

Activity 11: Relational Operations

Model 1 Select and Project

In relational databases, *data* is organized as tables. We use *SELECT* to work with rows and *PROJECT* to work with columns. The names of the columns are called the *schema*.

snacks

name	owner	calories	price
Snickers	Mars	215	1.25
Peanut M&M's	Mars	250	1.00
Twix	Mars	286	1.25
Reeses Pieces	Hershey	234	1.00
Butterfinger	Nestle	275	1.25
Milk Duds	Hershey	218	1.50
Milky Way	Mars	264	1.25
Baby Ruth	Ferrero	275	1.50
Doritos	Frito-Lay	140	0.75
Cheetos	Frito-Lay	160	0.75

Examples:

SELECT_{price ≥ 1.50} (snacks)

name	owner	calories	price
Milk Duds	Hershey	218	1.50
Baby Ruth	Ferrero	275	1.50

SELECT_{price < 0} (snacks)

name	owner	calories	price
------	-------	----------	-------

PROJECT_{name} (SELECT_{price = 0.75} (snacks))

name
Doritos
Cheetos

PROJECT_{owner, calories} (snacks)

owner	calories
Mars	215
Mars	250
Mars	286
Hershey	234
Nestle	275
Hershey	218
Mars	264
Ferrero	275
Frito-Lay	140
Frito-Lay	160

Questions (10 min)

Start time: _____

1. How many rows and columns are in:

- a) the original snacks table? 10 rows, 4 cols
- b) selecting price ≥ 1.50 ? 2 rows, 4 cols
- c) selecting price < 0 ? 0 rows, 4 cols
- d) projecting owner and calories? 10 rows, 2 cols

2. Which operation (SELECT or PROJECT) affects the schema? Justify your answer.

The PROJECT operation reduces the table to fewer columns.

3. The bottom-left example in Model 1 uses both SELECT and PROJECT. Describe the data source of each operation (the part in parentheses):

- a) SELECT ... (which data?) the original snacks table
- b) PROJECT ... (which data?) the result of the SELECT

4. In addition to the data source, what other information (the part in subscript) is required for:

- a) a SELECT operation? a condition for which rows to select
- b) a PROJECT operation? a list of columns to project (retain)

5. Explain what is wrong with this example: $\text{SELECT}_{\text{price} = 0.75} (\text{PROJECT}_{\text{name}} (\text{snacks}))$

The PROJECT operation removes all columns except for name, so you can't SELECT based on price anymore.

6. Write the following *queries* using SELECT and/or PROJECT:

- a) List the name and price of all snacks. $\text{PROJECT}_{\text{name}, \text{price}} (\text{snacks})$
- b) Find snacks with less than 200 calories. $\text{SELECT}_{\text{calories} < 200} (\text{snacks})$

c) Which company makes Twix? `PROJECT owner (SELECT name = 'Twix' (snacks))`

Model 2 Product and Join

Mathematically speaking, we combine tables by “multiplying” them. Every row in the right table is appended to every row in the left table:

A	B	A × B	
let	num	let	num
A	1	A	1
B	2	A	2
C		B	1
		B	2
		C	1
		C	2

In relational databases, a *join* operation is a product followed by a condition. The condition is used to specify which of the combined rows should be part of the result.

course			teach		professor		
cid	dept	num	cid	pid	pid	dept	name
13466	CS	101	13466	2774	2774	CS	Mayfield
13468	CS	149	13468	2774	9036	CS	Stewart
56482	MATH	231	13466	9036	1158	MATH	Taalman
			13468	9036	5241	SCOM	Hazard

course × teach				
cid	dept	num	cid	pid
13466	CS	101	13466	2774
13466	CS	101	13468	2774
13466	CS	101	13466	9036
13466	CS	101	13468	9036
13468	CS	149	13466	2774
13468	CS	149	13468	2774
13468	CS	149	13466	9036
13468	CS	149	13468	9036
56482	MATH	231	13466	2774
56482	MATH	231	13468	2774
56482	MATH	231	13466	9036
56482	MATH	231	13468	9036

JOIN_{course.cid = teach.cid} (course, teach)

cid	dept	num	cid	pid
13466	CS	101	13466	2774
13466	CS	101	13466	9036
13468	CS	149	13468	2774
13468	CS	149	13468	9036

Questions (10 min)

Start time: _____

7. How many rows and columns are in:

a) the course table? 3 rows, 3 cols

b) the teach table? 4 rows, 2 cols

c) course \times teach? 12 rows, 5 cols

8. Consider a table with i rows and j columns, and another table with k rows and l columns.

a) how many rows will be in the product? $i * k$

b) how many columns will be in the product? $j + l$

9. Discuss how the results of “course \times teach” are different from the JOIN operation. Then in Model 2, draw an arrow from each result in the JOIN to the corresponding row in the product.

There should be four arrows: row 1 \leftarrow row 1, row 3 \leftarrow row 2, row 6 \leftarrow row 3, row 8 \leftarrow row 4.

10. What is the result of $\text{JOIN}_{\text{teach.pid} = \text{professor.pid}}$ (teach, professor)? Don't forget to include the column names.

cid	pid	pid	dept	name
13466	2774	2774	CS	Mayfield
13468	2774	2774	CS	Mayfield
13466	9036	9036	CS	Stewart
13468	9036	9036	CS	Stewart

11. Describe what relational operations you would have to use to find the names of all professors who teach CS 101. (The results should have 2 rows and 1 column.)

You would first have to join the course, teach, and professor tables using the conditions “cid = cid” and “pid = pid”. Then you would need to apply a selection on the rows with “CS 101”. Finally, you would need to project the “name” column.