Database Management Systems

Vijaya Saradhi

IIT Guwahati

Fri, 07^{th} Feb 2020

Cartesian Product - $\times$

Definition

- Takes two relation as input
- Produces a new relation that has
  - Number of columns = sum of the individual relations columns
  - A set $\{(x, y) | x \in R \; y \in S\}$
  - Notation: $R \times S$

Example

Relation R_1

A	B
1	2
3	4

Relation S_1

B	C	D
2	5	6
4	7	8
9	10	11

Relation $R_1 \times S_1$

R_1.A	R_1.B	S_1.B	S_1.C	S_1.D
1	2	2	5	6
1	2	4	7	8
1	2	9	10	11
3	4	2	5	6
3	4	4	7	8
3	4	9	10	11

Natural Join

Definition

- Cross product 'joins' relations without conditions
- It is compute and memory intense operation
- Joining relations on a specified criteria
- Join tuples by attributes that are common to relations in question
- Join take place when the values are identical among common attributes of the two relations
- Produces a new relation that has
- Notation: $R \bowtie S$

Natural Join

Example

Relation R_1	
A	B
1	2
3	4

Relation S_1		
B	C	D
2	5	6
4	7	8
9	10	11

Relation $R_1 \bowtie S_1$				
R_1.A	R_1.B	S_1.B	S_1.C	S_1.D
1	2	2	5	6
3	4	4	7	8

one of the columns R_1.B or S_1.B is retained.

Theta Join

Definition

- Join condition in Natural join is based on common attributes
- Joins tuples from two relations based on a specified condition
- Produces a new relation that has tuples satisfying specified condition
- Notation: $R \underset{C}{\bowtie} S$

Example

Relation U

A	B	C
1	2	3
6	7	8
9	7	8

Relation V

B	C	D
2	3	4
2	3	5
7	8	10

Relation $U \underset{A<D}{\bowtie} V$

U.A	U.B	U.C	V.B	V.C	V.D
1	2	3	2	3	4
1	2	3	2	3	5
1	2	3	7	8	10
6	7	8	7	8	10
9	7	8	7	8	10

Renaming Relations

Definition

Rename a given relation and/or attribute of a given relation

Example - 1

Rename U(A, B, C) to $U_1(A, B, C)$: $\rho(U_1, U)$

Example - 2

Rename $U(A, B, C)$ to $U_1(A_1, B, C_1)$: $\rho\left(\underline{U_1(A_1, B, C_1)}, \overline{U(A, B, C)}\right)$

Combining Operations To Form Queries - 1

Composition

- Operators are applied on relations
- Every operator results in a relation
- Operators can be composed to form a bigger expression

Example

T

Title	Year	length	inColor	studioName	producerNo
Star Wars	1977	124	true	Fox	12345
Mighty Ducks	1991	104	true	Disney	67890
Wayne's World	1992	95	true	Paramount	99999

What are the titles and years of movies made by Fox that are at least 100 minutes long?

Combining Operations To Form Queries - 2

Example

1. Select those tuples from T that have length ≥ 100
2. Select those tuples from T that have studioName $=$ Fox
3. Compute intersection of (1) and (2)
4. Project the relation from (3) onto attributes title, year

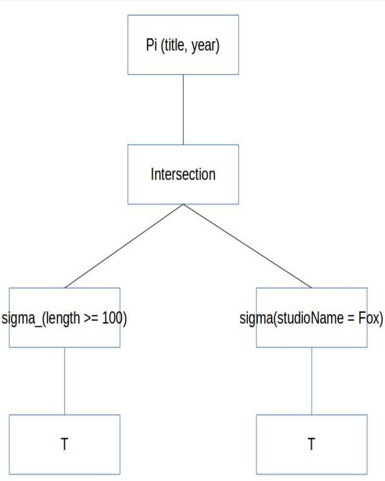
Example

1. $\rho(T_1, \sigma_{length \geq 100}(T))$
2. $\rho(T_2, \sigma_{studioName='Fox'}(T))$
3. $\rho(T_3, T_1 \cap T_2)$
4. $\pi_{\{title, year\}}(T_3)$

$$\pi_{\{title, year\}}\left(\left(\sigma_{length \geq 100}(T)\right) \cap \left(\sigma_{studioName='Fox'}(T)\right)\right)$$

Combining Operations To Form Queries - 3

Example



Combining Operations To Form Queries - 4

1. $\rho(T_1, \sigma_{length \geq 100}(T))$
2. $\rho(T_2, \sigma_{studioName='Fox'}(T))$
3. $\rho(T_3, T_1 \cap T_2)$
4. $\pi_{\{title, year\}}(T_3)$

$$\pi_{\{title, year\}}\left(\sigma_{length \geq 100\ AND\ studioName\ =\ 'Fox'}(T)\right)$$

Expression Equivalence

Definition

Two algebraic expressions are equivalent when both yield identical relations

Example - 1

$R \cap S = R - (R - S)$

Example - 2

$R \underset{C}{\bowtie} S = \sigma_C(R \times S)$

Example - 3

$R \bowtie S = \pi_L(\sigma_C(R \times S))$

Linear Notation

Notation

1. $R(t, y, l, i, s, p) := \sigma_{length \geq 100}(T))$
2. $S(t, y, l, i, s, p) := \sigma_{studioName='Fox'}(T))$
3. $T(t, y, l, i, s, p) := R \cap S$
4. $Answer(title, year) := \pi_{t,y}(T)$

Relational Operations on Bags

Definition

When a set is allowed to have multiuple occurences of a member

Example

A	B
1	2
3	4
1	2
1	2

Why Bags?

1. Duplicate elimination is an expensive operation
2. Aggregation operations results in incorrect computation
3. Example: $avg(A) = 2$ vs $avg(A) = 1.5$?

Example - Result of operation - a Bag

Relation Example

R		
A	B	C
1	2	5
3	4	6
1	2	7
1	2	8

$\pi_{\{A,B\}}(R)$

Revised Definitions

Union of Bags

- Let a tuple $t \in R$ appears n times
- Let $t \in S$ appears m times
- $t \in (R \cup S)$ appears $(n + m)$ times

Intersection of Bags

- Let a tuple $t \in R$ appears n times
- Let $t \in S$ appears m times
- $t \in (R \cap S)$ appears $\min(n, m)$ times

Difference of Bags

- Let a tuple $t \in R$ appears n times
- Let $t \in S$ appears m times
- $t \in (R - S)$ appears $\max(0, (n - m))$ times

Examples Using Bags

R		S	
A	B	A	B
1	2	1	2
3	4	3	4
1	2	3	4
1	2	5	6

Union of Bags

$R \cup S$	
A	B
1	2
1	2
1	2
1	2
3	4
3	4
3	4
5	6

Examples Using Bags

R		S	
A	B	A	B
1	2	1	2
3	4	3	4
1	2	3	4
1	2	5	6

Intersection of Bags

$R \cap S$	
A	B
1	2
3	4

Examples Using Bags

R		S	
A	B	A	B
1	2	1	2
3	4	3	4
1	2	3	4
1	2	5	6

Difference of Bags

$R - S$	
A	B
1	2
1	2

Examples Using Bags

R		S	
A	B	A	B
1	2	1	2
3	4	3	4
1	2	3	4
1	2	5	6

Projection of Bags

$\pi_{A,B}(R)$	
A	B
1	2
3	4
1	2
1	2

Selection on Bags

Example Relation

R		
A	B	C
~~1~~	~~2~~	~~5~~
3	4	6
1	2	7
1	2	7

Selection on Bags

$\sigma_{C \geq 6}(R)$		
A	B	C
3	4	6
1	2	7
1	2	7

Cross Product on Bags

Example Relations

R		S	
A	B	B	C
1	2	2	3
1	2	4	5
		4	5

Cross Product

R.A	R.B	S.B	S.C
1	2	2	3
1	2	4	5
1	2	4	5
1	2	2	3
1	2	4	5
1	2	4	5

Natural Join on Bags

Example Relations

R		S	
A	B	B	C
1	2	2	3
1	2	4	5
		4	5

Natural Join

R.A	R.B	S.B	S.C
1	2	2	3
1	2	2	3

Theta Join on Bags

Example Relations

R		S	
A	B	B	C
1	2	2	3
1	2	4	5
		4	5

Natural Join

$$R \underset{R.B<S.B}{\bowtie} S$$

R.A	R.B	S.B	S.C
1	2	4	5
1	2	4	5
1	2	4	5
1	2	4	5

Comprehensive List

- δ Duplicate elimination turns Bag into a Set
- **Aggregation** SUM, AVG, MIN, MAX, COUNT
- γ Grouping of tuples according to their value in one or more attributes; partitioning the tuples into groups
- τ Sorting - Sorts given list of columns
- **Extended Projection** - Projection includes expressions and columns that are not available in original relation
- **Outer Join** Takes into account dangling tuples - the tuples that do not satisfy specified condition

Duplicate Elimination

Example

R	
A	B
1	2
3	4
1	2
1	2

Example

$\delta(R)$	
A	B
1	2
3	4

Aggregation

Example

R	
A	B
1	2
3	4
1	2
1	2

SUM

$R_{SUM(B)}$
SUM(B)
10

Aggregation

Example

R	
A	B
1	2
3	4
1	2
1	2

AVG

$R_{AVG(A)}$
AVG(A)
1.5

Aggregation

Example

R	
A	B
1	2
3	4
1	2
1	2

MIN

$R_{MIN(A)}$
MIN(A)
1

Aggregation

Example

R	
A	B
1	2
3	4
1	2
1	2

MAX

$R_{MAX(A)}$
MAX(A)
4

Aggregation

Example

R	
A	B
1	2
3	4
1	2
1	2

COUNT

$R_{COUNT(A)}$
COUNT(A)
4

Grouping

	studioName	
	Disney	
	Disney	
	Disney	
	MGM	
	MGM	
	...	
	...	
	...	

Grouping

- γ is applied to an attribute of R
- Partitions tuples of R into groups
- Each group consists of all tuples having one particular assignment of values
- If there are no grouping attributes, entire relation is one group
- For each group produce one tuple consisting of
  - The grouping attributes' values for that group and
  - The aggregations over all tubples of that group for the aggregated attributes on list L

Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

$\gamma_{rating}(Sailors)$			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

$\gamma_{rating}(Sailors)$			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

$\gamma_{rating}(Sailors)$			
sid	sname	rating	age
22	Dustin	7	45.0
64	Horatio	7	35.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping Example

Example Relation

Sailors			
sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

$\gamma_{rating}(Sailors)$			
sid	sname	rating	age
22	Dustin	7	45.0
~~64~~	~~Horatio~~	~~7~~	~~35.0~~
29	Brutus	1	33.0
31	Lubber	8	55.5
~~32~~	~~Andy~~	~~8~~	~~25.5~~
58	Rusty	10	35.0
~~71~~	~~Zorba~~	~~10~~	~~16.0~~
74	Horatio	9	35.0
85	Art	3	25.5
~~95~~	~~Bob~~	~~3~~	~~63.5~~

Grouping Example

Example Relation

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Grouping

$\gamma_{rating}(Sailors)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
58	Rusty	10	35.0
74	Horatio	9	35.0
85	Art	3	25.5

Extended Projection

Definition

- In the projection operation L should be from list of attributes of R
- Extended projection can create new attributes from existing L
- Notation: $\pi_L(R)$

Extended Projection

Elements of Projection

- A single attribute of R
- An expression $x \rightarrow y$ renames x to y
- An expression$E \rightarrow z$; E is an experssion involving attributes of R
- Example: $a + b \rightarrow x$; Sums attribute values a and b and rename it to x

Extended Projection

Example Relation

R		
A	B	C
0	1	2
0	1	2
3	4	5

$\pi_{A,B+C\rightarrow X}(R)$

A	X
0	3
0	3
3	9

Extended Projection

Example Relation

R		
A	B	C
0	1	2
0	1	2
3	4	5

$\pi_{B-A \rightarrow X, C-B \rightarrow Y}(R)$

X	Y
1	1
1	1
1	1

Sorting Operator

Definition

Sort given attribute in ascending/descending order

Example Relation

R

A	B	C
3	4	5
1	1	2
7	1	2

$\tau_A(R)$

R

A	B	C
1	1	2
3	4	5
7	1	2

Outer Join

Variants

- Outer join: $\overset{\circ}{\bowtie}$
- Left outer join: $\overset{\circ}{\underset{L}{\bowtie}}$
- Right outer join: $\overset{\circ}{\underset{R}{\bowtie}}$

Outer join

Definition: In addition to the Natural join, add any dangling tuples from R or S.
Dangling tuples: The tuples that did not meet the Natural join criteria

Outer Join

Outer join - Example

U

A	B	C
1	2	3
4	5	6
7	8	9

V

B	C	D
2	3	10
2	3	11
6	7	12

$U \overset{\circ}{\bowtie} V$

A	B	C	D
1	2	3	10
1	2	3	11
4	5	6	$\perp$
7	8	9	$\perp$
$\perp$	6	7	12

Left/Right Outer Join

Left/Right Outer Join - Example

U

A	B	C
1	2	3
4	5	6
7	8	9

V

B	C	D
2	3	10
2	3	11
6	7	12

$U \underset{L}{\overset{\circ}{\bowtie}} V$

A	B	C	D
1	2	3	10
1	2	3	11
4	5	6	$\perp$
7	8	9	$\perp$

Left/Right Outer Join

Left/Right Outer Join - Example

U

A	B	C
1	2	3
4	5	6
7	8	9

V

B	C	D
2	3	10
2	3	11
6	7	12

$U \overset{\circ}{\underset{R}{\bowtie}} V$

A	B	C	D
1	2	3	10
1	2	3	11
$\perp$	6	7	12

Relational Algebra as a Constraint Language

Constraints on Relations

- Two ways of expressing constraints using relational algebra
- If R is an expression of relational algebra then $R = \Phi$. That is there are no tuples in the result of R
- If R and S are expressions of relational algebra then $R \subseteq S$ is a constraint
- States Every tuple in the result of R must also be a tuple in the result of S
- $R \subseteq S$ could just as well have been written as $R - S = \Phi$

Referential Integrity

Using Relational Algebra

- `Student(roll_number, name, email, phone)`
- `Course(cid, cname, credits)`
- `Registers(roll_number, cid, semester, year)`
- $\pi_{roll\_number}(Registers) \subseteq \pi_{roll\_number}(Student)$
- $\pi_{roll\_number}(Registers) - \pi_{roll\_number}(Student) = \Phi$