department
  - A Manager controls a Project
A Foreign Key

- To preserve relationships, you may need to create a foreign key.
- A foreign key is a primary key from one table placed into another table.
- The key is called a foreign key in the table that received the key.
Foreign Key Example 1

**Project**
- ProjID
- ProjName
- MgrID

**Manager**
- MgrID
- MgrName

Primary Key

Foreign Key
Foreign Key Example II

Department
- DeptID
- DeptName
- Location

Employee
- EmpID
- DeptID
- EmpName

Primary Key
- DeptID

Foreign Key
- DeptID
Referential Integrity

- Referential integrity states that every value of a foreign key must match a value of an existing primary key.

- Example (see previous slide):
  - If EmpID = 4 in EMPLOYEE has a DeptID = 7 (a foreign key), a Department with DeptID = 7 must exist in DEPARTMENT.
  - The primary key value must exist before the foreign key value is entered.
Referential Integrity (Cont’d)

- Another perspective…

The value of the Foreign Key EmployeeID in EQUIPMENT

must exist in

The values of the Primary Key EmployeeID in EMPLOYEE
The Null Value

- A **Null value** means that no data was entered.
- This is different from a zero, space character, or tab character.
The Problem of Null Values

- A Null is often ambiguous. It could mean...
  - The column value is not appropriate for the specific row.
  - The column value is not decided.
  - The column value is unknown.
- Each may have entirely different implications.
Functional Dependency

- Functional Dependency—A relationship between attributes in which one attribute (or group of attributes) determines the value of another attribute in the same table

- Illustration…
  - The price of one cookie can determine the price of a box of 12 cookies.

\[(\text{CookiePrice, Qty}) \rightarrow \text{BoxPrice}\]
Determinants

- The attribute (or attributes) that we use as the starting point (the variable on the left side of the equation) is called a *determinant*.

\[(\text{CookiePrice, Qty}) \rightarrow \text{BoxPrice}\]

*Determinant*
Candidate/Primary Keys and Functional Dependency

- By definition…
  A candidate key of a relation will functionally determine all other attributes in the row.

- Likewise, by definition…
  A primary key of a relation will functionally determine all other attributes in the row.
Primary Key and Functional Dependency Example

(EmployeeID) → (EmpLastName, EmpPhone)

(ProjectID) → (ProjectName, StartDate)
Chapter 6: summary

- basic concepts
- relational databases
- SQL
- entities
- relations
- keys