Welcome to MSDS 597!

Week 1 Schedule

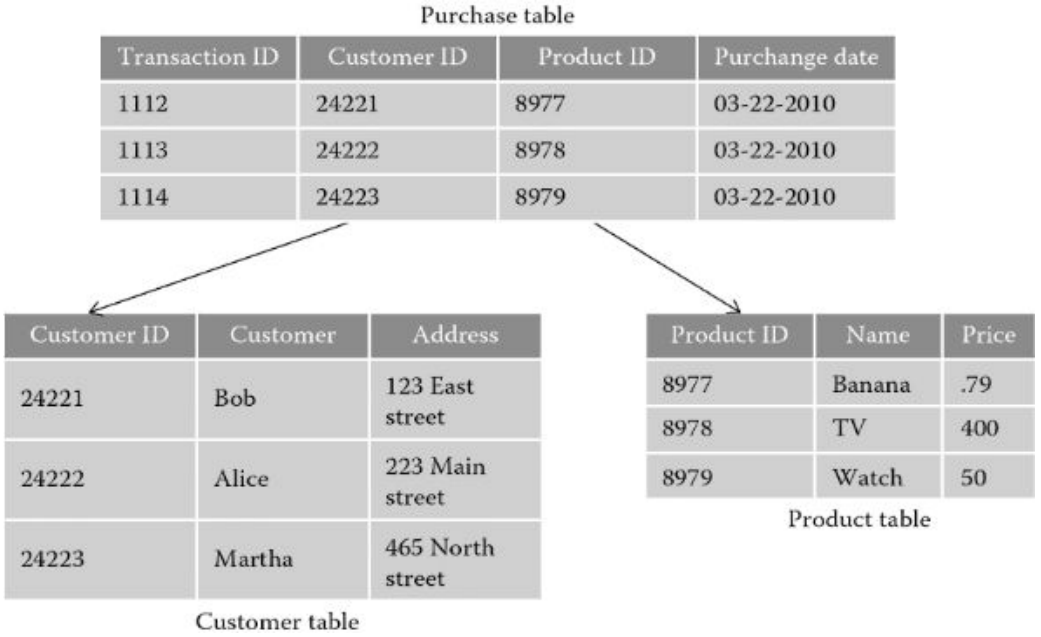
- [6:30 - 6:35] Introduction
- [6:35 - 7:05] Review Syllabus
  - SQL survey
- [7:10 - 7:30] Lecture Part 1: Intro to SQL
- [7:30 - 8:15] Lab 1 (Getting set up)
- [8:15 - 9:20] Lecture Part 2: SQL Fundamentals
- [9:20 - 9:30] Lookahead to Week 2
  - HW1
  - Reading
  - Next week schedule

Who am I?

- Call me Ken
- Graduate of USF's MSDS program (Class of 2015)
- Data Engineer at FB
  - Currently focused on recruiting analytics
- No data/analytics work experience prior to the Master's degree
- Goal is to make analytics as accessible and as fun as possible!

What is a relational database?

- Database: a collection of data
- Relational Database
  - Follows a specific data model known as the “Relational Data Model”
- Consist of the following
  - Rows
  - Columns
  - Tables
  - Schema
- Relational database is a collection of schemas



Purchase table

Transaction ID	Customer ID	Product ID	Purchange date
1112	24221	8977	03-22-2010
1113	24222	8978	03-22-2010
1114	24223	8979	03-22-2010

Customer ID	Customer	Address
24221	Bob	123 East street
24222	Alice	223 Main street
24223	Martha	465 North street

Customer table

Product ID	Name	Price
8977	Banana	.79
8978	TV	400
8979	Watch	50

Product table

What is a relational database?

- Strongly typed
  - Each column is a specific data type
    - Storage space
    - Rules
- Work well when you have structured data
  - Structured: Ex. Age, first name, last name
  - Not Structured: Ex. Freeform text (like essay responses)
- Accessed using SQL
- Use client-server systems
  - Client and server can be on the same physical computer
  - Server can be in the cloud and a user can be anywhere
  - Server can be on premises and the client could be physically proximal

What is SQL?

SQL ...

- stands for Structured Query Language
- can be pronounced S-Q-L or like Sequel
- lets you access and manipulate data in RDBMS (relational database management systems)
- developed at IBM in the early 1970s
- became a standard of ANSI in 1986 and ISO in 1987

What can you do with SQL?

- Execute queries against a database
- Retrieve data from a database
- Insert records into a database
- Update records in a database
- Delete records from a database
- Create new tables and new databases
- ... and more

Why learn SQL?

- Working with data in large databases
- Ubiquitous in industry
- Required in all data related work
- Frequently tested in technical interviews
- Concepts translate well to many facets of analytics

DDL vs DML

Data Definition Language

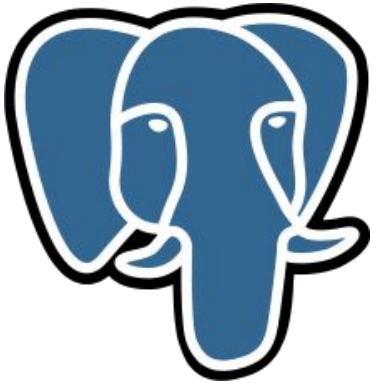
- Syntax for creating and modifying database objects (e.g. tables)
- CREATE TABLE
- ALTER TABLE
- DROP TABLE

Data Manipulation Language

- Syntax for adding, deleting, updating and reading data
- SELECT, INSERT INTO, UPDATE, DELETE FROM

What will we be using in this course?

- **PostgreSQL**
  - AKA Postgres
  - Free and open source RDBMS emphasizing SQL compliance
  - Second most used DB after MySQL
- **DataGrip IDE**
  - IDE = Integrated Development Environment
    - Acts as a dev environment for your SQL coding
    - Text editor, smart code completion, data loading, query console etc.
  - Install locally on your computer
  - Free for educational purposes



Lab 1

- Install PostgreSQL
- Install Datagrip

Find instructions in “Lab 1” in Modules

Selecting Columns

- Query Template

  SELECT
  <comma separated list of columns>
  FROM <table name>
  ;

- Query starts with SELECT
  - States which columns of data we are interested in
- FROM clause specifies the table to use
- SQL ignores white space and is case insensitive
  - However, there are some style rules we should follow to make the query more readable
- Use * to return **all** columns

Test it out

Write queries for the following

- Select all columns from ramen_ratings
- Select **brand** and **country** from ramen_ratings

Filtering Rows

```
SELECT
    name,
    age
FROM records
WHERE
    name = ‘John’  -- Use single quotes for strings.  Strings are case sensitive
```

- Use the WHERE clause to select specific rows
- Use single quotes for strings. Double quotes have a special meaning in SQL and are used for referencing DB objects like tables, columns, and schemas.
- Use any standard comparison operator for numbers (=, <, >, <=, >=)
- Inequality can be stated with either != or <>
- Combine multiple criteria using AND or OR
- Use parentheses to dictate the order of operations

Test it out

Write queries for the following

- Select all columns from ramen_ratings with greater than or equal to 4.5 stars
- Select all ramen from Nigeria
- Select all “Bowl” style ramen from Japan with less than 3 stars

NULL values

NULL values represent **missing** values. They are the absence of a value and thus are a special class of data in the DB.

Use IS NULL or IS NOT NULL to test for NULL values

Phrase A	Phrase B	A OR B	A AND B	A = B
True	True	True	True	True
True	False	True	False	False
True	Null	True	Null	Null
False	True	True	False	False
False	False	False	False	True
False	Null	Null	False	Null
Null	True	True	Null	Null
Null	False	Null	False	Null
Null	Null	Null	Null	Null

ORDER BY and LIMIT

```
SELECT
    *
FROM transactions
ORDER BY price DESC
LIMIT 5
```

Default ordering is ASC (ascending)
Multiple sort orders can be combined
LIMIT sets the number of rows to return

Test it out

Write queries for the following

- Top 5 Korean ramen by star rating
- Bottom 5 USA ramen by star rating

Column Numbering

```
SELECT
        product_name,
        price
FROM transactions
ORDER BY 2 DESC
LIMIT 5
```

Column number can be used in lieu of column name in the ORDER BY clause

Where are we?

Most relational databases use dot notation to specify objects within the database

Database.schema.table.column

It’s recommended to select your database and schema within DataGrip to reduce the amount of typing required to access columns and tables.