COS221
L11 - Relational Algebra 1
(Chapter 6 in Edition 6 and Chapter 8 in Edition 7)

Linda Marshall

23 March 2023

Why Relational Algebra (RA)?

- Developed before the SQL language.
- Represents the basic set of operations for the relational model.
- Basic retrieval operations are expressed using relational algebra expressions which results in a new relation being created. A relational algebra expression is defined by a sequence of relational algebra operations.

Why Relational Algebra (RA)?

So why?

- provides a formal mathematical foundation for the relational model
- provides a mechanism to optimise queries which is why we use a relational DBMS (RDBMS)
- concepts are defined within SQL

Why Relational Algebra (RA)?

Relation algebra operations fall into two groups:

- set operations from mathematical set theory
  - Relations are views as a set of tuples.
  - Therefore set operations such as UNION, INTERSECTION, SET DIFFERENCE and CARTESIAN/CROSS PRODUCT can be applied
- relational database specific operations
  - Relational database specific operations include operations such as SELECT, PROJECT and JOIN.
  - SELECT and PROJECT are unary operations while JOIN is a binary operation.

Aggregate operations are defined for instances where set or specific relational algebra operations are not enough.

The SELECT operation

- The SELECT (σ) operation is used to choose a subset of tuples which satisfy a selection condition.
- Syntax of the SELECT operation is given by:

  $\sigma_{<\text{selection\_condition}>}(R)$, where R is a relation

- The resulting relation will have the same attributes as R.
- For example:
  $\sigma_{Dno=4}(EMPLOYEE)$
  $\sigma_{Salary>30000}(EMPLOYEE)$

The SELECT operation

The selection condition is a boolean expression and is made up of a number of clauses:
$<$ attribute_name $><$ comparison_op $><$ constant_value $>$
$<$ attribute_name $><$ comparison_op $><$ attribute_value $>$

where:

$<$ attribute_name $>$ is the name of an attribute in R
$<$ constant_name $>$ is a constant attribute from the attribute domain
$<$ comparison_op $>$ is normally one of $\{=, <, \leq, >\geq, \neq\}$. All operators can be applied to *ordered values*. If the values are unordered, such as chars or strings, then only the first and last operator given above can be applied.

The SELECT operation

Clauses can be connected by the Boolean operators AND, OR and NOT. They follow their normal interpretation:

- AND is true is both conditions are true, otherwise it is false
- OR is true of one of the conditions is true, otherwise it is false
- NOT is true if its condition is false, otherwise it is true

The SELECT operation

- The *selection_condition* is applied independently to each tuple (t) in R. By substituting each attribute (A_i) in the selection condition with its value in the tuple $t[A_i]$, a TRUE result will result in tuple t to be placed in the resultant relation.
- The *degree*, number of attributes, of the resultant relation is the same as the original relation R.
- The number of tuples in the resultant relation is less than or equal to the number of tuples in R.
- Therefore, for any condition c, $|\sigma_c(R)| \leq |R|$

The SELECT operation

- SELECT is *commutative*,
  $\sigma_{<cond1>}(\sigma_{<cond2>}(R)) = \sigma_{<cond2>}(\sigma_{<cond1>}(R))$
- SELECT *cascades*,
  $\sigma_{<cond1>}(\sigma_{<cond2>}(R)) = \sigma_{<cond1> AND <cond2>}(R)$

The SELECT operation

For example:

$\sigma_{Dno=4}(\sigma_{Salary>25000}(EMPLOYEE)) =$

$\sigma_{Dno=4 \text{ AND } Salary>25000}(EMPLOYEE))$

which translates to the SQL query:

```
SELECT *
FROM EMPLOYEE
WHERE Dno = 4 AND Salary > 25000;
```

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

The PROJECT operation

- Where the SELECT operation selected 'rows', the PROJECT operation selects 'columns'. That is, the PROJECT (π) operation partitions the relation vertically.
- The PROJECT operation is used to select certain attributes and discard the rest.
- PROJECT has the following syntax:
  $\pi_{<attribute\_list>}(R)$

The PROJECT operation

- The order of the attributes in the resultant relation are the same as given in the $< attribute\_list >$.
- The degree of the relation of the same as the number of attributes in the $< attribute\_list >$.
- If the $< attribute\_list >$ does not include any key attributes, the duplicate tuples are removed - referred to as *duplicate elimination*
- Duplicate elimination is as a result of a relation being represented as a set of tuples and a *mathematical set not allowing duplicates*.

The PROJECT operation

For example:

$\pi_{Sex,Salary}(EMPLOYEE)$

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

(c)

Sex	Salary
M	30000
M	40000
F	25000
F	43000
M	38000
M	25000
M	55000

The PROJECT operation

- If a relation had been defined as a *multiset/bag*, then duplicates would have been allowed.
- RA considers a relation as a *set*, whereas SQL allows for the distinction between a set and a multiset.
  - As a set:

    ```
    SELECT DISTINCT Sex, Salary
    FROM EMPLOYEE;
    ```

  - As a multiset or bag:

    ```
    SELECT Sex, Salary
    FROM EMPLOYEE;
    ```

Sequences of operations and the RENAME operation

- Operations may be applied as a nested sequence or one at a time.
- If operations are applied one at a time, the resultant relation must be stored in an intermediate relation.
- This is written in relational algebra as follows:
  $DEPS\_EMPS \leftarrow \sigma_{Dno=5}(EMPLOYEE)$
  $RESULT \leftarrow \pi_{Fname,Lname,Salary}(DEPS\_EMPS)$
  This is equivalent to:
  $\pi_{Fname,Lname,Salary}(\sigma_{Dno=5}(EMPLOYEE))$

Sequences of operations and the RENAME operation

- Sometimes it is necessary to *rename* the attributes of the relation. Renaming is written as follows:
  $TEMP \leftarrow \sigma_{Dno=5}(EMPLOYEE)$
  $R(First\_name, Last\_name, Salary) \leftarrow$
  $\pi_{Fname,Lname,Salary}\, TEMP$

(b)
TEMP

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston,TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston,TX	M	40000	888665555	5
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble,TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5

R

First_name	Last_name	Salary
John	Smith	30000
Franklin	Wong	40000
Ramesh	Narayan	38000
Joyce	English	25000

Sequences of operations and the RENAME operation

A formal RENAME (ρ) operation definition for a relation of degree n is given by:

- for renaming the relation and attributes
  $\rho_{S(B_1,B_2,\ldots,B_n)}(R)$
- for renaming the relation only
  $\rho_S(R)$
- for renaming the attributes only
  $\rho_{B_1,B_2,\ldots,B_n}(R)$

In SQL

```
SELECT E.Fname AS First_name
       E.Lname AS Last_name
       E.Salary AS Salary
FROM EMPLOYEE AS E
WHERE DNO = 5;
```

Relational algebra operations from set theory

- UNION, INTERSECTION and MINUS are standard operations from set theory.
- All these operations are binary operations and therefore work on two relations (sets of tuples).
- These operations are applied to sets that are union/type compatible. That is, for two relations $R(A_1, A_2, ..., A_n)$ and $S(B_1, B_2, ..., B_n)$, the degrees of the relations must be the same (both n) and the $dom(A_i) = dom(B_i)$ for $1 \leq i \leq n$.

Relational algebra operations from set theory

For two union compatible relations R and S, the operations have the following meaning:

- UNION ($\cup$) - the union of R and S is the relation of all tuples in (R or S) or (R and S). Duplicates are eliminated.
- INTERSECTION ($\cap$) - the intersection of R and S is the relation of all tuples in both R and S.
- SET DIFFERENCE (-) (or MINUS) - the difference of R and S ($R - S$) is all tuples in R but not in S.

Relational algebra operations from set theory

Figure 6.4

The set operations UNION, INTERSECTION, and MINUS. (a) Two union-compatible relations. (b) STUDENT ∪ INSTRUCTOR. (c) STUDENT ∩ INSTRUCTOR. (d) STUDENT – INSTRUCTOR. (e) INSTRUCTOR – STUDENT.

(a) **STUDENT**

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

(b)

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert
John	Smith
Ricardo	Browne
Francis	Johnson

(c)

Fn	Ln
Susan	Yao
Ramesh	Shah

(d)

Fn	Ln
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

(e)

Fname	Lname
John	Smith
Ricardo	Browne
Francis	Johnson