1 Introduction

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 plan and perform a computational process to produce such a result.

In SQL, data is organized into tables. A table has a fixed number of named columns. A row of the table represents a single 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. Each of the eight rows represents an employee.

<table>
<thead>
<tr>
<th>Name</th>
<th>Division</th>
<th>records</th>
<th>Title</th>
<th>Salary</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben Bitdiddle</td>
<td>Computer</td>
<td>Wizard</td>
<td>60000</td>
<td>Oliver Warbucks</td>
<td></td>
</tr>
<tr>
<td>Alyssa P Hacker</td>
<td>Computer</td>
<td>Programmer</td>
<td>40000</td>
<td>Ben Bitdiddle</td>
<td></td>
</tr>
<tr>
<td>Cy D Fect</td>
<td>Computer</td>
<td>Programmer</td>
<td>35000</td>
<td>Ben Bitdiddle</td>
<td></td>
</tr>
<tr>
<td>Lem E Tweaktit</td>
<td>Computer</td>
<td>Technician</td>
<td>25000</td>
<td>Ben Bitdiddle</td>
<td></td>
</tr>
<tr>
<td>Louis Reasoner</td>
<td>Computer</td>
<td>Programmer Trainee</td>
<td>30000</td>
<td>Alyssa P Hacker</td>
<td></td>
</tr>
<tr>
<td>Oliver Warbucks</td>
<td>Administration</td>
<td>Big Wheel</td>
<td>150000</td>
<td>Oliver Warbucks</td>
<td></td>
</tr>
<tr>
<td>Eben Scrooge</td>
<td>Accounting</td>
<td>Chief Accountant</td>
<td>75000</td>
<td>Oliver Warbucks</td>
<td></td>
</tr>
<tr>
<td>Robert Cratchet</td>
<td>Accounting</td>
<td>Scrivener</td>
<td>18000</td>
<td>Eben Scrooge</td>
<td></td>
</tr>
</tbody>
</table>

For this discussion, you can test out your code at sql.cs61a.org. the records table should already be loaded in.

2 Creating Tables

We can use a SELECT statement to create tables. The following statement creates a table with a single row, with columns named “first” and “last”:

```sql
sqlite> SELECT "Ben" AS first, "Bitdiddle" AS last;
Ben|Bitdiddle
```

Given two tables with the same number of columns, we can combine their rows into a larger table with UNION:

```sql
sqlite> SELECT "Ben" AS first, "Bitdiddle" AS last UNION
       ...> SELECT "Louis", "Reasoner";
Ben|Bitdiddle
Louis|Reasoner
```

---

1 Example adapted from Structure and Interpretation of Computer Programs
To save a table, use CREATE TABLE and a name. Here we’re going to create the table of employees from the previous section and assign it to the name records:

```sql
sqlite> CREATE TABLE records AS
    ...> SELECT "Ben Bitdiddle" AS name, "Computer" AS division,
    ...> "Wizard" AS title, 60000 AS salary,
    ...> "Oliver Warbucks" AS supervisor UNION
    ...> SELECT "Alyssa P Hacker", "Computer",
    ...> "Programmer", 40000, "Ben Bitdiddle" UNION ... ;
```

We can SELECT specific values from an existing table using a FROM clause. This query creates a table with two columns, with a row for each row in the records table:

```sql
sqlite> SELECT name, division FROM records;
<table>
<thead>
<tr>
<th>Name</th>
<th>Division</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alyssa P Hacker</td>
<td>Computer</td>
</tr>
<tr>
<td>Ben Bitdiddle</td>
<td>Computer</td>
</tr>
<tr>
<td>Cy D Fect</td>
<td>Computer</td>
</tr>
<tr>
<td>Eben Scrooge</td>
<td>Accounting</td>
</tr>
<tr>
<td>Lem E Tweakit</td>
<td>Computer</td>
</tr>
<tr>
<td>Louis Reasoner</td>
<td>Computer</td>
</tr>
<tr>
<td>Oliver Warbucks</td>
<td>Administration</td>
</tr>
<tr>
<td>Robert Cratchet</td>
<td>Accounting</td>
</tr>
</tbody>
</table>
```

The special syntax SELECT * will select all columns from a table. It’s an easy way to print the contents of a table.

```sql
sqlite> SELECT * FROM records;
<table>
<thead>
<tr>
<th>Name</th>
<th>Division</th>
<th>Occupation</th>
<th>Salary</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alyssa P Hacker</td>
<td>Computer</td>
<td>Programmer</td>
<td>40000</td>
<td>Ben Bitdiddle</td>
</tr>
<tr>
<td>Ben Bitdiddle</td>
<td>Computer</td>
<td>Wizard</td>
<td>60000</td>
<td>Oliver Warbucks</td>
</tr>
<tr>
<td>Cy D Fect</td>
<td>Computer</td>
<td>Programmer</td>
<td>35000</td>
<td>Ben Bitdiddle</td>
</tr>
<tr>
<td>Eben Scrooge</td>
<td>Accounting</td>
<td>Chief Accountant</td>
<td>75000</td>
<td>Oliver Warbucks</td>
</tr>
<tr>
<td>Lem E Tweakit</td>
<td>Computer</td>
<td>Technician</td>
<td>25000</td>
<td>Ben Bitdiddle</td>
</tr>
<tr>
<td>Louis Reasoner</td>
<td>Computer</td>
<td>Programmer Trainee</td>
<td>30000</td>
<td>Alyssa P Hacker</td>
</tr>
<tr>
<td>Oliver Warbucks</td>
<td>Administration</td>
<td>Big Wheel</td>
<td>150000</td>
<td>Oliver Warbucks</td>
</tr>
<tr>
<td>Robert Cratchet</td>
<td>Accounting</td>
<td>Scrivener</td>
<td>18000</td>
<td>Eben Scrooge</td>
</tr>
</tbody>
</table>
```

We can choose which columns to show in the first part of the SELECT, we can filter out rows using a WHERE clause, and sort the resulting rows with an ORDER BY clause. In general the syntax is:

```
SELECT [columns] FROM [tables]
WHERE [condition] ORDER BY [criteria];
```

For instance, the following statement lists all information about employees with the “Programmer” title.

```sql
sqlite> SELECT * FROM records WHERE title = "Programmer";
<table>
<thead>
<tr>
<th>Name</th>
<th>Division</th>
<th>Occupation</th>
<th>Salary</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alyssa P Hacker</td>
<td>Computer</td>
<td>Programmer</td>
<td>40000</td>
<td>Ben Bitdiddle</td>
</tr>
<tr>
<td>Cy D Fect</td>
<td>Computer</td>
<td>Programmer</td>
<td>35000</td>
<td>Ben Bitdiddle</td>
</tr>
</tbody>
</table>
```

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

---

*Note: This worksheet is a problem bank—most TAs will not cover all the problems in discussion section.*
sqlite> SELECT name, salary FROM records
...> WHERE division = "Accounting" ORDER BY -salary;
Eben Scrooge|75000
Robert Cratchet|18000

Note that all valid SQL statements must be terminated by a semicolon (;). Additionally, you can split up your statement over many lines and add as much whitespace as you want, much like Scheme. But keep in mind that having consistent indentation and line breaking does make your code a lot more readable to others (and your future self)!

Questions

Our tables:

<table>
<thead>
<tr>
<th>records: Name</th>
<th>Division</th>
<th>Title</th>
<th>Salary</th>
<th>Supervisor</th>
</tr>
</thead>
</table>

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

2.2 Write a query that outputs all information about employees that supervise themselves.

2.3 Write a query that outputs the names of all employees with salary greater than 50,000 in alphabetical order.

3 Joins

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

<table>
<thead>
<tr>
<th>Division</th>
<th>Day</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>Monday</td>
<td>9am</td>
</tr>
<tr>
<td>Computer</td>
<td>Wednesday</td>
<td>4pm</td>
</tr>
<tr>
<td>Administration</td>
<td>Monday</td>
<td>11am</td>
</tr>
<tr>
<td>Administration</td>
<td>Wednesday</td>
<td>4pm</td>
</tr>
</tbody>
</table>

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 (the inner join) 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.

Note: This worksheet is a problem bank—most TAs will not cover all the problems in discussion section.
 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

**Questions**

Our tables:

<table>
<thead>
<tr>
<th>records:</th>
<th>Name</th>
<th>Division</th>
<th>Title</th>
<th>Salary</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>meetings:</td>
<td>Division</td>
<td>Day</td>
<td>Time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

2. Write a query that outputs the names of employees whose supervisor is in a different division.
3.3 Write a query that outputs the names of all pairs of employees that have a meeting at the same time. Make sure that if A|B appears in your output, B|A does not appear as well (A|A and B|B should additionally not appear).

3.4 (Extra question) Will the statement above filter out all redundant output in all cases? Why or why not?

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 minimum salary earned in each division of the company.

```
sqlite> SELECT division, MIN(salary) FROM records GROUP BY division;
Computer|25000
Administration|25000
Accounting|18000
```

Note: This worksheet is a problem bank—most TAs will not cover all the problems in discussion section.
These groupings can be additionally filtered by the `HAVING` clause. In contrast to the `WHERE` clause, which filters out rows, the `HAVING` clause filters out entire groups.

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;
```

**Questions**

Our tables:

```markdown
records: Name Division Title Salary Supervisor
```

```markdown
meetings: Division Day Time
```

4.1 Write a query that outputs each supervisor and the sum of salaries of all the employees they supervise.

4.2 Write a query that outputs the days of the week for which fewer than 5 employees have a meeting. You may assume no department has more than one meeting on a given day.

4.3 Write a query that outputs all divisions for which there is more than one employee, and all pairs of employees within that division that have a combined salary less than 100,000.

*Note: This worksheet is a problem bank—most TAs will not cover all the problems in discussion section.*
5 Extra Questions

Use the following table called courses for the questions below:

<table>
<thead>
<tr>
<th>Professor</th>
<th>Course</th>
<th>Semester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dan Garcia</td>
<td>CS 61C</td>
<td>Sp19</td>
</tr>
<tr>
<td>John DeNero</td>
<td>CS 61A</td>
<td>Fa18</td>
</tr>
<tr>
<td>Dan Garcia</td>
<td>CS 10</td>
<td>Fa18</td>
</tr>
<tr>
<td>Josh Hug</td>
<td>CS 61B</td>
<td>Sp18</td>
</tr>
<tr>
<td>John DeNero</td>
<td>CS 61A</td>
<td>Sp18</td>
</tr>
<tr>
<td>John DeNero</td>
<td>CS 61A</td>
<td>Fa17</td>
</tr>
<tr>
<td>Paul Hilfinger</td>
<td>CS 61A</td>
<td>Fa17</td>
</tr>
<tr>
<td>Paul Hilfinger</td>
<td>CS 61A</td>
<td>Sp17</td>
</tr>
<tr>
<td>John DeNero</td>
<td>Data 8</td>
<td>Sp17</td>
</tr>
<tr>
<td>Josh Hug</td>
<td>CS 61B</td>
<td>Sp17</td>
</tr>
<tr>
<td>Satish Rao</td>
<td>CS 70</td>
<td>Sp17</td>
</tr>
<tr>
<td>Nicholas Weaver</td>
<td>CS 61C</td>
<td>Sp17</td>
</tr>
<tr>
<td>Gerald Friedland</td>
<td>CS 61C</td>
<td>Sp17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1 Create a table called num_taught that contains three columns: professor, the course they taught, and the number of times they taught each course.

**Hint:** For this problem, it may help to GROUP BY multiple columns. Multiple columns and full expressions can appear in the group by clause, and groups will be formed for every unique combination of values that result.

5.2 Write a query that outputs two professors and a course if they have taught that course the same number of times. You may use the num_taught table you created in the previous question.

5.3 Write a query that outputs two professors if they co-taught (taught the same course at the same time) the same course more than once.

*Note: This worksheet is a problem bank—most TAs will not cover all the problems in discussion section.*
1. **Notation notation notation**

One annoying thing about Scheme is that it can only understand arithmetic operations that are written in prefix notation. That is, if I want to evaluate an expression, the arithmetic operator must come first, which really goes against everything you were taught as a child. Let’s leverage our interpreter skills to define a Scheme procedure that accepts arithmetic operations with infix notation, which places operators between operands as you’re used to. You only need to support the addition and multiplication operators * and +, but you need to support order of operations. Define the interpret procedure so that it passes the test cases below.

```scheme
(define (interpret expr)
  (cond
    (? expr (interpret expr))
    ((null? (cdr expr)) (if (caar expr))
     (interpret expr))
    ((null? (cadr expr)) (interpret expr))
    ((null? (cddr expr)) (interpret expr))
    (else (+ (interpret (caar expr)) (interpret (cadr expr))))))

(scm> (interpret '(1))
1
(scm> (interpret '(1 + 2))
3
(scm> (interpret '(1 * 2))
2
; Order of operations apply
(scm> (interpret '(3 + 2 * 5 + 4))
17
(scm> (interpret '(5 * 3 + 2 * 4 * 9))
53
; Parentheses should be handled properly
(scm> (interpret '(3 * (2 * 4)))
24
(scm> (interpret '(3 + (2 + 4)))
9
(scm> (interpret '((3 + 2) + 4))
9
; Parentheses are prioritized higher than order of operations
(scm> (interpret '(1 + 2 * (3 + 4)))
15
(scm> (interpret '(1 + 2 * (3 + 4 * (5 + 6))))
95

; Some helper procedures (optional)
(define (caar x) (car (car x)))
(define (cadr x) (car (cdr x)))
(define (cddr x) (cdr (cdr x)))

(define (interpret expr)
  (cond
    (? expr (interpret expr))
    ((null? (cdr expr)) (if (caar expr))
     (interpret expr))
    ((null? (cadr expr)) (interpret expr))
    ((null? (cddr expr)) (interpret expr))
    (else (+ (interpret (caar expr)) (interpret (cadr expr))))))
```

2. Don’t forget to check your quiz answers, which are on the last page of discussion solutions that are posted at the end of each week.