Nested Subqueries

Warmup

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What does this do?

```
sqlite> select * from
    (select name, ID from instructor order by name);
```

Warmup

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What does this do?

```
sqlite> select * from
    (select name, ID from instructor order by name);
```

```
sqlite> select * from (select name, ID from instructor order by name);
Brandt,83821
Califieri,58583
Crick,76766
Einstein,22222
"El Said",32343
Gold,33456
…
```

How to express this in SQL?

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the most any department pays in total instructor salaries?

Trick!

```
sqlite> select dept_name, sum(salary) as total_salary
        from instructor
        group by dept_name
        order by total_salary desc
        limit 1;
```

How to express this in SQL?

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the most any department pays in total instructor salaries?

Would the problem be easier if we had a table like this?

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

Idea: query in two steps

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

get sum of salaries for each dept.

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

get greatest total salary

232000

Exercise: what is the first query?

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

get sum of salaries for each dept.

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

```
sqlite> select dept_name, sum(salary)
          from instructor
         group by dept_name;
Biology,72000
Comp. Sci.,232000
Elec. Eng.,80000
Finance,170000
…
```

Exercise: what is the second query?

T

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

get greatest total salary

232000

```
sqlite> select max(total_salary) from T;
232000
```

Solution

What is the most any department pays in total instructor salaries?

```
sqlite> select max(total_salary)
          from (select dept_name, sum(salary) as total_salary
                  from instructor
                 group by dept_name);
232000
```

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

Alternative solution, using 'with'

What is the most any department pays in total instructor salaries?

```
sqlite> with T(dept_name, total_salary) as
  (select dept_name, sum(salary) as total_salary
     from instructor
     group by dept_name)
select max(total_salary) from T;
232000
```

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

How does this work?

1. every query produces a **table**

2. so a query can be used **wherever a table is accepted**

```
sqlite> select max(total_salary)
          from (select dept_name, sum(salary) as total_salary
                  from instructor
                 group by dept_name);
232000
```

Another example

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000

What are the average instructors' salaries in departments where the average salary is > $42,000?

```
sqlite> select dept_name, round(avg_salary)
          from (select dept_name, avg(salary) as avg_salary
                  from instructor
                group by dept_name)
          where avg_salary > 42000;
Biology,72000.0
Comp. Sci.,77333.0
Elec. Eng.,80000.0
Finance,85000.0
History,61000.0
Physics,91000.0
```

Where are tables expected in SQL?

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What are the instructors' names, aside from 'Mozart' and 'Einstein'?

a table can be used here

```
sqlite> sqlite> select distinct(name) from instructor where name not in ('Mozart', 'Einstein');
Srinivasan
Wu
El Said
Gold
Katz
Califieri
…
```

Nested queries with ‘in’, ‘not in’

section

course_id	sec_id	semester	year	building	room_number	time_slot_id
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	B
CS-347	1	Fall	2009	Taylor	3128	A

What courses are taught in Fall 2009 but not Spring 2010?

```
sqlite> select distinct course_id
          from section
          where semester = 'Fall' and year = 2009 and
                course_id not in (select course_id
                                  from section
                                  where semester = 'Spring' and year = 2010);
CS-347
PHY-101
```

Use Case 1 – using natural join

T

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

How to write this query?

Natural Join enforces equality on total_salary column

get dept_name and greatest total salary

dept_name	total_salary
Comp. Sci.	232000

total_salary

232000

select dept_name, max(total_salary) from T

```
sqlite> select dept_name, total_salary
          from T natural join (select max(total_salary) as total_salary
                               from T);
“Comp. Sci.”,232000
```

Use Case 2 – using IN

T

dept_name	total_salary
Biology	72000
Comp. Sci.	232000
Elec. Eng.	80000
Finance	170000
History	122000
Music	40000
Physics	182000

get dept_name and greatest total salary

dept_name	total_salary
Comp. Sci.	232000

```
sqlite> select dept_name, total_salary
        from T
        where total_salary IN (select max(total_salary) from T);
“Comp. Sci.”,232000
```

which wards contain both male and female patients?

```
sqlite> select ward from patient
        where sex="M"
        and ward IN (select ward from patient where sex="F");
```