Relational Algebra

Overview

- Relational algebra offers a concise way to express queries.
  - Form the basis for “real” query languages (SQL).
  - Much more concise than SQL.

- It is widely used by database professionals.

Role of Relational Algebra

- How does a relational DBMS work?
  - Queries are expressed by users in a language, e.g. SQL;
  - The DBMS translates an SQL query into *relational algebra*, and meanwhile looks for other algebra expressions that produce the same answers but saving the computational costs.
  - Based on the relational algebra, DBMS calculates the query results.

Basic operations

- Selection
- Projection
- Union
- Difference
- Rename
- Cartesian product
- Natural join

Example Tables

- "Sailors" and "Reserves" relations are used for our examples.

R1

sid	bid	day
22	101	10/10/96
58	103	11/12/96

S1

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

Selection

- Notation: $\sigma_p(r)$

- p is called the **selection predicate**, r is a relation
- Defined as:

$$\sigma_p(r) = \{t \mid t \in r \text{ and } p(t)\}$$

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

Each **term** is one of:

<attribute> *op* <attribute> or <constant>

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

Selection (1/3)

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$$\sigma_{rating>8}(S2)$$

sid	sname	rating	age
28	yuppy	9	35.0
58	rusty	10	35.0

Selection (2/3)

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$$\sigma_{rating>8 \wedge sid>40}(S2)$$

sid	sname	rating	age
58	rusty	10	35.0

$\wedge$ means logical AND.

Selection (3/3)

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$$\sigma_{rating>8 \vee sid>40}(S2)$$

sid	sname	rating	age
28	yuppy	9	35.0
44	guppy	5	35.0
58	rusty	10	35.0

$\vee$ means logical OR.

Question 1

S

sid	sname	gpa	age
42	David	4.0	21
15	Louis	2.8	19
98	Amy	1.7	20

- What are the results for the following queries?
  - $\sigma_{gpa > 2.6 \wedge gpa \leq 4.0} (\sigma_{age \geq 19 \wedge age \leq 20} (S))$
  - $\sigma_{age \geq 19 \wedge age \leq 20} (\sigma_{gpa > 2.6 \wedge gpa \leq 4.0} (S))$
  - $\sigma_{age \leq 19} (\sigma_{gpa > 2.8} (S))$

Projection

- Notation: $\pi_{A\_1,A\_2,\ldots,A\_k}(r)$
  where A_1, A_2 are attributes 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 are removed from result, since relations are sets.

Projection

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$\pi_{age}(S2)$

age
35.0
55.5

Duplicate rows are removed.

Operator composition

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$$\pi_{sname,rating}(\sigma_{rating>8}(S2))$$

sname	rating
yuppy	9
rusty	10

Composition of Selection and Projection.

Question 2

S

sid	sname	gpa	age
42	David	4.0	21
15	Louis	2.8	19
98	Amy	1.7	20

- What is the result for the following query?

$$\pi_{age}(\sigma_{gpa > 2.6 \wedge gpa \leq 4.0}(S))$$

Union

- Two input relations must be *union-compatible*.
  - Same number of attributes.
  - `Corresponding' attributes have the same type.

S1

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$S1 \cup S2$

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0
44	guppy	5	35.0
28	yuppy	9	35.0

Intersection and Difference

S1

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S2

sid	sname	rating	age
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

$S1 \cap S2$

sid	sname	rating	age
31	lubber	8	55.5
58	rusty	10	35.0

$S1 - S2$

sid	sname	rating	age
22	dustin	7	45.0

Renaming

S1

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

$\rho_{My\text{-}table(id,\ name,\ level,\ age)}$ (S1)

My-table

id	name	level	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Cartesian Product

- Each row of S1 is paired with each row of R1.

R1

sid	bid	day
22	101	10/10/96
58	103	11/12/96

S1

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S1 × R1 =

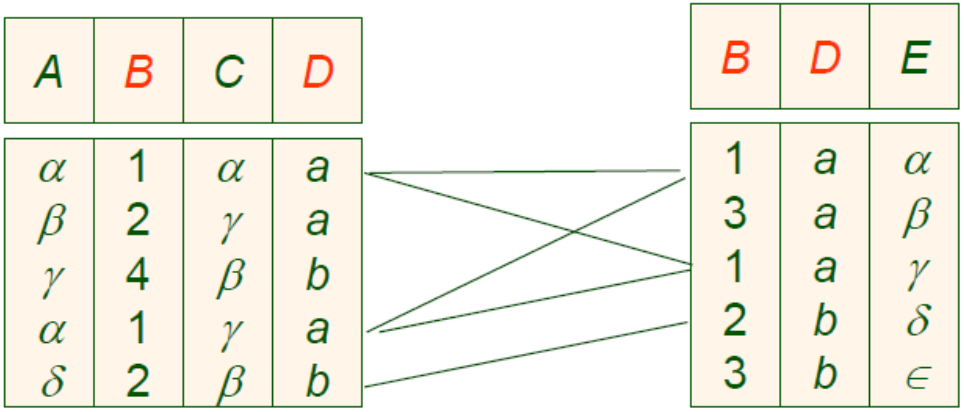
S1. sid	sname	rating	age	R1. sid	bid	day
22	dustin	7	45.0	22	101	10/10/96
22	dustin	7	45.0	58	103	11/12/96
31	lubber	8	55.5	22	101	10/10/96
31	lubber	8	55.5	58	103	11/12/96
58	rusty	10	35.0	22	101	10/10/96
58	rusty	10	35.0	58	103	11/12/96

Natural Join Operation

- $R = (A, B, C, D)$, $S = (E, B, D)$
- Equal on ***all common*** attributes

$$r \bowtie s = \Pi_{r.A,\ r.B,\ r.C,\ r.D,\ s.E} (\sigma_{r.B = s.B \wedge\ r.D = s.D} (r \times s))$$

Result schema = (A, B, C, D, E)



A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

r

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	$\in$

s

$r \bowtie s$

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

Natural Join Operation

S1. sid	sname	rating	age	R1. sid	bid	day
22	dustin	7	45.0	22	101	10/10/96
22	dustin	7	45.0	58	103	11/12/96
31	lubber	8	55.5	22	101	10/10/96
31	lubber	8	55.5	58	103	11/12/96
58	rusty	10	35.0	22	101	10/10/96
58	rusty	10	35.0	58	103	11/12/96

$$\sigma_{S1.sid=R1.sid}(S1\times R1)$$

S1. sid	sname	rating	age	R1. sid	bid	day
22	dustin	7	45.0	22	101	10/10/96
58	rusty	10	35.0	58	103	11/12/96

Question 3

S

sid	sname	gpa	age
42	David	4.0	21
15	Louis	2.8	19
98	Amy	1.7	20

E

sid	cid	day
15	2016	01/09/18
15	2006	12/01/18
42	4035	11/01/20

- What are the results for the following queries?
  - $S \bowtie_{S.sid = E.sid} E$
  - $S \bowtie_{S.sid = E.sid} (\sigma_{cid = 4035} (E))$

Solution to Question 1

sid	sname	gpa	age
15	Louis	2.8	19

sid	sname	gpa	age
15	Louis	2.8	19

- Empty

Solution to Question 2

-

age
21
19

Solution to Question 3

-

sid	sname	gpa	age	cid	day
42	David	4.0	21	4035	11/01/20
15	Louis	2.8	19	2016	01/09/18
15	Louis	2.8	19	2006	12/01/18

-

sid	sname	gpa	age	cid	day
42	David	4.0	21	4035	11/01/20