Fpt University

Chapter 5

ALGEBRAIC QUERY LANGUAGE

Objectives

Understand why must use Bags concept (multi-set)

Know relational operations on bags

Know extended operations on bags

Relational Operations on Bags

We consider relations as bags (multi-sets) rather than sets, that is

- We allow the same tuples to appear more than once in a relation

We need to make changes to the definition of some relational operations

Relational Operations on Bags

Example

- In this figure, the tuple (1,2) appears 3-times

A	B
1	2
3	4
1	2
1	2

- As a set-valued relation, we would have to eliminate 2-occurences of the tuple (1,2)
- As a bag-valued relation, we allow multiple occurrences of the same tuple, but like sets, the order of tuples does not matter

Why Bags?

As mentioned, relations in commercial DBMS are implemented relations as bags rather than sets

Some relational operations are considerably more efficient if we use the bag model

- Union
- Projection

Why Bags?

Example

- Have a look at a projection on A and B
- What happens if we'd like to take the average of the A-components?
- What do you consider between treating as set-valued and treating as bag-valued relation?

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

Union, Intersection, and Difference of Bags

Suppose R and S are bags, and t is the tuple that appears n-times and m-times in R and S ($n \geq 0$, $m \geq 0$). Then we have:

- In $\{R \cup S\}$, *t* appears (n + m) times
- In $\{R \cap S\}$, *t* appears MIN(n, m) times
- In $\{R \setminus S\}$, *t* appears MAX(0, n - m) times

Union, Intersection, and Difference of Bags

Example

R

A	B
1	2
3	4
1	2
1	2

S

A	B
1	2
3	4
3	4
5	6

R∪S

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

R∩S

A	B
1	2
3	4

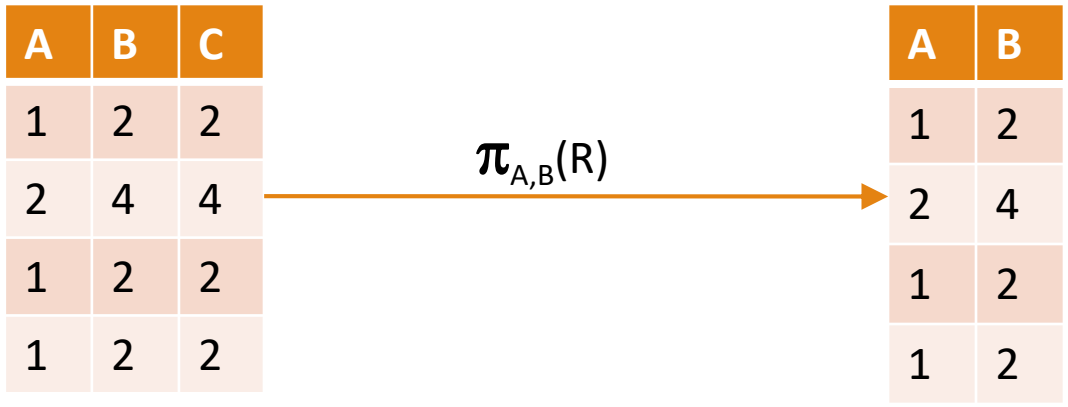
R\S

A	B
1	2
1	2

Projection of Bags

Each tuple is processed independently during the projection

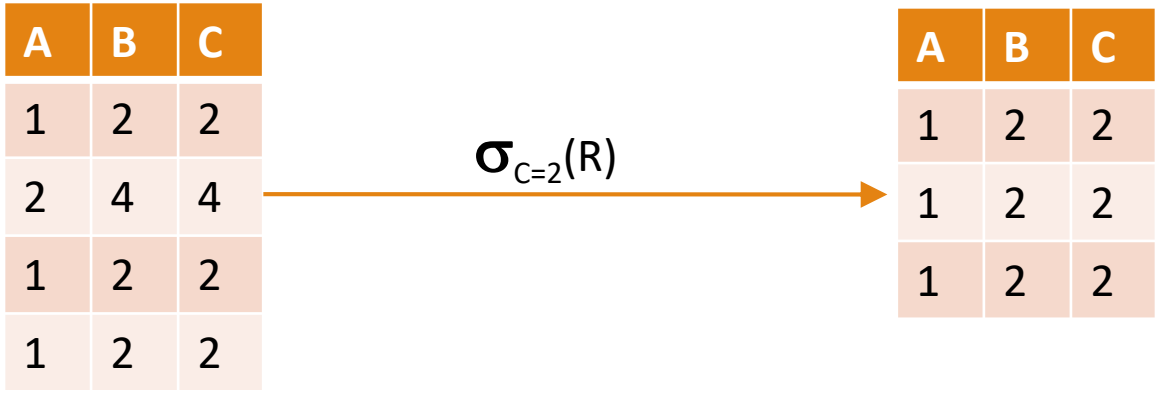
Don’t eliminate duplications



Selection on Bags

To apply a selection to a bag, we apply the selection condition to each tuple independently

Don’t eliminate duplications



Product of Bags

Each tuple of one relation is paired with each tuple of the other, regardless of whether it is a duplicate or not

If a tuple r appears in relation R m-times, and tuple s appears in relation S n-times, then in the product RxS, the tuple rs will appear mn-times

Product of Bags

Example

R

A	B
1	2
1	2

S

B	C
2	3
4	5
4	5

RxS

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

Joins of Bags

Each tuple of one relation is paired with each tuple of the other if they satisfy the join condition whether there are duplicates or not

Joins of Bags

Example

R

A	B
1	2
1	2

S

B	C
2	3
4	5
4	5

$R \bowtie_{R.B<S.B} S$

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

EXTENDED OPERATIONS OF RELATIONAL ALGEBRA

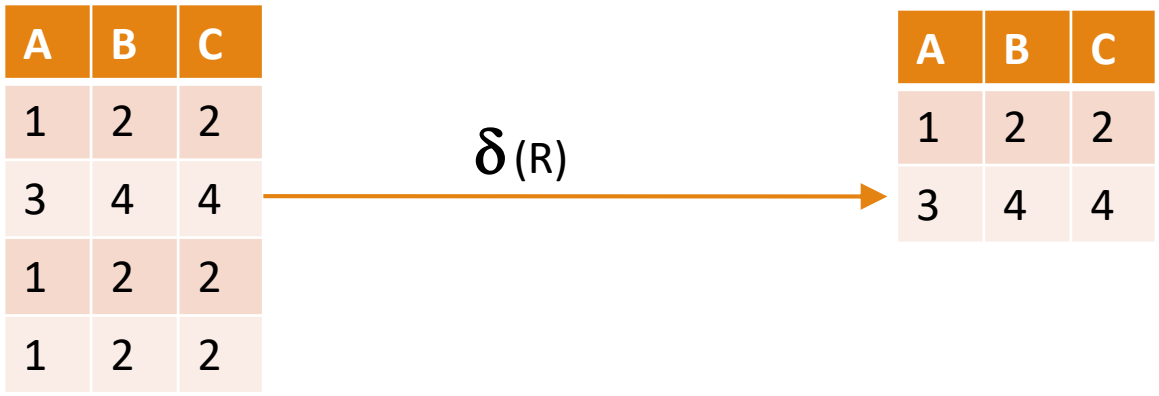
Extended Operators of Relational Algebra

There are some extended operators

- Duplicate-elimination operator δ
- Aggregation operators (used by the grouping operator)
- Grouping operators
- Extended projection
- Sorting operators
- Outer join operators

Duplicate Elimination

$\delta(R)$ is used to convert a bag to a set



Aggregation Operators

These operators are used to summarize or *aggregate* the values in one column of a relation

The standard aggregation operators

- SUM
- AVG
- MIN
- MAX
- COUNT

Aggregation Operators

Example

- SUM(B)=2+4+2+2=10
- AVG(A)=(1+3+1+1)/4=1.5
- MIN(A)=1
- MAX(B)=4
- COUNT(A)=4

A	B	C
1	2	2
3	4	4
1	2	2
1	2	2

How do we ...

Compute the total number of stars of each movie

title	*year*	*length*	*genre*	*studioName*	*starName*
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With The Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

- First, list all stars for each movie, and store in temporary relation (there are three such relations)
- Then, for each relation, use the COUNT aggregation operator to count the number of tuples

This approach is so complicated

So how we do?

title	*year*	*length*	*genre*	*studioName*	*starName*
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With The Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

First, we group the tuple of Movies

Then, we apply the aggregation COUNT to each group independently

That is, we use the grouping operator

Grouping and Grouping Operator

The grouping operator is denoted by $\gamma_L(R)$, where L is a list of elements:

- An attribute of the relation R to which the γ is applied, this attribute is one of the attributes by which R will be grouped, is called by *grouping attribute*
- An aggregation operator applied to an attribute of the relation, this attribute is said to be an *aggregating attribute*

The result of $\gamma_L(R)$ is constructed as follow:

- Partition the tuples of R into groups on *grouping attributes* in list L
- 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

Grouping and Grouping Operator

Example: Compute the total number of stars of each movie?

title	*year*	*length*	*genre*	*studioName*	*starName*
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With The Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

$\gamma_{title->movieTitle,\ year->movieYear,\ COUNT(starName)->totalStars}$(Movies)

movieTitle	movieYear	totalStars
Star Wars	1977	3
Gone With The Wine	1939	1
Wayne's World	1992	2

Extending the Projection Operator

An extended projection, denoted $\pi_L(R)$, L list can have the following kinds of elements:

- A single attribute of R
- An expression $x \rightarrow y$, where x, y are attributes, means that rename x attribute of R to y
- An expression $E \rightarrow z$, where E is an expression involving attributes of R, constants, arithmetic operators, string operators, and z is new attribute that results from E

Extending the Projection Operator

Example

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

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

A	X
0	3
0	3
3	9

The Sorting Operator

Sort the tuples of a relation by one or more of its attributes

$\tau_L(R)$, where R is a relation and L is a list of some of R's attributes

Example

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

$\tau_A(R)$

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

Outer joins

A property of the join is that it is possible for certain tuples that can match any tuple of the other relation in the common attributes

Example:

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

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

A	B	C	D
1	2	3	10
1	2	3	11

- What do you consider?

Natural Outer Joins

Natural Outer Joins, is denoted as $R \overset{O}{\bowtie} S$, is on equated values of all attributes in common to the two relations (like $R \bowtie S$), and adding any dangling tuples from R or S

The added tuples must be padded with a special *null* symbol in all the attributes that they do not possess but that appear in the join result

Natural Outer Joins

Example

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

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

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

Outer Joins

There are many variants of the basic outer join

- The left outer join $R \bowtie_L S$
- The right outer join $R \bowtie_R S$

Left, right, full outer join are also denoted that:

- Left outer join ($⟕$)
- Right outer join ($⟖$)
- Full Outer join ($⟗$)