SQL 13

COMPUTER SCIENCE 61A

August 4, 2015

1 Introduction

Over the semester, we have been using **imperative programming** – a programming style where code is written as a set of instructions for the computer. In this section, we introduce **declarative programming** – code that declares *what* we want, not *how* to compute it. SQL is an example of a declarative programming language. Statements do not describe computations directly, but instead describe the desired result of some computation. It is the role of the query interpreter of the database system to design and perform a computational process to produce such a result.

A table, also called a relation, has a fixed number of named and typed columns. Each row of a table represents a data record and has one value for each column.

For example, we have a table named `records` that stores information about the employees at a small company[1]:

Name	Division	Title	Salary	Supervisor
Ben Bitdiddle	Computer	Wizard	60000	Oliver Warbucks
Alyssa P Hacker	Computer	Programmer	40000	Ben Bitdiddle
Cy D Fect	Computer	Programmer	35000	Ben Bitdiddle
Lem E Tweakit	Computer	Technician	25000	Ben Bitdiddle
Louis Reasoner	Computer	Programmer Trainee	30000	Alyssa P Hacker
Oliver Warbucks	Administration	Big Wheel	150000	Oliver Warbucks
DeWitt Aull	Administration	Secretary	25000	Oliver Warbucks
Eben Scrooge	Accounting	Chief Accountant	75000	Oliver Warbucks
Robert Cratchet	Accounting	Scrivener	18000	Eben Scrooge

[1] Example adapted from Structure and Interpretation of Computer Programs

2 Creating Tables

We can use a `select` statement to create tables. We can define a new table by listing the values in a single row and joining rows together with `union`:

```
sqlite> select "Ben" as first, "Bitdiddle" as last union
   ...> select "Louis",        "Reasoner";
Ben|Bitdiddle
Louis|Reasoner
```

To save a table for use later, use `create table` and the name we want to give the table.

```
sqlite> create table records as
   ...>     select "Ben Bitdiddle" as name, ...
   ...
```

We can also project an existing table using a `from` clause, for example:

```
sqlite> select * from records where name = "Ben Bitdiddle";
Ben Bitdiddle|Computer|Wizard|60000|Oliver Warbucks
```

The following statement lists the names and salaries of each employee under the accounting division, sorted in descending order by their salaries.

```
sqlite> select name, salary from records
   ...>   where division = "Accounting" order by -salary;
Eben Scrooge|75000
Robert Cratchet|18000
```

We can choose which columns to show, filter using a `where` clause, and sort with an `order by` clause using the following syntax:

```
select [columns] from [tables] where [condition] order by [order]
```

2.1 Questions

1. Write a query that outputs the names of employees that Oliver Warbucks directly supervises.

2. Write a query that outputs all information about self-supervising employees.

3. Write a query that outputs the names of all employees with salary greater than 50000 in alphabetical order.

3 Joins

Suppose we have another table `meetings` which records the divisional meetings.

Division	Day	Time
Accounting	Monday	9am
Computer	Wednesday	4pm
Administration	Monday	11am
Administration	Thursday	1pm

Data are combined by joining multiple tables together into one, a fundamental operation in database systems. There are many methods of joining, all closely related, but we will focus on just one method in this class. When tables are joined, the resulting table contains a new row for each combination of rows in the input tables. If two tables are joined and the left table has m rows and the right table has n rows, then the joined table will have mn rows. Joins are expressed in SQL by separating table names by commas in the `from` clause of a `select` statement.

```
sqlite> select name, day from records, meetings;
Ben Bitdiddle|Monday
Ben Bitdiddle|Wednesday
...
Alyssa P Hacker|Monday
...
```

Tables may have overlapping column names, and so we need a method for disambiguating column names by table. A table may also be joined with itself, and so we need a

method for disambiguating tables. To do so, SQL allows us to give aliases to tables within a `from` clause using the keyword `as` and to refer to a column within a particular table using a dot expression. In the example below we find the name and title of Louis Reasoner's supervisor.

```
sqlite> select b.name, b.title from records as a, records as b
   ...>   where a.name = "Louis Reasoner" and
   ...>         a.supervisor = b.name;
Alyssa P Hacker|Programmer
```

3.1 Questions

1. Write a query that creates a table with columns: employee, salary, supervisor and supervisor's salary, containing all supervisors who earn more than twice as much as the employee.

2. Write a query that outputs the names of employees whose supervisor is in a different division.

3. Write a query that outputs the meeting days and times of all employees directly supervised by Oliver Warbucks.

3.2 Extra Questions

1. A middle manager is a person who is both supervising someone and is supervised by someone different. Write a query that outputs the names of all middle managers.

2. What is the output of the query in the previous part? Explain the output you get.

3. Write a query that results in the names of all employees that have a meeting on the same day as their supervisor.

4 Aggregation

So far, we have joined and manipulated individual rows using `select` statements. But we can also perform aggregation operations over multiple rows with the same `select` statements.

We can use the `max`, `min`, `count`, and `sum` functions to retrieve more information from our initial tables.

If we wanted to find the name and salary of the employee who makes the most money, we might say

```
sqlite> select name, max(salary) from records;
Oliver Warbucks|150000
```

Using the special `count(*)` syntax, we can count the number of rows in our table to see the number of employees at the company.

```
sqlite> select count(*) from records;
9
```

These commands can be performed on specific sets of rows in our table by using the `group by [column name]` clause. This clause takes all of the rows that have the same value in `column name` and groups them together.

We can find the miniumum salary earned in each division of the company.

```
sqlite> select division, min(salary) from records
   ...>   group by division;
Computer|25000
Administration|25000
Accounting|18000
```

These groupings can be additionally filtered by the `having` clause, which is very similar to the `where` clause.

To find all titles that are held by more than one person, we say

```
sqlite> select title from records
   ...>   group by title having count(*) > 1;
Programmer
```

4.1 Questions

1. Write a query that outputs each supervisor and the sum of salaries of all of each supervisor's employees.

2. Write a query that outputs all salaries that appear more than once in the employee records.