Formal Query Languages: Relational Algebra

- Set Theory Operations
- Specific Relational Operations
- Write Queries in Relational Algebra

1

Steps in Processing a Query

SQL statement
↓
Scan, Parse, Validate
↓
Intermediate form of query
↓
Query Optimizer
↓
Execution plan
↓
Query Code Generator
↓
Code to execute query
↓
Runtime DB Processor
↓
Results of running query

2

Relational Algebra − + × ÷

- Operations on entire relations
  - Operands are (constant or variable) relations
  - Result is a relation
- Set theory operations:
  - Union, Intersection, Difference and Cartesian Product (product for short)
- Specific relational operations:
  - Selection, Projection, Join and Division
- Complete set of relational algebra operations:
  - Select, project, product, union and difference
- SQL is <u>based</u> on concepts from relational algebra

3

Selection σ

- Unary operator <u>Select</u>, **σ**:

$$\sigma_{\text{selection-condition}}(r)$$

- E.g., $\sigma_{Name='John' \vee Name='Susan'}(STUDENT)$
  - result = $\{t \mid t \in r \text{ and } (t[Name] = 'John' \text{ or } t[Name] = 'Susan')\}$
- *Selection condition* any logical expression on attributes of **r** involving any applicable comparison operator $\{=, <, \leq, >, \geq, \neq\}$

4

Example of Selection σ

- $\sigma_{\text{Name='Bob'} \vee \text{Major = 'Math'}}(S)$ = ?
- $\sigma_{\text{Name='Bob'} \wedge \text{Major = 'Math'}}(S)$ = ?
- How can I get a copy of S?
- How can I get an empty copy of S?

Relation **S**

SID	Name	Major
1	Bob	CS
3	Ann	CoE
4	Bob	Math

5

Example of Selection σ

- How can I get a copy of S?
- $\sigma_{\text{true}}(S)$ =

RSLT:

SID	Name	Major
1	Bob	CS
3	Ann	CoE
4	Bob	Math

- How can I get an empty copy of S?
- $\sigma_{\text{false}}(S)$ =

RSLT:

SID	Name	Major

6

Projection π

- Unary operator <u>Project</u>, π:

  $\pi_{\textbf{attribute-list}}(\textbf{r})$

  - Attribute-list $\subseteq$ R
- E.g., $\pi_{\text{Name, Major}}$(STUDENT)
  - result = {t | t$\in$r and t[Name, Major]}
- What about $\pi_{\text{SID, Major}}(S)$ = ?
- What about $\pi_{\text{Name, Major}}(S)$ = ?

Relation **S**

SID	Name	Major
1	Bob	CS
3	Ann	CoE
4	Bob	Math

7

Example of Projection π

Relation **S**

SID	Name	Major
1	Bob	CS
3	Ann	CoE
4	Bob	Math

5	Bob	CS

- $\pi_{\text{SID, Major}}(S)$ = ?

RSLT:

SID	Major
1	CS
3	CoE
4	Math

- $\pi_{\text{Name, Major}}(S)$ = ?

RSLT:

Name	Major
Bob	CS
Ann	CoE
Bob	Math

8

Relational Algebra Expressions

- Query: *List the QPA of all students (SID) in CSD whose QPA is greater than 3.5*
- STUDENT (<u>SID</u>, FName, SName, Dept, Major, QPA)
- **Nesting** the operations

  $\pi_{\text{SID, QPA}} (\sigma_{\text{Dept = 'CSD'} \wedge \text{QPA>3.5}} (\text{STUDENT}))$
- **Sequence** of operations

  $\text{HS} \leftarrow \sigma_{\text{Dept = 'CSD'} \wedge \text{QPA>3.5}} (\text{STUDENT})$

  $\text{RESULT} \leftarrow \pi_{\text{SID, QPA}} (\text{HS})$
- **Query tree**
  - *leaf* nodes are relations and *internal* nodes are operations

Renaming Operator

- Renaming attributes of the result

  $\text{RSLT(StudentID, GPA)} \leftarrow \pi_{\text{SID, QPA}} (\text{HS})$
- Change the name of Attributes (in general):

  $\rho(a1,a2,a3,..an)(r)$
- Example:

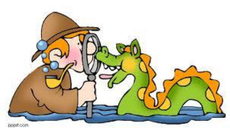
  $\rho(\text{StudentID, GPA}) ($

  $\pi_{\text{SID, QPA}} (\sigma_{\text{Dept = 'CSD'} \wedge \text{QPA>3.5}} (\text{STUDENT}))$

Properties of σ and π

- $\sigma_{\text{cond1}} (\sigma_{\text{cond2}} (R)) = \sigma_{\text{cond2}} (\sigma_{\text{cond1}} (R))$
- $\sigma_{\text{cond1}} (\sigma_{\text{cond2}} (R)) = \sigma_{\text{cond2} \wedge \text{cond1}} (R)$

  $= \sigma_{\text{cond1} \wedge \text{cond2}} (R)$
- $\pi_{\text{list1}} (\pi_{\text{list2}} (R)) = \pi_{\text{list1}} (R)$ When?



Efficient / Optimized Queries

- Reduce cost of computing (a.k.a, *time-complexity*)
  - Short-circuit (fast computing logical expressions)
  - Execute faster comparisons first
- Reduce memory needs (a.k.a., *space-complexity*)
  - Execute Selections with high *selectivity* (i.e., with more strict conditions) to reduce the size of intermediate tables.
  - Execute Projects as early as possible to reduce tuple size

Selectivity

- Selectivity = The **ratio** of the number of records that satisfy a condition to the total number of records
- Let assume that Students
  - Female = 55% & Male 45%
  - CS majors = 5% & Non-CS majors = 95%
- Which is more efficient? [Poll]
  a. $\sigma_{\text{Major='Non-CS'} \wedge \text{Gender = 'Female'}}$ (STUDENT)
  b. $\sigma_{\text{Gender = 'Female'} \wedge \text{Major='Non-CS'}}$ (STUDENT)
  c. $\sigma_{\text{Major='CS'} \wedge \text{Gender = 'Female'}}$ (STUDENT)

13

Basic Set Operations

- $r \cup s$
- $r \cap s$
- $r - s$

relation **r**

A	B	C
a	b	c
d	a	f
c	b	d

relation **s**

A	B	C
b	g	a
d	a	f

- Can we perform $\cup$, $\cap$, $-$ between any two relations?
  - They need to be *union compatible*
    - |R| = |S| and
    - corresponding attributes have same domains
- Properties
  - Both $\cup$ and $\cap$ are commutative operations
  - Difference is not commutative

Attribute Names?

14

Basic Set Operations

- $r \cup s$
- $r \cap s$
- $r - s$

relation **r**

A	B	C
a	b	c
d	a	f
c	b	d

relation **s**

D	E	F
b	g	a
d	a	f

- Can we perform $\cup$, $\cap$, $-$ between any two relations?
  - They need to be *union compatible*
    - |R| = |S| and
    - corresponding attributes have same domains
- Properties
  - Both $\cup$ and $\cap$ are commutative operations
  - Difference is not commutative

Attribute Names?

15

Cartesian Product

- r x s

relation **r**

A	B	C
a	b	c
d	a	f
c	b	d

$\underbrace{\hspace{3em}}_{\alpha_r}$

relation **s**

A	B
b	g
d	a

$\underbrace{\hspace{2em}}_{\alpha_s}$

- Let p(P) = r(R) x s(S)
- |P| = ? and |p| = ?
  - $|P| = |R| + |S| = \alpha_r + \alpha_s$
  - $|p| = |r| * |s|$
- Name conflicts are resolved by using the relations names as prefixes: r.A, r.B, S.A, S.B

16