MSiA-413 Introduction to Databases and Information Retrieval

Lecture 7
SQLite Tutorial

Instructor: Nikos Hardavellas

Slides adapted from Steve Tarzia

Last Lecture

- Extended ER Diagrams
- SELECT query steps
  - `GROUP BY`
  - Aggregation functions: `AVG()`, `SUM()`

In this tutorial

- We will practice SELECT statements
- We will introduce:
  - DB Browser for SQLite
  - Aggregation functions: `COUNT()`, `MIN()`, `MAX()`, `SUM()`
  - More complex filters: `WHERE` *`cond`* `IN (`*`cond1`*`,` *`cond1`*`, …)`
  - `DISTINCT`
  - `LIMIT <count> OFFSET <skip>`
  - Subqueries
  - String manipulation (pattern matching, concatenation)

SQLite

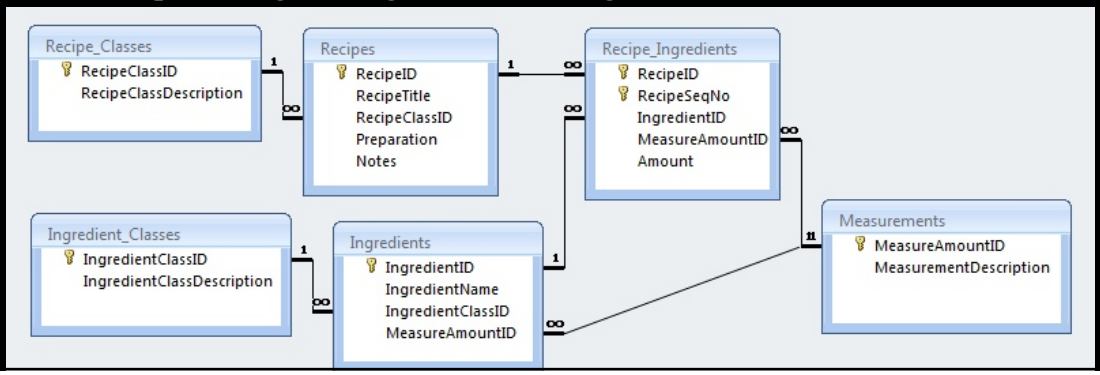
- A lightweight and easy-to-setup database
- Similar to Microsoft Access, but free and more portable
- Can handle very large data sets (terabytes)
- The whole database is stored in a single file (.db or .sqlite)
- But SQLite *does not* handle remote access from multiple users

A good choice for an individual needing to set up his/her own database

- Download it from http://sqlitebrowser.org

Recipes.sqlite (download it from Canvas)

- Print an alphabetically sorted list of ingredients
- How many times is butter used as an ingredient?
- How many ingredients are in the Yorkshire Pudding recipe?
- What percentage of ingredients are used in recipes?
- What percentage of ingredients are vegan?



Recipes.sqlite (answers)

- Print an alphabetically sorted list of ingredients

```
SELECT IngredientName FROM Ingredients ORDER BY IngredientName;
```

- How many times is butter used as an ingredient?

```
SELECT IngredientID FROM Ingredients WHERE IngredientName="Butter";
```

(it answers 47)

```
SELECT COUNT(*) FROM Recipe_Ingredients WHERE IngredientID = 47;
```

Another way (using a **subquery**):

```
SELECT COUNT(*)
FROM Recipe_Ingredients
WHERE IngredientID =
   (SELECT IngredientID FROM Ingredients
    WHERE IngredientName="Butter");
```

- How many ingredients are in the Yorkshire Pudding recipe?

```
SELECT COUNT(*) FROM Recipe_Ingredients WHERE RecipeID=10;
```

Recipes.sqlite (answers)

- What percentage of ingredients are used in recipes?

```
SELECT 100.0 * COUNT(DISTINCT Recipe_Ingredients.IngredientID)
        / (SELECT COUNT(Ingredients.IngredientID)
           FROM Ingredients)
FROM Recipe_Ingredients;
```

- What percentage of ingredients are vegan?

```
SELECT 100.0 * COUNT(IngredientID)
        / (SELECT COUNT(IngredientID)
           FROM Ingredients)
FROM Ingredients
WHERE IngredientClassID NOT IN (2, 8, 10, 11, 12);
```


Recipes.sqlite – string manipulation

- How many recipes have multi-word names?
- How many recipes have two-word names?
- How many recipes have nine-letter names?

Recipes.sqlite (string manipulation answers)

- How many recipes have multi-word names?

```
SELECT COUNT(*) FROM Recipes
WHERE RecipeTitle LIKE "% %";
```

- How many recipes have two-word names?

```
SELECT COUNT(*) FROM Recipes
WHERE RecipeTitle LIKE "% %" AND RecipeTitle NOT LIKE "% % %";
```

- How many recipes have nine-letter names?
  -
```
SELECT COUNT(*) FROM Recipes
WHERE RecipeTitle LIKE "_________";
```
  -
```
SELECT COUNT(*) FROM Recipes
WHERE LENGTH(RecipeTitle) = 9;
```


SalesOrders.sqlite (download it from Canvas)

- List all customers in California (CA)
- List all customers in a west coast state (CA, OR, WA)
- Count the unique customer area codes in California (CA)
- What is the full address of customer John Viescas?
- What is the most expensive product? Cheapest 5?
- What is the value of the product inventory on hand? Bike inventory?

SalesOrders.sqlite (answers)

- List all customers in California (CA)

```
SELECT * FROM Customers WHERE CustState = "CA";
```

- List all customers in a west coast state (CA, OR, WA)

```
SELECT * FROM Customers
WHERE CustState IN ("CA","OR", "WA");
```

- Count the unique customer area codes in California (CA)

```
SELECT COUNT(DISTINCT CustAreaCode)
FROM Customers
WHERE CustState = "CA";
```


SalesOrders.sqlite (answers)

- What is the full address of customer John Viescas?

```
SELECT CustStreetAddress || ", " || CustCity || ", " ||
       CustState || " " || CustZipCode
 AS FullAddress
FROM Customers
WHERE CustFirstName = "John" AND CustLastName = "Viescas";
```

  Note: concatenation works with any data type

- What is the most expensive product? Cheapest 5?
  - `SELECT ProductNumber, ProductName FROM Products WHERE RetailPrice = (SELECT MAX(RetailPrice) FROM Products);`
  - `SELECT ProductName, RetailPrice FROM Products ORDER BY RetailPrice LIMIT 5;`
- What is the value of the product inventory on hand? Bike inventory?
  - `SELECT SUM(RetailPrice * QuantityOnHand) FROM Products;`
  - `SELECT SUM(RetailPrice * QuantityOnHand) FROM Products WHERE CategoryID=2;`

SchoolScheduling.sqlite (download it from Canvas)

- What is the mean average classroom capacity? Median?
- How much classroom capacity is there in each building?
  (Hint: use "`GROUP BY BuildingCode`")
- How many classes does each instructor teach on average?
- What is the average grade earned by students?

SchoolScheduling.sqlite (answers)

- What is the average classroom capacity? Median?
  - `SELECT AVG(Capacity) FROM Class_Rooms;`
  - `SELECT Capacity FROM Class_Rooms ORDER BY Capacity LIMIT 1 OFFSET (SELECT COUNT(*)/2 FROM Class_Rooms);`
- How much classroom capacity is there in each building?

```
SELECT BuildingCode, SUM(Capacity) FROM Class_Rooms
   GROUP BY BuildingCode;
```

- How many classes does each instructor teach on average?

```
SELECT AVG(NumClasses) FROM
    (SELECT COUNT(*) AS NumClasses
     FROM Faculty_Classes GROUP BY StaffID);
```

- What is the average grade earned by students?

```
SELECT AVG(Grade) FROM Student_Schedules WHERE Grade > 0;
```