

CSC 742
Database Management Systems

Topic #7:
Relational Algebra

Spring 2002

CSC 742: DBMS by Dr. Peng Ning

1

Relational Algebra

Offers operations defined at a high level

- input ≥ 1 relations; output 1 relation
- independent of implementation
- procedural
- nesting of an expression defines the *essential* order of operations independent of implementation
- intuitively, like a virtual machine for databases

Spring 2002

CSC 742: DBMS by Dr. Peng Ning

2

Achtung!

- Operations are
 - ◆ written in terms of R, S, etc. referring to schema names
 - ◆ applied to specific instances

Select Operation

- Denoted $\sigma_{\langle \text{selection condition } c \rangle}(\mathbf{R})$
 - ◆ Filters out tuples from r that don't match c
 - ◆ c is a boolean combination of comparison clauses
 - ◆ Terms in c are attributes or constants
 - ◆ c makes sense only for one tuple at a time

<u>SSN</u>	Lname	Fname	Salary
111-22-3333	Smith	John	30000
121-23-3333	Wong	Frank	45000
153-32-1342	Wallace	Jennifer	43000
154-33-3333	Borg	James	56000
555-44-5555	English	Joyce	53000

$\sigma_{\text{salary} > 50000}(\text{Salary})$

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 5

Achtung!

- σ is not the same as SELECT in SQL
- A series of σ 's can be reordered or joined into one
 - ◆ $\sigma_{\langle \text{cond1} \rangle}(\sigma_{\langle \text{cond2} \rangle}(\dots (\sigma_{\langle \text{condn} \rangle}(\mathbf{R}))\dots)) = \sigma_{\langle \text{cond1} \rangle \text{ AND } \langle \text{cond2} \rangle \text{ AND } \langle \text{condn} \rangle}(\mathbf{R}).$

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 6

Project

- Denoted $\pi_{\langle \text{attribute list} \rangle}(\mathbf{R})$
- Returns a relation that has
 - ◆ only the attributes in the list
 - ◆ all the tuples in r restricted to these attributes
 - ◆ except that duplicates are removed

Salary

<u>SSN</u>	Lname	Fname	Salary
111-22-3333	Smith	John	30000
121-23-3333	Wong	Frank	45000
153-32-1342	Wallace	Jennifer	43000
154-33-3333	Borg	James	56000
555-44-5555	English	Joyce	53000

$\pi_{\text{Lname, Fname, Salary}}(\text{Salary})$

$\pi_{\text{Lname, Fname}}(\pi_{\text{Lname, Fname, Salary}}(\text{Salary}))$

$\pi_{\text{Lname, Fname}}(\sigma_{\text{salary} > 50000}(\text{Salary}))$

Conventions

- Relations may be
 - ◆ temporary (named or anonymous)
 - ◆ predefined in the schema
- Attributes may be renamed
- Define a new relation
 - ◆ $A \leftarrow B$

Rename Operation

- Change the name of a relation, the names of attributes, or all of the above.
- $\rho_S(R)$
- $\rho_{(B_1, B_2, \dots, B_n)}(R)$
- $\rho_{S(B_1, B_2, \dots, B_n)}(R)$

<u>SSN</u>	Lname	Fname	Salary
111-22-3333	Smith	John	30000
121-23-3333	Wong	Frank	45000
153-32-1342	Wallace	Jennifer	43000
154-33-3333	Borg	James	56000
555-44-5555	English	Joyce	53000

■ $\rho_{\text{CurrentSalary}}(\text{Salary})$
 ■ $\rho_{(B1, B2, B3, B4)}(\text{Salary})$
 ■ $\rho_{\text{CurrentSalary}(B1, B2, B3, B4)}(\text{Salary})$

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 11

Set Operations

- Union
- Intersection
- Difference
- Apply only on "(union) compatible" arguments
 - ◆ same degree
 - ◆ same domains for each matching pair of attributes
- Attribute names come from first argument

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 12

<u>SSN</u>	Lname	Fname	Salary
111-22-3333	Smith	John	30000
121-23-3333	Wong	Frank	45000
153-32-1342	Wallace	Jennifer	43000

AccountingSalary

<u>SSN</u>	Lname	Fname	Salary
154-33-3333	Borg	James	56000
555-44-5555	English	Joyce	53000

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 13

Cartesian Product

- Also called cross product or cross join
- Notated $R \times S$
- All possible pairings
- All the attributes, renamed if necessary
- lots of tuples: $|r \times s| = |r| \times |s|$
- Not to be used except in defining or simplifying expressions

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 14

Employee			Dependent		
Fname	Lname	SSN	Fname	Lname	ESSN
Alice	Zelaya	999-88-7777	Eric	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333	Alex	Wallace	111-22-3333
Joyce	White	222-33-4444			
Employee × Dependent					
Fname	Lname	SSN	D_FName	D_Lname	D_ESSN
Alice	Zelaya	999-88-7777	Eric	Zelaya	999-88-7777
Alice	Zelaya	999-88-7777	Alex	Wallace	111-22-3333
Jennifer	Wallace	111-22-3333	Eric	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333	Alex	Wallace	111-22-3333
Joyce	White	222-33-4444	Eric	Zelaya	999-88-7777
Joyce	White	222-33-4444	Alex	Wallace	111-22-3333
Spring 2002					
CSC 742: DBMS by Dr. Peng Ning					
15					

<h1>Theta Join</h1>	
<ul style="list-style-type: none"> ■ $R \bowtie_{\text{condition}} S$ ■ Equivalent to a Cartesian product nested within a select <ul style="list-style-type: none"> ◆ $\sigma_{\text{condition}} (R \times S)$ ■ Tuples with null values for any attribute in condition are eliminated 	
Spring 2002	
CSC 742: DBMS by Dr. Peng Ning	
16	

Employee			Dependent		
Fname	Lname	SSN	Fname	Lname	ESSN
Alice	Zelaya	999-88-7777	Eric	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333	Alex	Wallace	111-22-3333
Joyce	White	222-33-4444			

Employee |><|_{SSN=ESSN} Dependent

Fname	Lname	SSN	D_FName	D_Lname	D_ESSN
Alice	Zelaya	999-88-7777	Eric	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333	Alex	Wallace	111-22-3333

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 17

Equijoin

- Theta join where only = may be used as the comparison operator in the join condition

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 18

Natural Join

- Denoted $R * S$ or $R \bowtie S$
- Basically an equijoin followed by removal of the superfluous attributes.
 - ◆ Equijoin results in pairs of attributes with identical values
- Natural join eliminates the second of each pair
 - ◆ Matching attributes must have the same name
 - ◆ Equality for all matching attributes is required

Employee

Fname	Lname	SSN
Alice	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333
Joyce	White	222-33-4444

Dependent

Fname	Lname	ESSN
Eric	Zelaya	999-88-7777
Alex	Wallace	111-22-3333

Employee \bowtie $\rho_{(DFname, DLname, SSN)}$ **Dependent**

Fname	Lname	SSN	DFName	DLname
Alice	Zelaya	999-88-7777	Eric	Zelaya
Jennifer	Wallace	111-22-3333	Alex	Wallace

Join

- The heart of the relational model
- Helps combine information from multiple tables
- Essential because there are no direct pointers or references in the model

Outer Join

- Includes tuples in result even if no match is found
- The result tuples are padded with nulls
- Variants: left, right, full

Employee			Dependent		
Fname	Lname	SSN	Fname	Lname	ESSN
Alice	Zelaya	999-88-7777	Eric	Zelaya	999-88-7777
Jennifer	Wallace	111-22-3333	Alex	Wallace	111-22-3333
Joyce	White	222-33-4444			

Employee $\bowtie_{\rho_{(DFname, DLname, SSN)}} \text{Dependent}$

Fname	Lname	SSN	DFName	DLname
Alice	Zelaya	999-88-7777	Eric	Zelaya
Jennifer	Wallace	111-22-3333	Alex	Wallace
Joyce	White	222-33-4444	NULL	NULL

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 23

Outer Union

- Applies on incompatible relations
- For attributes in only one input relation, tuples of the other are padded with nulls

Spring 2002 CSC 742: DBMS by Dr. Peng Ning 24

