CS 493
Secure Software Systems

Ch 5 The Database Environment

Dr. Williams
Central Connecticut State University

Objectives

- The purpose of conceptual, logical, and physical modeling
- The purpose of normalization
- Methods for gathering initial requirements in support of the overall database design process
- Understand the difference between Data Definition Language (DDL) and Data Manipulation Language (DML)

Goals

- Identify the differences between data and information.
- Explain normalization of relations.
- Define a data model.
- Understand Relational Theory.
- Use basic SQL commands.

Database Fundamentals

Modern organizations rely heavily on the transformation of data to information in order to make critical business decisions.

- **Data** is simply defined as “raw facts.”
- **Information** is data that has been organized into some useful format to help the entity make critical business decisions.

Database Fundamentals

- An **attribute** or **field** is the property used to describe a characteristic of the entity.
- The **database management system (DBMS)** is the software used to manage, create, and maintain the database.
- A **DBMS** could host several databases for an organization.
- The **primary key** is a unique identifying attribute or a combination of two attributes that identify the row in a relation.

Database Fundamentals

- Combining two attributes to form a primary key, this is referred to as a **composite key**.
- A **foreign key** is a field in the database that serves as a primary key in a different table or entity in the same database.
- The **schema** identifies the logical mapping of the entire database.
- The **subschema** is the visualization of the database as seen by the database user.

Database Fundamentals

- The first part is the **Data Definition Language (DDL),** which is used to develop the schema.
- **Data Manipulation Language (DML)** is used to manipulate the data in the database.
- In 1970, E. F. Codd of IBM developed the seminal paper titled “A Relational Model of Data for Large Shared Data Banks.” This paper produced the theory for the modern relational data model.

Database Fundamentals

- The relational model provided several objectives:
- A method to deal with consistency and redundancy issues relating to data management
- Data independence to keep an application separate from the data representation
- The concept of normalized data

Conceptual Design

- The database conceptual design is critical to any successful information system development project.
- **Conceptual modeling** is the process used to construct the architectural components of the database.
- In this phase the organization's data requirements are collected, detailed definitions are developed, and diagrams produced.

The Logical Design

- During the **logical design,** the developer takes conceptual design and implements the database into a logical data model.
- The **entity-relationship model** is the conceptual representation of the data in an organization.

Database Normalization

- **Normalization** is defined as a methodology used to develop well-organized entities based on the organization's information needs.
- This method is used to eliminate redundancy, anomalies, and inconsistency in the database.
- There are five normal forms. The first three, referred to as **First Normal Form (1NF)**, **Second Normal Form (2NF)**, and **Third Normal Form (3NF)**; they are the most common.

Unnormalized data

C_NUM	C_NAME	BOAT_NUM	LOCATION	LEASE_ST	LEASE_END	LEASE_C	OWNER_ID	OWNER_N
78	John Smith	NY102	11 East River	1-Jul-08	2-Aug-08	1200	BO 28	Jane Doe
		FL106	12 Dinner Key	1-Jul-07	6-Aug-07	860	BO 30	Jack James
81	Christy Jones	NY105	12 East River	30-Jun-09	29-Jul-09	1800	BO 28	Jane Doe
		NY102	11 East River	28-May-08	15-Jul-08	1200	BO 28	Jane Doe
		NY106	12 Dinner Key	4-Jul-10	4-Aug-10	860	BO 30	Jack James

Figure 5.3

Unnormalized data in the `BoatRental` table.

Database Normalization

- We want to normalize the table by ensuring that there is single value at each row and column.

First Normal Form:

- Eliminate repeating groups.
- Identify each entity with a primary key.

1NF

Eliminate repeating groups.
Identify each entity with a primary key.

C_NUM	C_NAME	BOAT_NUM	LOCATION	LEASE_ST	LEASE_END	LEASE_C	OWNER_ID	OWNER_N
78	John Smith	NY102	11 East River	1-Jul-08	2-Aug-08	1200	BO 28	Jane Doe
78	John Smith	FL106	12 Dinner Key	1-Jul-07	6-Aug-07	860	BO 30	Jack James
81	Christy Jones	NY105	12 East River	30-Jun-09	29-Jul-09	1800	BO 28	Jane Doe
81	Christy Jones	NY102	11 East River	28-May-08	15-Jul-08	1200	BO 28	Jane Doe
81	Christy Jones	NY106	12 Dinner Key	4-Jul-10	4-Aug-10	860	BO 30	Jack James

Figure 5.4

Normalized `BoatRental` relation.

Database Normalization

Second Normal Form: The Second Normal Form aims to remove partial dependencies and to avoid update anomalies.

- Relation or table is in 1NF
- Eliminate partial dependencies

2NF

Eliminate partial dependencies

RENTER

C_NUM	C_NAME
78	John Smith
81	Christy Jones

RENTAL

C_NUM	BOAT_NUM	LEASE_ST	LEASE_END
78	NY102	1-Jul-08	2-Aug-08
78	FL106	1-Jul-07	6-Aug-07
81	NY105	30-Jun-09	29-Jul-09
81	NY107	28-May-08	15-Jul-08
81	NY103	4-Jul-10	4-Aug-10

RENTAL OWNER

BOAT_NUM	LOCATION	OWNER_ID	LEASE_C	OWNER_N
NY102	11 East River	BO 28	1200	Jane Doe
FL106	12 Dinner Key	BO 30	860	Jack James
NY105	12 East River	BO 28	1800	Jane Doe

Figure 5.5

(2NF) Second Normal Form.

Database Normalization

Third Normal Form: Your relations are in 3NF if they are all in 2NF and all transitive dependencies have been removed.

- Relations or tables are in 2NF.
- All transitive dependencies have been removed.

3NF

All transitive dependencies have been removed.

RENTER

C_NUM	C_NAME
78	John Smith
81	Christy Jones

RENTAL

C_NUM	BOAT_NUM	LEASE_ST	LEASE_END
78	NY102	1-Jul-08	2-Aug-08
78	FL106	1-Jul-07	6-Aug-07
81	NY105	30-Jun-09	29-Jul-09
81	NY107	28-May-08	15-Jul-08
81	NY103	4-Jul-10	4-Aug-10

BoatRental

BOAT_NUM	LOCATION	OWNER_ID	LEASE_C
NY102	11 East River	BO 28	1200
FL106	12 Dinner Key	BO 30	860
NY105	12 East River	BO 28	1800

BoatOwner

OWNER_ID	OWNER_N
BO 28	Jane Doe
BO 30	Jack James

Figure 5.7

(3NF) New normalized relations.

The Physical Design

- The **physical design** of a database management system is the portion of the database design process when the developer decides on the hardware needed, DBMS platform, indexing and file organization, and transformation of the entity relationship diagrams into relations.
- A performance and tuning plan would be developed at this stage.

Introduction to SQL

- SQL (Structured Query Language) commands. SQL was developed as a direct result of the System R project at IBM's San Jose research labs, known as SEQUEL.
- The **nonprocedural language** allows the database administrator to conduct powerful tasks using relatively simple commands, unlike procedural languages where many steps must be developed toward the execution of tasks.

Introduction to SQL

SQL is practically the same in all commercial RDBMS platforms with the exception of a few syntax differences. There are three sections:

- Data Definition Language (DDL)
  - Creation, modification, deletion of tables and keys
- Data Manipulation Language (DML)
  - Creation, modification, deletion of records
- Data Control Language (DCL)
  - Manipulation of of access/authorization

Sample Table Commands

Syntax used to create a new relation:

```
SQL> CREATE TABLE EMPLOYEES (
  2  CUST_NUM              CHAR(15) NOT NULL,
  3  F_NAME                VARCHAR2 (19) NOT NULL,
  4  ADDRESS               VARCHAR2 (25),
  5  CITY                  CHAR(15),
  6  STATE                 CHAR(2),
  7  ZIP                   CHAR(5),
  8  PHONE                 VARCHAR2(8)
  9  );

Table created.

SQL> DROP TABLE EMPLOYEES;

Table dropped.

SQL>
```

Data manipulation

Select SQL cheat sheet

```
select * from PERSON

select id,firstname from PERSON

select * from PERSON where id=?

select * from PERSON where
  firstname like ?
```

? - "Bob%"

Join tables:

```
select P.firstname, A.city from
  PERSON P, ADDRESS A where P.ID =
  A.PERSONID
```

Insert statement cheatsheet

```
insert into PERSON values
  (?,?,?)
```

Note very problematic as depends on order of database columns if DBA changes these it breaks all of your code or worse

```
insert into PERSON
  (id,firstname) values(?,?)
```

Update statement cheatsheet

```
update PERSON set firstname=?,
  lastname=? WHERE id=?

update PERSON set firstname=?
  WHERE lastname=? AND age=?
```

Note like delete, very dangerous if WHERE condition isn't specific enough

Delete statement cheatsheet

```
delete from person where id=?

delete from person where
  firstname=? AND lastname=?
```

Very dangerous if WHERE condition isn't specific enough

The User Interface

The user interface, also known as the front end of the application, is the portion of the information system that the end user interacts with to access the information within the database management system.

```
<?php

$conn = oci_connect('myusername', 'mypassword', 'myhost/XE');

if (!$conn) {
trigger_error("Could not connect to database", E_USER_ERROR);
}

?>
```

Web Applications and the Internet

- A web application is an application that end users access by using the Internet.
- The Internet is a worldwide network of computer networks using the TCP/IP addressing scheme to interconnect millions of users.
- Tim Berners-Lee developed the application that we know as the World Wide Web using the Objective-C® language.

Summary

- The purpose of this chapter has been to familiarize you with the overall method of database design.
- The SQL portion of this chapter covers the fundamental foundation of the language.
- Key concept from chapter is database environment holds application's data and information
  - Without protection, anything that can be done with DDL, DML, DCL could potentially be attacked from the user interface

In class exercises revisited

Bob's Pizza Shack

In groups consider this scenario and identify in general what data elements are going to have security concerns

- Bob is a small business owner of Bob's Pizza Shack and wants to create a website to allow online credit card delivery orders

Alice's Online Bank

In groups consider this scenario and identify in general what data elements are going to have security concerns

- Alice opens Alice's Online Bank (AOB)
- What are Alice's security concerns from a network perspective?
  - Consider ones in own environment
  - Consider ones interoperating with another bank
  - Consider ones interacting with customer
    - Web
    - Mobile app

Location based social media app

In groups consider this scenario and identify in general what data elements are going to have security concerns

- Open source group wants to create a mobile app to allow groups to communicate/find each other in public demonstrations/protests
  - Communication internet, as well as, P2P (WiFi/Bluetooth) in case internet cut off – so if person you want to contact is on other side of crowd and no internet as long as P2P network can be established with app can reach somebody outside your immediate vicinity via the P2P network
  - Should be able to communicate securely messages and images to people you identify within your group (group as whole or direct)
  - Should be able to share GPS location with people in group (group as whole or direct)
- Broader scope – Who are potential threats and associated ways could attack/weaken system?