Lecture 23

THE NORTHCAP UNIVERSITY

Relational Algebra

30 YEARS OF EXCELLENCE

NAAC Accredited

Union Operation

THE NORTHCAP UNIVERSITY

- The result of this operation, denoted by R ∪ S, is a relation that includes all tuples that are either in R or in S or in both R and S.
  - The operand relations $R_1(A_1, A_2, ..., A_n)$ and $R_2(B_1, B_2, ..., B_n)$ must have the **same number of attributes, and the domains of corresponding attributes must be compatible**; that is, $dom(A_i)=dom(B_i)$ for i=1, 2, ..., n.
- Duplicate tuples are automatically eliminated.

30 YEARS OF EXCELLENCE | NAAC Accredited

Union Operator Example

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

- To retrieve the social security numbers of all employees who either work in department 5 or directly supervise an employee who works in department 5.

Query:

$$\textbf{DEP5\_EMPS} \leftarrow \sigma_{\textbf{DNO=5}}\ \textbf{(EMPLOYEE)}$$

$$\textbf{RESULT1} \leftarrow \pi_{\textbf{SSN}}\textbf{(DEP5\_EMPS)}$$

$$\textbf{RESULT2(SSN)} \leftarrow \pi_{\textbf{SUPERSSN}}\textbf{(DEP5\_EMPS)}$$

$$\textbf{RESULT} \leftarrow \textbf{RESULT1} \cup \textbf{RESULT2}$$

- Another way to write the same query:

$$\textbf{Result} \leftarrow \pi_{\textbf{Ssn}} (\sigma_{\textbf{Dno=5}} (\textbf{EMPLOYEE})) \cup$$

$$\pi_{\textbf{Super\_ssn}} (\sigma_{\textbf{Dno=5}} (\textbf{EMPLOYEE}))$$

- **Output:**

RESULT1

Ssn
123456789
333445555
666884444
453453453

RESULT2

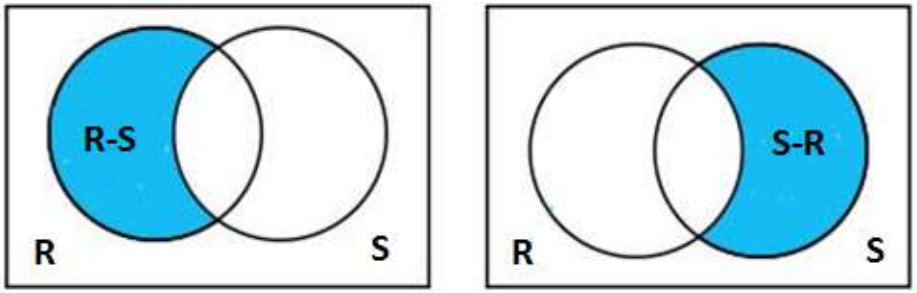
Ssn
333445555
888665555

RESULT

Ssn
123456789
333445555
666884444
453453453
888665555

Set Difference (or Minus) Operation

- The result of this operation, denoted by R - S, is a relation that includes all tuples that are in R but not in S.
- The two operands must be "type compatible".



Set Difference (or Minus) Operation Example

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Query:

$$PROJECT\_SUBSET \leftarrow \pi_{Pnumber,\ Plocation}(PROJECT)$$

$$Result \leftarrow PROJECT\_SUBSET - DEPT\_LOCATIONS$$

Output:

Pnumber	Plocation
1	Bellaire
2	Sugarland
3	Houston
10	Stafford
20	Houston
30	Stafford

JOIN Operation

- **Join** is a derived operator that uses a sequence of cartesian product followed by selection of related tuples from two relations and then projection of distinct attributes. It is denoted by a ⋈.
- This operation is very important for any relational database with more than a single relation, because it allows us to process relationships among relations.
- The general form of a join operation on two relations $R(A_1, A_2, \ldots, A_n)$ and $S(B_1, B_2, \ldots, B_m)$ is:

$$R \bowtie_{\text{<join condition>}} S$$

where R and S can be any relations that result from general *relational algebra expressions*.

JOIN Operation Example

THE NORTHCAP UNIVERSITY

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

30 YEARS OF EXCELLENCE

NAAC Accredited

JOIN Operation Example Cont.

- To retrieve the name of the manager of each department. To get the manager's name, we need to combine each DEPARTMENT tuple with the EMPLOYEE tuple whose SSN value matches the MGRSSN value in the department tuple.

Query:

$$\text{DEPT\_MGR} \leftarrow \text{DEPARTMENT} \bowtie_{\text{MGRSSN=SSN}} \text{EMPLOYEE}$$

Output:

DEPT_MGR

Dname	Dnumber	Mgr_ssn	...	Fname	Minit	Lname	Ssn	...
Research	5	333445555	...	Franklin	T	Wong	333445555	...
Administration	4	987654321	...	Jennifer	S	Wallace	987654321	...
Headquarters	1	888665555	...	James	E	Borg	888665555	...

JOIN Operation Example

$$EMP\_DEPENDENTS \leftarrow EMPNAMES \times DEPENDENT$$

$$ACTUAL\_DEPENDENTS \leftarrow \sigma_{Ssn=Essn}(EMP\_DEPENDENTS)$$

= = = = = = = = = =

$$ACTUAL\_DEPENDENTS \leftarrow EMPNAMES \bowtie_{Ssn=Essn} DEPENDENT$$

Variations of Join Operation (Inner Joins)

- **EQUIJOIN Operation**
  - The most common use of join involves join conditions with equality comparisons only. Such a join, where the only comparison operator used is =, is called an EQUIJOIN.
  - In the result of an EQUIJOIN we always have one or more pairs of attributes (whose names need not be identical) that have *identical values* in every tuple.

- **NATURAL JOIN Operation**
  - Because one of each pair of attributes with identical values is superfluous, a new operation called natural join—denoted by *—was created to get rid of the second (superfluous) attribute in an EQUIJOIN condition.
  - Natural join requires that the two join attributes, or each pair of corresponding join attributes, have the **same name** in both relations. If this is not the case, a renaming operation is applied first.

Natural Join Example

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Query:

DEPT_LOCS ← DEPARTMENT * DEPT_LOCATIONS

Output:

DEPT_LOCS

Dname	Dnumber	Mgr_ssn	Mgr_start_date	Location
Headquarters	1	888665555	1981-06-19	Houston
Administration	4	987654321	1995-01-01	Stafford
Research	5	333445555	1988-05-22	Bellaire
Research	5	333445555	1988-05-22	Sugarland
Research	5	333445555	1988-05-22	Houston

Inner Joins

- **Note:** NATURAL JOIN is basically an EQUIJOIN followed by the removal of the superfluous attributes.

- The NATURAL JOIN or EQUIJOIN operation can also be specified among multiple tables, leading to an *n-way join*

Example:

$$((\text{PROJECT} \bowtie_{\text{Dnum=Dnumber}} \text{DEPARTMENT}) \bowtie_{\text{Mgr\_ssn=Ssn}} \text{EMPLOYEE})$$

Each project tuple is combined with its controlling department tuple into a single tuple, and then that tuple is combined with an employee tuple that is the department manager. The net result is a consolidated relation in which each tuple contains this project-department-manager combined information.

Relational Algebra Operators

Operation	Purpose	Notation
SELECT	Selects all tuples that satisfy the selection condition from a relation *R*.	σ<selection condition>(*R*)
PROJECT	Produces a new relation with only some	π<attribute list>(*R*)
CARTESIAN/ CROSS PRODUCT	Produces a relation that has the attributes of *R*1 and *R*2 and includes as tuples all possible combinations of tuples from *R*1 and *R*2.	*R*1 × *R*2
UNION	Produces a relation that includes all the tuples in *R*1 or *R*2 or both *R*1 and *R*2; *R*1 and *R*2 must be union compatible.	*R*1 ∪ *R*2
SET-DIFFERENCE/ MINUS	Produces a relation that includes all the tuples in *R*1 that are not in *R*2; *R*1 and *R*2 must be union compatible.	*R*1 – *R*2

Relational Algebra Operators Cont.

Operation	Purpose	Notation
EQUIJOIN	Produces all the combinations of tuples from *R*1 and *R*2 that satisfy a join condition with only equality comparisons.	*R*1 $\bowtie_{\text{<join condition>}}$ *R*2, OR *R*1 $\bowtie_{\text{(<join attributes 1>), (<join attributes 2>)}}$ *R*2
NATURAL JOIN	Same as EQUIJOIN except that the join attributes of *R*2 are not included in the resulting relation; if the join attributes have the same names, they do not have to be specified at all.	*R*1 $*_{\text{<join condition>}}$ *R*2, OR *R*1 $*_{\text{(<join attributes 1>), (<join attributes 2>)}}$ *R*2 OR *R*1 * *R*2
THETA JOIN	Produces all combinations of tuples from *R*1 and *R*2 that satisfy the join condition.	*R*1 $\bowtie_{\text{<join condition>}}$ *R*2

THE NORTHCAP UNIVERSITY

30 YEARS OF EXCELLENCE

NAAC Accredited

Question

1. **Cross Product in relation algebra is a**

a) Unary Operator

b) Binary Operator

c) Ternary Operator

d) Not defined

2. **Relational Algebra is a _____ query language that takes two relation as input and produces another relation as output of the query.**

a) Relational

b) Procedural

c) Structural

d) Fundamental

Answer

1. Answer: B

2. Answer: B

Question

3. Which of the following is used to denote the selection operation in relational algebra?

a) π

b) σ

c) α

d) ρ

4. A relational operator that yields values from all rows found in a table is known as the ____ operator.

a) Selection

b) Projection

c) Difference

d) Cross-product

Answer

3. Answer: A

4. Answer: A

Question

5. If relation A has 6 rows and 4 attributes and relation B has 3 rows and 2 attributes and the cartesian product operation was carried out on relation A and relation B, What would be the cardinaility of the new relation C where C = A X B?

a) 12

b) 6

c) 24

d) 18

Solution

5. Answer: D

THE NORTHCAP UNIVERSITY

Thanks!!

30 YEARS OF EXCELLENCE

NAAC Accredited