Relational Algebra 2

Unit 3.4

Relational Algebra - Example

Consider the following SQL to find which departments have had employees on the `Further Accounting` course.

```
SELECT DISTINCT dname
FROM department, course, empcourse, employee
WHERE cname = `Further Accounting'
  AND course.courseno = empcourse.courseno
  AND empcourse.empno = employee.empno
  AND employee.depno = department.depno;
```

The equivalent relational algebra is:

$PROJECT_{dname}$ (department JOIN $_{depno\ =\ depno}$ (
$PROJECT_{depno}$ (employee JOIN $_{empno\ =\ empno}$ (
$PROJECT_{empno}$ (empcourse JOIN $_{courseno\ =\ courseno}$ (
$PROJECT_{courseno}$ (SELECT $_{cname\ =\ \text{`Further Accounting' course}}$)
))))))

Symbolic Notation

From the example, one can see that for complicated cases a large amount of the answer is formed from operator names, such as PROJECT and JOIN. It is therefore commonplace to use symbolic notation to represent the operators.

- SELECT ->σ (sigma)
- PROJECT -> π (pi)
- PRODUCT -> × (times)
- JOIN -> |x| (bow-tie)
- UNION -> ∪ (cup)
- INTERSECTION -> ∩ (cap)
- DIFFERENCE -> - (minus)
- RENAME ->ρ (rho)

Usage

The symbolic operators are used as with the verbal ones. So, to find all employees in department 1:

SELECT $_{depno = 1}$ (employee)
becomes: $\sigma_{depno = 1}$ (employee)

Conditions can be combined together using ^ (AND) and v (OR). For example, all employees in department 1 called `Smith':

SELECT $_{depno = 1 \wedge surname = `Smith'}$ (employee)
becomes: $\sigma_{depno = 1 \wedge surname = `Smith'}$ (employee)

Usage Cont…

The use of the symbolic notation can lend itself to brevity. Even better, when the JOIN is a natural join, the JOIN condition may be omitted from |x|. The earlier example resulted in:

PROJECT $_{\text{dname}}$ (department JOIN $_{\text{depno = depno}}$ (
PROJECT $_{\text{depno}}$ (employee JOIN $_{\text{empno = empno}}$ (
PROJECT $_{\text{empno}}$ (empcourse JOIN $_{\text{courseno = courseno}}$ (
PROJECT $_{\text{courseno}}$ (SELECT $_{\text{cname = `Further Accounting'}}$ course)))))))

becomes

π_{dname} (department |x| (
π_{depno} (employee |x| (
π_{empno} (empcourse |x| (
π_{courseno} ($\sigma_{\text{cname = `Further Accounting'}}$ course)))))))

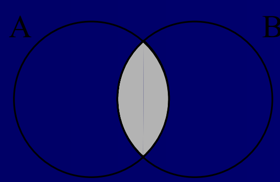
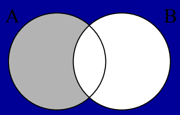
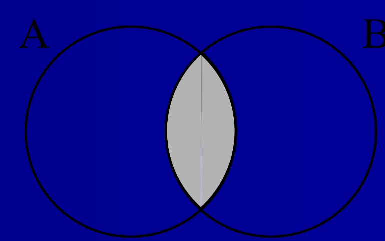
Rename Operator

The rename operator returns an existing relation under a new name. σA(B) is the relation B with its name changed to A. For example, find the employees in the same Department as employee 3.

$$\pi_{\text{emp2.surname,emp2.forenames}} (\sigma_{\text{employee.empno = 3 ^ employee.depno = emp2.depno}} (\text{employee} \times (\rho_{\text{emp2}} \text{employee})))$$

Derivable Operators

- Fundamental operators: σ, π, $\times$, $\cup$, -, ρ
- Derivable operators: $|x|$, $\cap$

$A \cap B$	$\Leftrightarrow$	A -	(A – B)
			
			

Deriving |x| from π, σ, and $\times$

- $A |x|_{c} B \Leftrightarrow \pi_{a1,a2,\ldots aN} (\sigma_c (A \times B))$
  - where c is the join condition (eg A.a1 = B.a1),
  - and a1,a2,...aN are all the attributes of A and B without repetition.
- c is called the join-condition, and is usually the comparison of primary and foreign key.
- Where there are N tables, there are usually N-1 join-conditions.
- In the case of a natural join, the conditions can be missed out, but otherwise missing out conditions results in a cartesian product (a common mistake to make).

Equivalences

The same relational algebraic expression can be written in many different ways. The order in which tuples appear in relations is never significant.

- $A \times B <=> B \times A$
- $A \cap B <=> B \cap A$
- $A \cup B <=> B \cup A$
- (A - B) is not the same as (B - A)
- $\sigma_{c1} (\sigma_{c2} (A)) <=> \sigma_{c2} (\sigma_{c1} (A)) <=> \sigma_{c1 \wedge c2} (A)$
- $\pi_{a1}(A) <=> \pi_{a1}(\pi_{a1,etc}(A))$, where etc is any attributes of A.
- ...

While equivalent expressions always give the same result, some may be much easier to evaluate that others.

When any query is submitted to the DBMS, its query optimiser tries to find the most efficient equivalent expression before evaluating it.

Comparing RA and SQL

- Relational algebra:
  - is closed (the result of every expression is a relation)
  - has a rigorous foundation
  - has simple semantics
  - is used for reasoning, query optimisation, etc.
- SQL:
  - is a superset of relational algebra
  - has convenient formatting features, etc.
  - provides aggregate functions
  - has complicated semantics
  - is an end-user language.

Comparing RA and SQL

Any relational language as powerful as relational algebra is called relationally complete.

A relationally complete language can perform all basic, meaningful operations on relations.

Since SQL is a superset of relational algebra, it is also relationally complete.