Exercise 10A: 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 $_{\text{price} \geq 1.50}$ (snacks)

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

SELECT $_{\text{price} < 0}$ (snacks)

name	owner	calories	price

PROJECT $_{\text{owner, calories}}$ (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

PROJECT $_{\text{name}}$ (SELECT $_{\text{price} = 0.75}$ (snacks))

name
Doritos
Cheetos

1. How many rows and columns are in:

a) the original snacks table?

b) selecting price ≥ 1.50?

c) selecting price < 0?

d) projecting owner and calories?

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

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 $_{\ldots}$ (which data?)

b) PROJECT $_{\ldots}$ (which data?)

4. In addition to the data source, what other information (the part in $_{\text{subscript}}$) is required for:

a) a SELECT operation?

b) a PROJECT operation?

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

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

a) List the name and price of all snacks.

b) Find snacks with less than 200 calories.

c) Which company makes Twix?

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

let
A
B
C

B

num
1
2

A × B

let	num
A	1
A	2
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

cid	dept	num
13466	CS	101
13468	CS	149
56482	MATH	231

teach

cid	pid
13466	2774
13468	2774
13466	9036
13468	9036

professor

pid	dept	name
2774	CS	Simmons
9036	CS	Normand
1158	MATH	Taalman
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 $_{\text{course.cid = teach.cid}}$ (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

7. How many rows and columns are in:

a) the course table?

b) the teach table?

c) course $\times$ teach?

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?

b) how many columns will be in the product?

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.

10. What is the result of JOIN $_{\text{teach.pid = professor.pid}}$ (teach, professor)? Don't forget to include the column names. It is not necessary to draw with horizontal and vertical lines, instead line up column data under column names.

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.)