IIT Madras
BSc Degree

Module 16

Partha Pratim Das

Database Management Systems

Module 16: Formal Relational Query Languages/1

Partha Pratim Das

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

ppd@cse.iitkgp.ac.in

Week Recap

- SQL Examples have been practiced for basic query structures
- Nested Subquery in SQL
- Data Modification
- SQL expressions for Join and Views
- Transactions
- Integrity Constraints
- More data types in SQL
- Authorization in SQL
- Functions and Procedures in SQL
- Triggers

Module Objectives

- To understand formal query language through relational algebra

Module Outline

- Relational Algebra

Formal Relational Query Language

- Relational Algebra
  - Procedural and Algebra based
- Tuple Relational Calculus
  - Non-Procedural and Predicate Calculus based
- Domain Relational Calculus
  - Non-Procedural and Predicate Calculus based

Relational Algebra

Relational Algebra

- Created by Edgar F Codd at IBM in 1970
- Procedural language
- Six basic operators
    - select: σ
    - project: Π
    - union: $\cup$
    - set difference: $-$
    - Cartesian product: x
    - rename: ρ
- The operators take one or two relations as inputs and produce a new relation as a result

Select Operation

- Notation: $\sigma_p(r)$
- p is called the selection predicate
- Defined as:

$$\sigma_p(\mathbf{r}) = \{t | t \in r \textbf{ and } p(t)\}$$

where p is a formula in propositional calculus consisting of terms connected by : $\wedge$ (**and**), $\vee$ (**or**), $\neg$ (**not**)
Each terms is one of:

$$< attribute > op < attribute > or < constant >$$

where op is one of: $=, \neq, >, \geq . < . \leq$

- Example of selection:

$$\sigma_{dept\_name = 'Physics'}(instructor)$$

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

A	B	C	D
α	α	1	7
β	β	23	10

$\sigma_{A=B \wedge D > 5}(r)$

Project Operation

- Notation: $\Pi_{A_1, A_2, \ldots A_k}$ (r)
  where A_1, A_2 are attribute names and r is a relation
- The result is defined as the relation of k columns obtained by erasing the columns that are not listed
- Duplicate rows removed from result, since relations are sets
- Example: To eliminate the *dept_name* attribute of *instructor*

  $$\Pi_{ID, name, salary}(instructor)$$

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

A	C
α	1
α	1
β	1
β	2

=

A	C
α	1
β	1
β	2

Union Operation

- Notation: $r \cup s$
- Defined as: $r \cup s = \{t | t \in r \text{ or } t \in s\}$
- For $r \cup s$ to be valid.
  a) r, s must have the *same* arity (same number of attributes)
  b) The attribute domains must be compatible (example: 2nd column of r deals with the same type of values as does the 2nd column of s)
  c) Example: to find all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or in both

$$\Pi_{course\_id}(\sigma_{semester=\text{``Fall''} \wedge year=2009}(section)) \cup \Pi_{course\_id}(\sigma_{semester=\text{``Spring''} \wedge year=2010}(section))$$

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

A	B
α	1
α	2
β	1
β	3

r ∪ s

Difference Operation

- Notation $r - s$
- Defined as: $r - s = \{t | t \in r \text{ and } t \notin s\}$
- Set differences must be taken between **compatible** relations
  - r and s must have the same arity
  - attribute domains of r and s must be compatible
- Example: to find all courses taught in the Fall 2009 semester, but not in the Spring 2010 semester

$$\Pi_{course\_id}(\sigma_{semester=\text{"Fall"} \wedge year=2009}(section)) -$$
$$\Pi_{course\_id}(\sigma_{semester=\text{"Spring"} \wedge year=2010}(section))$$

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

A	B
α	1
β	1

r − s

Intersection Operation

- Notation: $r \cap s$
- Defined as:

$$r \cap s = \{t | t \in r \textbf{ and } t \in s\}$$

- Assume:
  - r, s have the *same arity*
  - attributes of r and s are compatible
- Note: $r \cap s \ = r - (r - s)$

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

A	B
α	2

$r \cap s$

Cartesian-Product Operation

- Notation $r \times s$
- Defined as:

$$r \times s = \{t\ q | t \in r \textbf{ and } q \in s\}$$

- Assume that attributes of $r(R)$ and $s(S)$ are disjoint. (That is, $R \cap S = \phi$)
- If attributes of $r(R)$ and $s(S)$ are not disjoint, then renaming must be used

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

r x s

Rename Operation

- Allows us to name, and therefore to refer to, the results of relational-algebra expressions.
- Allows us to refer to a relation by more than one name.
- Example:

  $\rho_X(E)$

  returns the expression E under the name X
- If a relational-algebra expression E has arity n, then

$$\rho_{X(A_1, A_2, \cdots, A_n)}(E)$$

returns the result of expression E under the name X, and with the attributes renamed to

$$A_1, A_2, \ldots, A_n$$

.

Division Operation

- The division operation is applied to two relations
- R(Z) $\div$ S(X), where X subset Z. Let Y = Z - X (and hence Z = X $\cup$ Y); that is, let Y be the set of attributes of R that are not attributes of S
- The result of DIVISION is a relation T(Y) that includes a tuple t if tuples t_R appear in R with t_R [Y] = t, and with
  - $t_R[X] = t_s$ for every tuple t_s in S.
- For a tuple t to appear in the result T of the DIVISION, the values in t must appear in R in combination with every tuple in S
- Division is a derived operation and can be expressed in terms of other operations
- $r \div s \equiv \Pi_{R-S}(r) - \Pi_{R-S}((\Pi_{R-S}(r) \times s) - \Pi_{R-S,S}(r))$

Division Examples

- R

Lecturer	Module
Brown	Compilers
Brown	Databases
Green	Prolog
Green	Databases
Lewis	Prolog
Smith	Databases

S

Subject
Prolog

R | S

Lecturer
Green
Lewis

Division Examples (2)

- R

Lecturer	Module
Brown	Compilers
Brown	Databases
Green	Prolog
Green	Databases
Lewis	Prolog
Smith	Databases

S

Subject
Databases
Prolog

R | S

Lecturer
Green

Division Examples (3)

A

sno	*pno*
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

B1

pno
p2

B2

pno
p2
p4

B3

pno
p1
p2
p4

A/B1

sno
s1
s2
s3
s4

A/B2

sno
s1
s4

A/B3

sno
s1

Division Example (4)

- Relations r, s:

A	B
α	1
α	2
α	3
β	1
γ	1
δ	1
δ	3
δ	4
ϵ	6
ϵ	1
β	2

r

B
1
2

s

e.g. A is customer name
B is branch-name
1 and 2 here show two specific branch-names
(Find customers who have an account in all branches of the bank)

- $r \div s$:

A
α
β

Division Example (5)

- Relations r, s:

A	*B*	*C*	*D*	*E*
α	a	α	a	1
α	a	γ	a	1
α	a	γ	b	1
β	a	γ	a	1
β	a	γ	b	3
γ	a	γ	a	1
γ	a	γ	b	1
γ	a	β	b	1

r

D	*E*
a	1
b	1

s

e.g. Students who have taken both "a" and "b" courses, with instructor "1"

(Find students who have taken all courses given by instructor 1)

- $r \div s$:

A	*B*	*C*
α	a	γ
γ	a	γ

Source: db.fcngroup.nl/silberslides/Divsion

Module Summary

- Discussed relational algebra with examples

Slides used in this presentation are borrowed from http://db-book.com/ with kind permission of the authors.
Edited and new slides are marked with "PPD".