Relational Databases and SQL

An Introduction

Acknowledge: Chris Smith

Outline: Relational Model and SQL

1. The Relational Model
   - History
   - The Relational Model Summarized
   - Tables and Keys
   - Relational Algebra
2. SQL
   - History
   - Data Manipulation Language
   - Data Definition Language
3. Relational Databases.
   - What are they?
   - Why use one?

The Relational Model: History

- Derivability, Redundancy and Consistency of Relations Stored in Large Data Banks. E.F.Codd, IBM Research Report RJ599 (August 1969)
- A Relational Model of Data for Large Shared Data Banks. E.F.Codd, CACM 13 No. 6 (June 1970)
- Research and systems developed in the 1970's. (e.g. Ingres, Oracle)

The Relational Model

- Summary of Codd's work:

  Data should be represented as relations (tables).

item_table				
item_no	description	cost	price	on_hand
011654	Mug	3.50	9.75	150
011665	Cup	2.75	6.54	225
011776	Bowl	5.98	12.34	112
011887	Serving bowl	10.59	27.00	40

Properties of Tables

- A table has a unique name (in some scope).
- Each cell of the table can contain an “atomic” value only.
- Each column has a unique name (within the table).
- Values in a column all come from the same domain.
- Each row in the table is distinct.
  - Part of the model but not actually enforced!

Relational Model: Jargon

Relational Model (Formal)	Alternative 1	Alternative 2
Relation	Table	File (not common)
Tuple	Row	Record
Attribute	Column	Field

We will generally use Alternative 1.

Defining a Table

- A table is defined by giving a set of attribute and domain name pairs.
- This is called a Table Schema (or Relation Schema).

Keys

- For practical purposes we want to be able to identify rows in our tables.
  - We use keys for this.
- A key is just a set of columns in the table.
- Quite frequently just one column is enough, and quite often it is obvious what it should be.

Keys: Jargon

Superkey	A set of columns that uniquely identifies a row.
Candidate Key	An irreducible superkey (no subset of the columns uniquely identifes the table rows).
Primary Key	A selected candidate key.
Foreign Key	A set of columns within one table that are a candidate key for some other table.

NULL Values

- A special value “NULL” is provided to allow for cells in a table that have an unspecified value.
- NULL is not the same as zero or the empty string, but represents complete absence of a value.
- No part of a primary key may be NULL.

Relational Algebra

- We have seen how to define tables (relations). We want to be able to manipulate them too.
- “The relational algebra is a theoretical language with operations that work on one or more relations to define another relation without changing the original relation(s).” (“Database Systems” Connolly and Begg.)

Relational Algebra: Unary Operations

- Selection
  - Take a subset of rows from a table (on some criterion).
- Projection
  - Take a subset of columns from a table.

Relational Algebra: Binary Operations 1

- Union
  - Return all rows from two tables.
  - The two tables must have columns with the same domains (union compatibility).
- Intersection
  - Return all matching rows from two tables.
- Difference
  - Return all rows from one table not in another.
  - The two tables must be union compatible.

Relational Algebra: Binary Operations 2

- Cartesian Product
  - Concatenate every row from one table with every row from another.
- Join
  - Not really a separate operation: can be defined in terms of cartesian product and selection.
  - Is very important.

Relational Database Management System (RDBMS)

- Implements the relational model and relational algebra (under the covers).
- Provides a language for managing relations.
- Provides a language for accessing and updating data.
- Provides other services:
  - Security
  - Indexing for efficiency.
  - Backup services (maybe).
  - Distribution services (maybe).

RDBMS Implementation

- An RDBMS is usually implemented as a server program.
- Client programs communicate with the server (typically using TCP/IP).

SQL History

- Structured Query Language.
- Officially pronounced S-Q-L, but many people say “sequel”.
- Has its roots in the mid-1970’s.
- Standardized in 1986 (ANSI), 1987 (ISO)
- Further standards in 1992 (ISO SQL2 or SQL-92), 1999 (ISO SQL3).

SQL Today

- SQL is the only database language to have gained broad acceptance.
- Nearly every database system supports it.
- The ISO SQL standard uses the “Table, Row, Column” terminology rather than “Relation, Tuple, Attribute”.
- Some debate about how closely SQL adheres to the relational model.

SQL

- SQL is divided into two parts:
  - Data Manipulation Language
  - Data Definition Language
- Originally designed to be used from another language and not intended to be a complete programming language in its own right.
- Non-procedural. Define what you want, not how to get it.
- Supposed to be “English Like”!

SQL

- Go to SQL note

Creating a Database

- Creation of an entire database tends to depend on the RDBMS being used.
- Usually allow multiple named databases to be accessed through a single instance of a database server.

When to Use an RDBMS?

- Good for large amounts of data.
  - Indexing capabilities.
- Frequent updates:
  - Insertions of new values
- Many different views of the data wanted.
- Associations between different entities (foreign keys).
- Data integrity.
  - Constraints.
  - Transactions.
    - ACID = Atomicity, Consistency, Isolation, Durability.
- Integration with other systems e.g. web pages.
- Sharing data between users.

Plain Old Text Files

- Can be perfect (even for largish amounts of data).
- Easier to hand over to someone else.
  - Don’t have to say “first install database X”.
- Not great for updates to existing values.
- No integrity checks (can be made in code).

References

- “Database Systems”, Connolly and Begg, Addison Wesley, 3rd Edition, 2002
- “PostgreSQL”, Douglas and Douglas, SAMS Publishing, 2003
- “MySQL”, DuBois, SAMS Publishing, 2nd Edition, 2003
- http://www.postgresql.org
  - Recommended website for further reading.
- http://www.mysql.com

Summary

- RDBMS's are good at manipulating data.
- Need to decide if you need one.
- SQL is the standard language.
  - Standard up to a point.