Relational Model 3: Relational Algebra (Part II)

Yufei Tao

Department of Computer Science and Engineering
Chinese University of Hong Kong

Relational Algebra (Review)

We have learned the 6 fundamental operations of relational algebra:

- Rename ρ
- Selection σ
- Projection Π
- Set union $\cup$
- Set difference $-$
- Cartesian product $\times$

The operators of the previous slide can express all queries in relational algebra. However, if we rely on only those operators, some queries common in practice require lengthy expressions. To shorten those expressions, people identified the following 4 operators, each of which can be implemented using only the 6 fundamental operators, and can be used to simplify many queries:

- Assignment $\leftarrow$
- Set intersection $\cap$
- Natural join $\bowtie$
- Division $\div$

Assignment

Denoted by $T \leftarrow [expression]$

- where $[expression]$ is a relational algebra expression, and T is a table variable.
- The assignment stores in T the table output by $[expression]$.

Assignments are often used to increase clarity by cutting a long query into multiple steps, each of which can be described by a short line.

PROF

pid	**name**	**dept**	**rank**	**sal**
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500
$p6$	Frank	CS	full	9000

$$
\begin{aligned}
T_1 &\leftarrow \Pi_{\mathrm{rank}}(\sigma_{\mathrm{sal} \geq 8000}(\mathrm{PROF})) \\
T_2 &\leftarrow \Pi_{\mathrm{rank}}(\sigma_{\mathrm{sal} \geq 9000}(\mathrm{PROF})) \\
T_1 &- T_2
\end{aligned}
$$

returns:

rank
asso

Set intersection

Denoted by $T_1 \cap T_2$

- where T_1 and T_2 are tables with the same schema.
- The output of the operation is a table T' such that
  - T' has the same schema as T_1 (and hence, T_2).
  - T' contains all and only the tuples that appear in both T_1 and T_2.

PROF

pid	**name**	**dept**	**rank**	**sal**
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500
$p6$	Frank	CS	full	9000

$\sigma_{\text{sal} \geq 8500}(\text{PROF}) \cap \sigma_{\text{dept} = \text{CS}}(\text{PROF})$ returns:

pid	**name**	**dept**	**rank**	**sal**
$p3$	Calvin	CS	full	10000
$p6$	Frank	CS	full	9000

In general:

$$T_1 \cap T_2 \quad = \quad T_1 - (T_1 - T_2)$$

Natural join

Denoted by $T_1 \bowtie T_2$

- where T_1 and T_2 are tables.
- The output of the operation is a table T' such that
  - The schema of T' includes all the distinct columns of T_1 and T_2.
  - T' contains all and only the tuples t satisfying the following conditions:
    - $t[T_1]$ belongs to T_1, where $t[T_1]$ is the part of t after trimming the attributes that do not exist in T_1;
    - $t[T_2]$ belongs to T_2, where $t[T_2]$ is defined similarly with respect to T_2.

PROF

pid	name	dept	rank	sal
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500

TEACH

pid	cid	year
$p1$	$c1$	2011
$p2$	$c2$	2012
$p1$	$c2$	2012

PROF $\bowtie$ TEACH returns:

pid	name	dept	rank	sal	cid	year
$p1$	Adam	CS	asst	6000	c_1	2011
$p2$	Bob	EE	asso	8000	c_2	2012
$p1$	Adam	CS	asst	6000	c_2	2012

In general:

$$T_1 \bowtie T_2 \quad = \quad \Pi_S\Big(\sigma_{T_1.A_1 \,=\, T_2.A_1 \,\wedge\, ... \,\wedge\, T_1.A_d \,=\, T_2.A_d}(T_1 \times T_2)\Big)$$

where

$$S = (S_1 - S_2) \cup \{T_1.A_1, ..., T_1.A_d\} \cup (S_2 - S_1)$$

where S_1 and S_2 are the schemas of T_1 and T_2 respectively, and $A_1, ..., A_d$ are the common attributes of T_1 and T_2.

Division

Denoted by $T_1 \div T_2$

- where T_1 and T_2 are tables such that the schema of T_2 is a subset of the schema of T_1.
- The output of the operation is a table T' such that
  - The schema of T' includes all the columns that are in T_1, but not in T_2.
  - T' contains all and only the tuples t such that:
    - for every tuple $t_2 \in T_2$, $t_1 = (t, t_2)$ is a tuple in T_1, where (t, t_2) represents a tuple that concatenates the attributes of t with those of t_2.

T_1

pid	**cid**
$p1$	$c1$
$p1$	$c2$
$p1$	$c3$
$p2$	$c2$
$p2$	$c3$
$p3$	$c1$
$p4$	$c1$
$p4$	$c2$
$p4$	$c3$

T_2

cid
$c1$
$c2$
$c3$

$T_1 \div T_2$ returns:

pid
$p1$
$p4$

In general:

$$T_1 \div T_2 \quad = \quad \Pi_{S_1 - S_2}(T_1) - \Pi_{S_1 - S_2}\Big(\Pi_{S_1 - S_2}(T_1) \times T_2 - T_2\Big)$$

where S_1 and S_2 are the schemas of T_1 and T_2 respectively.