EECS-317 Data Management and Information Processing

## Lecture 3 – More SQL

Steve Tarzia

Spring 2019

Northwestern

Announcements

- First HW assignment was posted, due Monday night.
- Datacamp online lessons actually seem to require a $20 payment.
  - I updated the assignment to give an alternative free options which was suggested by a student:
    - https://pgexercises.com/questions/basic/
    - https://pgexercises.com/questions/joins/
    - https://pgexercises.com/questions/aggregates/

Last lecture: SQL Basics

- Showed syntax diagram for SELECT.
- Described my 7-step process for building a SELECT query.
  - Start with a short query:
    ```
    SELECT * FROM some_table
    ```
  - Gradually refine the results, making the query more complex.
  - Choose table, filter, choose columns, apply mathematic operations, sort, etc.
- Demonstrated how to build a query to answer a few questions about recipes.
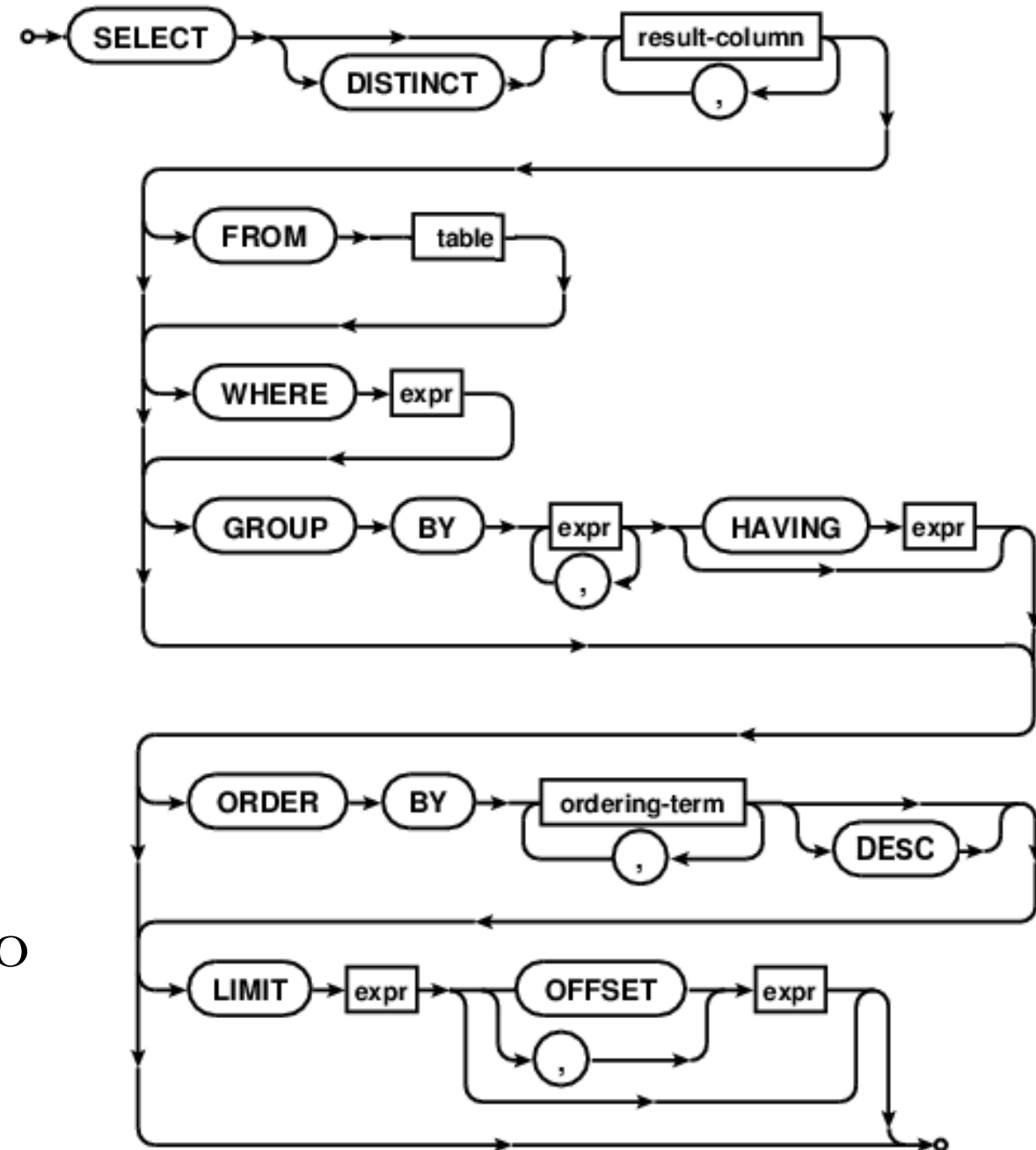
SELECT steps (abbreviated)

1. FROM chooses the table of interest
2. WHERE throws out irrelevant rows
3. GROUP BY identifies rows to combine
4. SELECT tells what values to return (allowing math and aggregation)
5. HAVING throws out irrelevant rows (after aggregation)
6. ORDER BY sorts
7. LIMIT throws out rows based on their position in the results

Each step gets closer to the specific result you want.

Integer vs. *floating point* division

- Computers store numbers in two basic ways:
    - **Integers** are whole numbers (0, 3, -40,921)
    - **Floating Point** numbers (*floats*) can be fractional (1.234, 0.0, -9.9×10^{-4})
- When doing arithmetic on two integers, an integer is always produced.
    - 1+1 = 2, 2-1=1, 4*3=12, **13/4=3**
- When doing arithmetic involving at least one float, a float is produced.
    - 1.0 + 1.0 = 2.0, 1.5 * 2 = 3.0, 13/4.0=3.25
- *Integer division is weird* – it always rounds down: **2/3 = 0**, **-5/2 = -3**
- Usually you need floating-point (not integer) division in your queries.
    - Just precede the expression with a floating point operation to force the division to be floating point: **1.0 * -5 / 2 = -2.5**

Aggregation functions

- COUNT, SUM, MIN, MAX, AVG
- These can be used to print out values that depend on multiple rows.
- For example, how many ounces of ingredients are used?
  - We have to add up the "Amount" from many rows to get this answer:
  ```
  SELECT SUM(Amount) FROM Recipe_Ingredients
    WHERE MeasureAmountID=1;
  ```
  - ("ounce" corresponds to MeasureAmountID=1)

- Normally, aggregation applies to all the rows, but…
- GROUP BY causes aggregations to occur on subsets of rows, where rows are grouped according to some rule.
  - Each group contains rows having the same value for the grouping expression

  ```
  SELECT SUM(Amount) FROM Recipe_Ingredients
    GROUP BY MeasureAmountID;
  ```
  - Same as above, but list amounts of all ingredients

Subqueries

- Any single value, list of values, or table can be replaced by a subquery

- A **subquery** is a query that appears inside of parentheses.
    - The subquery is computed first and its result is "plugged into" the parent expression.

```
SELECT SUM(Amount) FROM Recipe_Ingredients
  WHERE MeasureAmountID=
    (SELECT MeasureAmountID FROM Measurements
     WHERE MeasurementDescription="Ounce");
```

SalesOrders.sqlite

- List all customers in California (CA). Count them.
- 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? Cheapest 5?
- What is the value of the product inventory on hand? Bike inventory?

List all customers in California (CA). Count them.

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 single most expensive product? Cheapest? 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";
  ```

- What is the full address of customer John Viescas?
  ```
  SELECT CustStreetAddress || "
  " || CustCity || " " || CustState || " " || CustZipCode AS FullAddress FROM
  Customers
  WHERE CustFirstName = "John" AND CustLastName = "Viescas
  ```

- What is the single 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

- 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?

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 mean 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;
```