Database Management Systems

Vijaya Saradhi

IIT Guwahati

Mon, 10^{th} Feb 2020

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

Grouping on
Department attribtue

	Department	
	EEE	
	CSE	
	EEE	
	CSE	
	...	
	...	
	CSE	
	...	

Grouping

On grouping on
Department attribtue

	Department	
	CSE	
	CSE	
	CSE	
	EEE	
	EEE	
	...	
	...	
	...	

Grouping

On grouping on Department attribtue

	Department	
	CSE	
	~~CSE~~	
	~~CSE~~	
	EEE	
	~~EEE~~	
	...	
	...	
	...	

Grouping

- Grouping operator allows us to group a relation
- It also allows to apply aggregate operations on some attributes
- The notation is $\gamma_L(R)$
- Where L is a list of attributes each of which is either
  - An attribute of R to which γ is applied known as grouping attribute
  - An aggregation operator applied to an attribtue of R
- Example: $\gamma_{\text{dept,avg(cpi)}}(R)$
- In this example, attribtue `dept` is the grouping attribute
- The attribute `cpi` is not a grouping attribute but is an attribute of R on which aggregate operation avg is performed

Grouping

- The relation returned by the expression $\gamma_L(R)$ is constructed as follows:
- Partitions tuples of R into groups
- Each group consists of all tuples having one particular assignment of values to grouping attributes in L
- 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 tuples of that group, for the aggregated attributes on list L
- Example: $\gamma_{\text{dept},\text{avg(cpi)}}(R)$
- Department attribute takes several values {BT, CH, CE, CS, DD, EC, EE, MA, ME, PH}
- For each of the value, produce one tuple
- That is for BT produce one record (no constraint on which record within BT is produced)
- Apply aggregation to the specified attribute within the group
- That is within BT compute the average `cpi`

Grouping Example

Example Relation

Student			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping

$\gamma_{dept,avg(cpi)}(Student)$			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping Example

Example Relation

Student			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping

$\gamma_{dept,avg(cpi)}(Student)$			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping Example

Example Relation

Student			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping

$\gamma_{dept,avg(cpi)}(Student)$			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
64	Horatio	CSE	8.8
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping Example

Example Relation

Student			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping

$\gamma_{dept,avg(cpi)}(Student)$			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
~~64~~	~~Horatio~~	~~CSE~~	~~8.8~~
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
~~32~~	~~Andy~~	~~EEE~~	~~9.2~~
58	Rusty	PH	8.7
~~71~~	~~Zorba~~	~~PH~~	~~9.5~~
74	Horatio	CH	7.0
85	Art	CE	10.0
~~95~~	~~Bob~~	~~CE~~	~~9.9~~

Grouping Example

Example Relation

Student			
roll	sname	dept	cpi
22	Dustin	CSE	9.7
29	Brutus	ECE	9.3
31	Lubber	EEE	9.0
32	Andy	EEE	9.2
58	Rusty	PH	8.7
64	Horatio	CSE	8.8
71	Zorba	PH	9.5
74	Horatio	CH	7.0
85	Art	CE	10.0
95	Bob	CE	9.9

Grouping

$\gamma_{dept,avg(cpi)}(Student)$	
dept	avg(cpi)
CSE	9.25
ECE	9.3
EEE	9.1
PH	9.1
CH	7.0
CE	9.95

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

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

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

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

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

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

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

Division Operator

Definition

- Useful for expressing certain kinds of queries

Division Operator

Definition

- Useful for expressing certain kinds of queries
- Example: Find the names of sailors who have reserved all boats

Division Operator

Definition

- Useful for expressing certain kinds of queries
- Example: Find the names of sailors who have reserved all boats
- Let A and B be two relation instances

Division Operator

Definition

- Useful for expressing certain kinds of queries
- Example: Find the names of sailors who have reserved all boats
- Let A and B be two relation instances
- A has exactly two attributes x, y

Division Operator

Definition

- Useful for expressing certain kinds of queries
- Example: Find the names of sailors who have reserved <u>all</u> boats
- Let A and B be two relation instances
- A has exactly two attributes x, y
- B has exactly one attribute y

Division Operator

Definition

- Useful for expressing certain kinds of queries
- Example: Find the names of sailors who have reserved all boats
- Let A and B be two relation instances
- A has exactly two attributes x, y
- B has exactly one attribute y
- $A/B = \{x | \forall\ y \in B,\ \ \exists (x, y) \in A\}$ For every value of $y \in B$ there exists (x, y) $\in$ A

Division Operator

Alternate Definition

$A/B = \pi_x(A) - \pi_x(\pi_x(A) \times B - A)$

Explanation

- Compute all x values in A that are not disqualified

Division Operator

Alternate Definition

$A/B = \pi_x(A) - \pi_x(\pi_x(A) \times B - A)$

Explanation

- Compute all x values in A that are not disqualified
- Disqualification: By attaching y value from B to $x \in A$ such that $(x, y) \notin A$

Division Operator

Alternate Definition

$A/B = \pi_x(A) - \pi_x(\pi_x(A) \times B - A)$

Explanation

- Compute all x values in A that are not disqualified
- Disqualification: By attaching y value from B to $x \in A$ such that $(x, y) \notin A$
- Step 1: attaching y value from B to $x \in A$: $\pi_x(A) \times B$

Division Operator

Alternate Definition

$A/B = \pi_x(A) - \pi_x(\pi_x(A) \times B - A)$

Explanation

- Compute all x values in A that are not disqualified
- Disqualification: By attaching y value from B to $x \in A$ such that $(x, y) \notin A$
- Step 1: attaching y value from B to $x \in A$: $\pi_x(A) \times B$
- Step 2: check the pair (x, y) is not in A

$$(\pi_x(A) \times B) - A$$

Division Operator

A

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

B
y
p2

- Step 1: attaching y value from B to every $x \in A$: $\pi_x(A) \times B$

$\pi_x(A) \times B$

x	y
s1	p2
s1	p2
s1	p2
s1	p2
s2	p2
s2	p2
s3	p2
s4	p2
s4	p2

Division Operator

A	
x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

B
y
p2

- Step 2: check the pair (x, y) is not in A
  $(\pi_x(A) \times B) - A$

$\pi_x(A) \times B$	
x	y
s1	p2
s1	p2
s1	p2
s1	p2
s2	p2
s2	p2
s3	p2
s4	p2
s4	p2

$-$

A	
x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

$=$

$(\pi_x(A) \times B) - A$	
x	y
s1	p2
s1	p2
s1	p2
s2	p2
s4	p2

Division Operator

$\pi_x(A)$

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

$-$

$\pi_x((\pi_x(A) \times B) - A)$

x	y
s1	p2
s1	p2
s1	p2
s2	p2
s4	p2

$\pi_x A$ - $\pi_x(\pi_x(A) \times B - A)$

y
s1
s2
s3
s4

Division Operator

$A/B =$

A

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

B
y
p2
p4

A/B

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

Division Operator

$A/B =$

A

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

B
y
p2
p3
p4

A/B

x	y
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

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)`
- $\pi_{roll\_number}(Registers) \subseteq \pi_{roll\_number}(Student)$
- $\pi_{roll\_number}(Registers) - \pi_{roll\_number}(Student) = \Phi$

Example Database

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

Reserves		
sid	bid	day
22	101	10-Oct-2019
22	102	10-Oct-2019
22	103	08-Oct-2019
22	104	07-Oct-2019
31	102	10-Nov-2019
31	103	06-Nov-2019
31	104	12-Nov-2019
64	101	05-Sep-2019
64	102	08-Sep-2019
74	103	08-Sep-2019

Boats		
bid	bname	color
101	Interlake	blue
102	Interlanke	red
103	Clipper	gree
104	Marine	red

Queries on Example Database

Queries

Q1 Find the names of the Sailors who have reserved Boat 103

Queries on Example Database

Queries

Q1 Find the names of the Sailors who have reserved Boat 103

Q2 Find the names of the Sailors who reserved a red boat

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat
- Q5 Find the names of Sailors who have reserved a red or a green Boat

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat
- Q5 Find the names of Sailors who have reserved a red or a green Boat
- Q6 Find the names of Sailors who have reserved a red AND a green Boat

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat
- Q5 Find the names of Sailors who have reserved a red or a green Boat
- Q6 Find the names of Sailors who have reserved a red AND a green Boat
- Q7 Find the names of Sailors who have reserved at least two boats

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat
- Q5 Find the names of Sailors who have reserved a red or a green Boat
- Q6 Find the names of Sailors who have reserved a red AND a green Boat
- Q7 Find the names of Sailors who have reserved at least two boats
- Q8 Find the sids of Sailors with age over 20 who have not reserved a red boat

Queries on Example Database

Queries

- Q1 Find the names of the Sailors who have reserved Boat 103
- Q2 Find the names of the Sailors who reserved a red boat
- Q3 Find the colors of boats reserved by Lubber
- Q4 Find the names of Sailors who have reserved at least one boat
- Q5 Find the names of Sailors who have reserved a red or a green Boat
- Q6 Find the names of Sailors who have reserved a red AND a green Boat
- Q7 Find the names of Sailors who have reserved at least two boats
- Q8 Find the sids of Sailors with age over 20 who have not reserved a red boat
- Q9 Find the names of sailors who have reserved <u>all</u> boats

Queries on Example Database

Q2: Find the names of the Sailors who reserved a red boat

$\rho(Temp1, \sigma_{color='red'}(Boats))$

Schema for Temp1: `Temp1(bid, bname, color)`

$\rho(Temp2, Temp1 \bowtie Reserves \bowtie Sailors)$

Schema for Temp2: `Temp2(bid, bname, color, sid, day, sname, rating, age)`

$\pi_{sname}(Temp2)$

Schema for π_{sname}(Temp2): `Temp2(sname)`

Queries on Example Database

Q2: Find the names of the Sailors who reserved a red boat

$$\pi_{sname}((\sigma_{color='red'}(Boats)) \bowtie Reserves \bowtie Sailors)$$

Queries on Example Database

Q3: Find the colors of boats reserved by Lubber

$\pi_{color}((sigma_{sname='Lubber'}(Sailors)) \bowtie Reserves \bowtie Boats)$