ECE:5995 Modern Databases

Intro to RDBMS, Postgres, and CRUD

Relational databases

- Relational database management systems (RDBMS) store data in tables or relations
- A table is a collection of related data held in a structured format within a database.
- Tables are uniquely identified by their names
- Tables (relations) are comprised of columns (attributes) and rows (tuples).
  - Columns contain the column name, data type, and any other attributes.
  - Rows contain the records or data for the columns.

Last Name	First Name	Address	City
Jones	Sarah	123 Main Street	Orlando
Smith	Penny	567 First Street	Tampa
Reed	Gary	890 Third Street	Jacksonville

Structured Query Language (SQL)

- Standard for RDBMS
- Originally proposed by E.F.Codd in 1970
- Oracle adopted in late 1970s
- IBM's SQL/DS in 1981, and DB2 in 1983

- ANSI and ISO published SQL standards in 1986, called SQL-1, later revised SQL-89, SQL-2 (1999), SQL-3 (2003, 2008, 2011, and 2016)

Used today in PostgreSQL, Microsoft SQL Server, MySQL, Informix, Sybase, dBase, Paradox, r:Base, FoxPro, and others

Most vendors support standard, but have slight variations of their own

Components of SQL

- Data definition language – DDL

CREATE TABLE	CREATE INDEX
ALTER TABLE	RENAME TABLE
DROP TABLE	DROP INDEX
CREATE VIEW	

- Data manipulation language – DML
  - Create, Read, Update, Delete (CRUD)

    INSERT
    SELECT
    UPDATE
    DELETE

- Authorization language – grant privileges to users

Relationships between tables

- Larger tables can be broken into smaller tables and still maintain a relationship
- We can create relationships between two tables, through matching the data from one column in a table with the data in a column in the second table
- Two SQL relational concepts that allows to do this:
  - **Primary Key**. Primary Key is a column or a combination of columns that uniquely identifies each row in a table.
  - **Foreign Key.** Foreign Key is a column or a combination of columns whose values match a Primary Key in a different table.
- Three types of relations between tables:
  1. **One-To-Many** (Most Common) - a row in one of the tables can have many matching rows in the second table, but a row the second table can match only one row in the first table
  2. **Many-To-Many** - many rows from the first table can match many rows in the second and the other way around
  3. **One-To-One** - each row in the first table may match only one row in the second and the other way around

Data Integrity

- Data integrity refers to the overall completeness, accuracy and consistency of data.
- Data integrity is usually imposed during the database design phase through the use of standard procedures and rules as part of normalization.

- Three integrity constraints to achieve data integrity:
  1. **Entity Integrity**: Every table must have its own primary key and that has to be unique and not null
  2. **Referential Integrity**: This is the concept of foreign keys. The rule states that the foreign key value can be in two states. The first state is that the foreign key value would refer to a primary key value of another table, or it can be null. Being null could simply mean that there are no relationships, or that the relationship is unknown.
  3. **Domain Integrity**: This states that every column in a relational database is in a defined domain.

- The concept of **data integrity ensures that all data in a database can be traced and connected to other data**.
- **Normalization** ensures data integrity.

PostgreSQL

- PostgreSQL is an enterprise-class, open-source RDBMS
- More than 20 years of development by the open-source community.
- PostgreSQL project started in 1986 at Berkeley, named POSTGRES, in reference to the older Ingres database which also developed at Berkeley.
- PostgreSQL support most popular programming languages:
  - Python, Java, C#, C/C+, Ruby, JavaScript (Node.js), Perl, Go, and others
- PostgreSQL offer many advanced features:
  - User-defined types, table inheritance
  - Foreign key referential integrity, views, rules, subquery
  - Sophisticated locking mechanism, nested transactions (savepoints)
  - Multi-version concurrency control (MVCC), asynchronous replication
- Many companies have built products and solutions based on PostgreSQL. Some featured companies are Apple, Fujitsu, Red Hat, Cisco, Juniper Network, Instagram, etc.

Let's give it a try

Webserver:

https://s-l112.engr.uiowa.edu/phpPgAdmin/

NOTE: The Postgres databases are hosted on a campus subnet and will only be accessible through eduroam or VPN.

Check ICON for your group number under the postgres groups, i.e. if you are in group 1, your username will be student01 and password engr-2020-01)
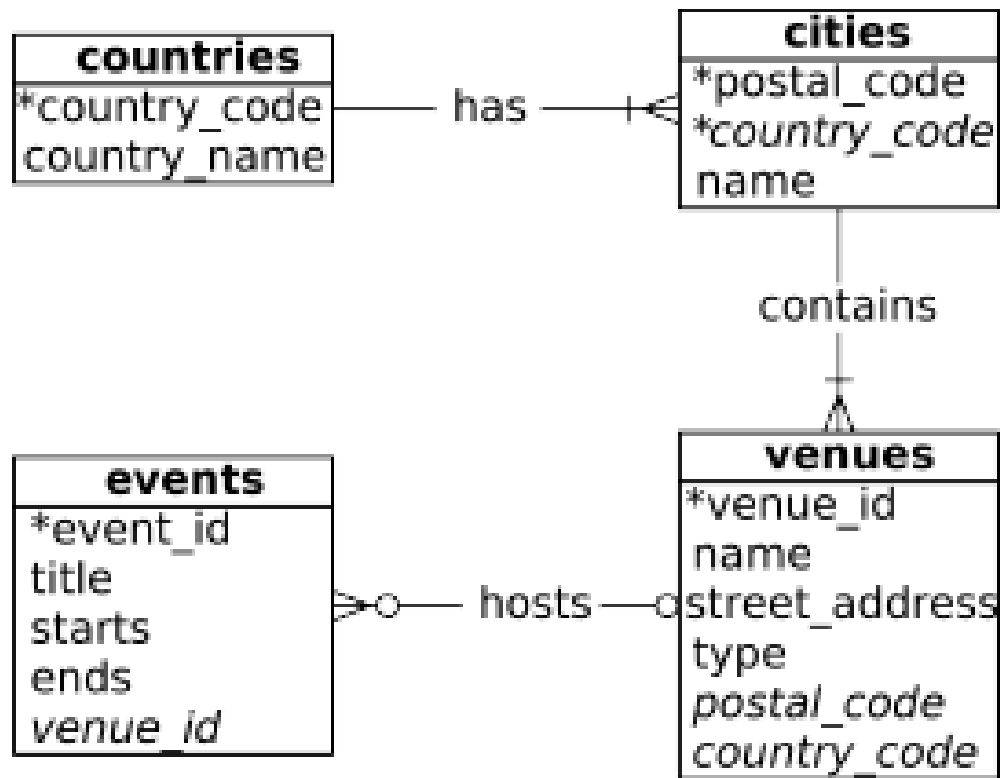
username: studentXX
password: engr-2020-XX

We will be using phpPgAdmin to interact with the database.

If you want to setup psql or pgAdmin in your own computer:
https://www.enterprisedb.com/postgres-tutorials/connecting-postgresql-using-psql-and-pgadmin

hostname: s-l112.engr.uiowa.edu

Today's example

countries
*country_code
country_name

— has —

cities
*postal_code
*country_code
name

contains

venues
*venue_id
name
street_address
type
postal_code
country_code

events
*event_id
title
starts
ends
venue_id

— hosts —

Create a new table named **events**. Your events table should have these columns: a SERIAL integer event_id (primary key), a title, starts and ends (of type timestamp), and a venue_id (foreign key that references **venues**).

After creating the **events** table, INSERT the following values (timestamps are inserted as a string like *2018-02-15 17:30*) :

title	starts	ends	venue_id	event_id
Fight Club	2018-02-15 17:30:00	2018-02-15 19:30:00	2	1
April Fools Day	2018-04-01 00:00:00	2018-04-01 23:59:00		2
Christmas Day	2018-02-15 19:30:00	2018-12-25 23:59:00		3

Today's terms

Term	Definition
Column	A domain of values of a certain type, sometimes called an *attribute*
Row	An object comprised of a set of column values, sometimes called a *tuple*
Table	A set of rows with the same columns, sometimes called a *relation*
Primary key	The unique value that pinpoints a specific row
Foreign key	A data constraint that ensures that each entry in a column in one table uniquely corresponds to a row in another table (or even the same table)
CRUD	Create, Read, Update, Delete
SQL	Structured Query Language, the *standard* language of a relational database
Join	Combining two tables into one by some matching columns