



Oregon State  
University

COLLEGE OF ENGINEERING

School of Electrical Engineering  
and Computer Science

# CS 340 Intro to DATABASES

Relational Algebra Operations  
Part 3

Some material from Harrington, Jan L..

*Relational Database Design and Implementation: Clearly Explained*, 2016.

Some of this material comes from the course textbook:

Chapter 6 on Relational Algebra

which you can read for free from the OSU Library.

Harrington, Jan L.. *Relational Database Design and Implementation : Clearly Explained*, Elsevier Science & Technology, 2016. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/osu/detail.action?docID=4509772>.

There is also a companion **Exploration Relax RA Simulator activity** for students to complete in Canvas

## Relational Algebra Set Operators

- Relational algebra
  - Defines theoretical way of manipulating table contents using relational operators
  - Do the most common things that we need to do with relations in a database
  - Use of relational algebra operators on existing relations produces new relations:
    - SELECT
    - PROJECT
    - JOIN
    - INTERSECT
    - UNION
    - DIFFERENCE
    - PRODUCT
    - DIVIDE

2

Database systems are software programs that have to follow rules. The rules that define how a database will behave are a narrative model with mathematical symbols.

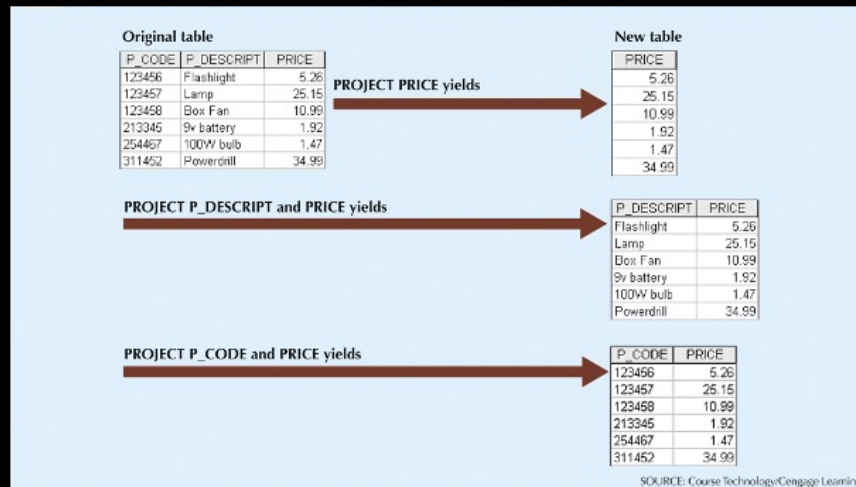
The relational algebra:

Defines theoretical way of manipulating table contents using relational operators. Use of relational algebra operators on existing relations produces new relations.

In short, these operators are designed to do the most common things that we need to do with relations in a database.

SELECT  
UNION  
PROJECT  
DIFFERENCE  
JOIN  
PRODUCT  
INTERSECT  
DIVIDE

# PROJECT



3

Project returns the attributes. Project does not limit the number of rows that are returned. We use the

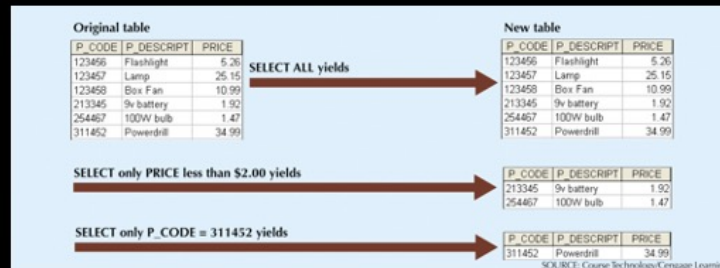
A Project is denoted by the letter pi

$\pi$  p\_code, price (parts)

A SELECT can be added to project to specify the columns and conditions desired.

$\pi$  p\_code, price ( $\sigma < 2000$  (parts))

# SELECT



SELECT is a unary operator, it works on one table. It is used to obtain a subset of the tuples (attributes) of a relation that satisfy a *select condition*

Select give us a way to apply comparison based filtering.

4

SELECT is used to obtain a subset of the tuples (attributes) of a relation that satisfy a *select condition*

Select give us a way to filter on a condition. We use the sigma operator for RA expressions

$\sigma_{p\_code < 2000}$  (parts)

OR

SELECT all parts that are less than \$2000

## Union & Intersection

P_CODE	P_DESCRIPTION	PRICE	UNION			yields		
123456	Flashlight	5.26	P_CODE	P_DESCRIPTION	PRICE	123456	Flashlight	5.26
123457	Lamp	25.15	345678	Microwave	160.00	123457	Lamp	25.15
123458	Box Fan	10.99	345679	Dishwasher	500.00	123458	Box Fan	10.99
213345	9v battery	1.92	123458	Box Fan	10.99	213345	9v battery	1.92
254467	100W bulb	1.47				254467	100W bulb	1.47
311452	Powerdrill	34.99				311452	Powerdrill	34.99
						345678	Microwave	160
						345679	Dishwasher	500

SOURCE: Course Technology/Cengage Learning

STU_FNAME	STU_LNAME	INTERSECT		yields	
George	Jones	EMP_FNAME	EMP_LNAME	STU_FNAME	STU_LNAME
Jane	Smith	Franklin	Lopez	Franklin	Johnson
Peter	Robinson	William	Turner		
Franklin	Johnson	Franklin	Johnson		
Martin	Lopez	Susan	Rogers		

SOURCE: Course Technology/Cengage Learning

5

Union combines all rows from two tables and excludes duplicates. In order for two relations to be union-compatible or *unifiable*, both must have the same number of attributes (columns) and corresponding attributes (columns) must have the same set of allowable values for that attribute (e.g. integer or character).

Intersect yields only the rows that appear in both tables, and the tables also have to be unifiable.

The U symbol is used for union and the  $\cap$  symbol for intersection

## DIFFERENCE & PRODUCT

STU_FNAME	STU_LNAME	DIFFERENCE	EMP_FNAME	EMP_LNAME	yields	STU_FNAME	STU_LNAME
George	Jones		Franklin	Lopez		George	Jones
Jane	Smith		William	Turner		Jane	Smith
Peter	Robinson		Franklin	Johnson		Peter	Robinson
Franklin	Johnson		Susan	Rogers		Martin	Lopez
Martin	Lopez						

SOURCE: Course Technology/Cengage Learning

P_CODE	P_DESCRIPTION	PRICE	PRODUCT	STORE	aisle	shelf	yields	P_CODE	P_DESCRIPTION	PRICE	STORE	aisle	shelf
123456	Flashlight	5.26		23	W	5		123456	Flashlight	5.26	23	W	5
123457	Lamp	25.15		24	K	9		123456	Flashlight	5.26	24	K	9
123458	Box Fan	10.99		25	Z	6		123456	Flashlight	5.26	25	Z	6
213345	9v battery	1.92						123457	Lamp	25.15	23	W	5
254467	100W bulb	1.47						123457	Lamp	25.15	24	K	9
311452	Powerdrill	34.99						123457	Lamp	25.15	25	Z	6
								123458	Box Fan	10.99	23	W	5
								123458	Box Fan	10.99	24	K	9
								123458	Box Fan	10.99	25	Z	6
								213345	9v battery	1.92	23	W	5
								213345	9v battery	1.92	24	K	9
								213345	9v battery	1.92	25	Z	6
								311452	Powerdrill	34.99	23	W	5
								311452	Powerdrill	34.99	24	K	9
								311452	Powerdrill	34.99	25	Z	6
								254467	100W bulb	1.47	23	W	5
								254467	100W bulb	1.47	24	K	9
								254467	100W bulb	1.47	25	Z	6

SOURCE: Course Technology/Cengage Learning

6

Difference yields all rows in one table that are not found in the other table, or to put it another way, it subtracts one table from the other.

The Cartesian Product yields all possible pairs of rows from two tables. So if one table has 3 rows and the other table has 6 rows, the resulting set is  $3 \times 6 = 18$  rows.

The – symbol is used for subtraction and the  $\times$  for product

## JOINS

- JOIN retrieves more than one table at a time
- The ⋈ symbol is used for join
- Natural join
  - Links tables by selecting rows with common values in common attributes (join columns)
- Inner join
  - Only returns matched records from the tables that are being joined
- Outer join
  - Matched pairs are retained, and any unmatched values in other table are left null

7

JOIN is performed when data are retrieved from more than one table at a time. The ⋈ symbol is used for join. There are three types of Join:

### Natural join

Links tables by selecting rows with common values in common attributes (join columns). Natural join no need to specify condition. If there is no condition specifies then it returns the rows based on the common column

```
π movies.name, movies_genres.genre, movies.rank ( movies ⋈ movies_genres )
```

Here is the corresponding SQL

```
select movies.name, movies_genres.genre, movies.rank  
from movies natural join movies_genres
```

### Inner join

Only returns matched records from the tables that are being joined. Inner Join joins two table on the basis of the column which is explicitly specified. In practice inner Join is used most times we need to write a Join query.

```
π movies.name, movies_genres.genre, movies.rank ( movies ⋈ movies.id =  
movies_genres.movie_id movies_genres )
```

Here is the corresponding SQL

```
select movies.name, movies_genres.genre, movies.rank
```

from movies inner join movies\_genres on movies.id = movies\_genres.movie\_id

#### Outer join

Matched pairs are retained, and any unmatched values in other table are left null

$\pi$  movies.name, movies\_genres.genre ( movies  $\bowtie$  movies.id =  
movies\_genres.movie\_id movies\_genres )

Heres the corresponding SQL:

```
SELECT movies.name, movies_genres.genre FROM movies FULL OUTER JOIN  
movies_genres ON movies.id=movies_genres.movie_id;
```



## Relational Algebra Summary

- The rules of RA help databases be compatible with one another.
  - RA is the basis for the SQL language.
- You will get to practice making RA expressions
  - But for the more complex queries we will use SQL
- Relax is a web RA calculator that you can use to practice RA expressions

8

To summarize, we spent time talking about relational algebra because it is the formal set of rules that define how a database should work. The fact that all relational databases follow these rules a tremendous benefit. It simplifies transferring data between different software vendor versions of DBs.

The rules of RA help databases be compatible with one another.  
RA is the formal definition for the SQL language.

We will use Relax to get some practice formulating simple RA expressions.  
<https://dbis-uibk.github.io/relax/>

And more complex queries will be covered in SQL.