Lecture 3: Relational Algebra

CS3402 Database Systems

What is Relational Algebra?

- Relational algebra is a formal language for the relational model
- The operations in relational algebra enable a user to specify basic retrieval requests (i.e., queries)
- Relational algebra consists of a set of operations on relations to generate relations
- The result of an operation is a new relation that can be further manipulated using operations
- A sequence of relational algebra operations forms a relational algebra expression

Importance of Relational Algebra

- Relational algebra provides a formal foundation for relational model
- It is used as a basis for implementing and optimizing queries in query processing and optimization
- Its concepts are incorporated into SQL standard language for relational database management systems
- The internal modules of most commercial RDBMS are based on relational algebra

Relational Algebra Overview (1/2)

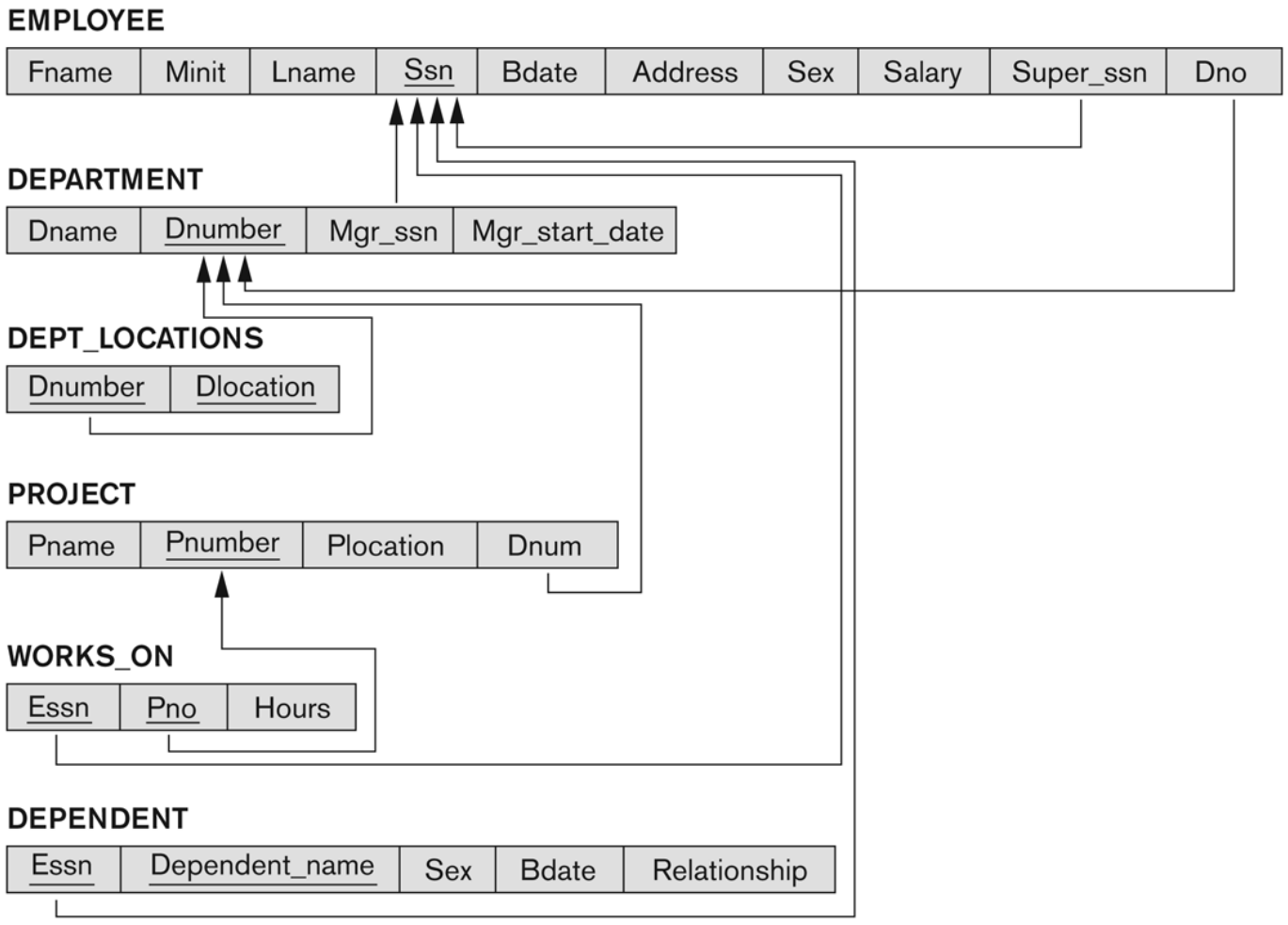
- Relational algebra consists of several groups of operations
- Unary relational operations
  - SELECT (denoted by σ (pronounced as “sigma”))
  - PROJECT (π (“pi”))
  - RENAME (ρ (“rho”))
- Relational algebra operations from set theory
  - UNION (∪), INTERSECTION (∩), DIFFERENCE (or MINUS, –)
  - CARTESIAN PRODUCT (**x**)

Relational Algebra Overview (2/2)

- Binary relational operations
  - JOIN (THETA JOIN, EQUIJOIN and NATURAL JOIN)
  - DIVISION
- Additional relational operations
  - OUTER JOINS, OUTER UNION
  - AGGREGATE FUNCTIONS (These compute summary of information: for example, SUM, COUNT, AVG, MIN, and MAX)

Relational Model of COMPANY Database

- The COMPANY database consists of 6 relations.
- In each relation, underlined attribute(s) is its primary key and each arrow indicates a foreign key (i.e., the primary key in another relation)



Database State for COMPANY

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

DEPT_LOCATIONS

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

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

PROJECT

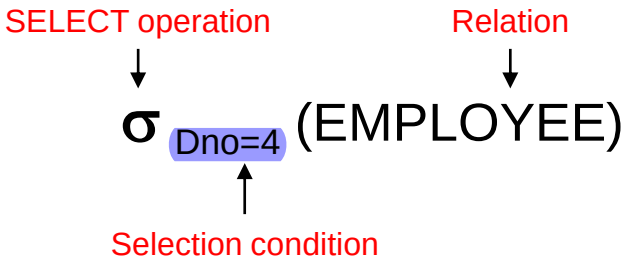
Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

SELECT Operation (1/3)

- The SELECT operation, denoted by σ (sigma), is used to select a subset of the tuples from a relation based on a selection condition
  - The selection condition acts as a filter to keep only those tuples that satisfy the qualifying condition
- Example 1
  - Select the EMPLOYEE tuples whose department number is 4:



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

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
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
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4

The result relation of the SELECT operation

SELECT Operation (2/3)

- Example 2
  - Select the employee tuples whose salary is greater than $40,000:

$$\sigma_{\text{Salary}>40,000}(\text{EMPLOYEE})$$

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

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

The result relation of the SELECT operation

SELECT Operation (3/3)

- Example 3
  - Select the employee tuples whose department number is 4 and salary is greater than $25,000 or department number is 5 and salary is greater than $30,000

$$\sigma_{(Dno=4\ \textbf{AND}\ Salary>25{,}000)\ \textbf{OR}\ (Dno=5\ \textbf{AND}\ Salary>30{,}000)}(EMPLOYEE)$$

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

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
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

The result relation of the SELECT operation

SELECT Operation Properties (1/2)

- The SELECT operation $\sigma_{<\text{selection condition}>}(R)$ produces a relation S that has the same schema (i.e., same attributes) as R
- SELECT σ is commutative:
  - $\sigma_{<\text{cond1}>}(\sigma_{<\text{cond2}>}(R)) = \sigma_{<\text{cond2}>}(\sigma_{<\text{cond1}>}(R))$
- Because of commutativity property, a cascade (sequence) of SELECT operations may be applied in any order:
  - $\sigma_{<\text{cond1}>}(\sigma_{<\text{cond2}>}(\sigma_{<\text{cond3}>}(R)) = \sigma_{<\text{cond2}>}(\sigma_{<\text{cond3}>}(\sigma_{<\text{cond1}>}(R)))$

SELECT Operation Properties (2/2)

- A cascade of SELECT operations may be replaced by a single selection with a conjunction (and) of all the conditions:
  - $\sigma_{<cond1>}(\sigma_{<cond2>}(\sigma_{<cond3>}(R))) = \sigma_{<cond1> \text{ AND } <cond2> \text{ AND } <cond3>}(R)$

- The number of tuples in the result of a SELECT operation is less than (or equal to) the number of tuples in the input relation R
- The fraction of tuples selected by a selection condition is called the selectivity of the condition

Unary Relational Operations: PROJECT (1/3)

- PROJECT Operation is denoted by π (pi)
- This operation keeps certain attributes from a relation and discards the other attributes
  - The list of specified attributes is kept in each tuple and the other attributes in each tuple are discarded
- The general form of the project operation is:

$$\pi_{<\text{attribute list}>}(R)$$

  - π is the symbol used to represent the project operation
  - <attribute list> is the desired list of attributes from relation R

Unary Relational Operations: PROJECT (2/3)

- The project operation removes any duplicate tuples
  - This is because the result of the project operation must be a set of tuples
  - Mathematical sets do not allow duplicate elements
- Example 1
  - Retrieve the first name, last name and salary of each employee

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

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

Lname	Fname	Salary
Smith	John	30000
Wong	Franklin	40000
Zelaya	Alicia	25000
Wallace	Jennifer	43000
Narayan	Ramesh	38000
English	Joyce	25000
Jabbar	Ahmad	25000
Borg	James	55000

The result relation of the PROJECT operation

Unary Relational Operations: PROJECT (3/3)

- Example 2
  - Retrieve the sex and salary of each employee

    $\pi_{Sex, Salary}$ (EMPLOYEE)

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

Duplicate tuple

Sex	Salary
M	30000
M	40000
F	25000
F	43000
M	38000
M	25000
M	55000

The result relation of the PROJECT operation

PROJECT Operation Properties

- The number of tuples in the result of projection $\pi_{<list>}(R)$ is always less (duplicates are removed) or equal (unique values) to the number of tuples in R
- If the list of attributes includes a key of R, then the number of tuples in the result of PROJECT is equal to the number of tuples in R
- PROJECT is not commutative
  - $\pi_{<list1>}(\pi_{<list2>}(R)) \neq \pi_{<list2>}(\pi_{<list1>}(R))$
  - $\pi_{<list1>}(\pi_{<list2>}(R)) = \pi_{<list1>}(R)$ if <list2> contains the attributes in <list1>, for example, list1 = Fname, Lname; list2 = Fname, Lname, Salary

Relational Algebra Expressions (1/2)

- We may want to apply several relational algebra operations one after the other
  - Either we can write the operations as a single relational algebra expression by nesting the operations, or
  - We can apply one operation at a time and create intermediate result relations
- In the latter case, we must give names (rename) to the relations that hold the intermediate results

Relational Algebra Expressions (2/2)

- To retrieve the first name, last name, and salary of all employees who work in department number 5, we must apply a SELECT and a PROJECT
- We can write a single relational algebra expression as follows:
  - $\pi_{\text{Fname, Lname, Salary}}(\sigma_{\text{DNO=5}}(\text{EMPLOYEE}))$
- Or, we can explicitly show the sequence of operations, giving a name to each intermediate relation:
  - $\text{DEP5\_EMPS} \leftarrow \sigma_{\text{Dno=5}}(\text{EMPLOYEE})$
  - $\text{RESULT} \leftarrow \pi_{\text{Fname, Lname, Salary}}(\text{DEP5\_EMPS})$

RENAME Operation (1/2)

- The RENAME operation is denoted by ρ (rho)
- The general RENAME operation ρ can be expressed by any of the following forms:
  - $\rho_S(R)$ changes
    - The relation name only to S
  - $\rho_{(B1, B2, \ldots, Bn)}(R)$ changes
    - The attribute names only to B1, B2, …, Bn
  - $\rho_{S(B1, B2, \ldots, Bn)}(R)$ changes both
    - The relation name to S, and
    - The attribute names to B1, B2, …, Bn

RENAME Operation (2/2)

- For convenience, we also use a shorthand for renaming attributes in an intermediate relation:
  - If we write:
    - $TEMP \leftarrow \pi_{Fname, Lname, Salary}$ (EMPLOYEE)
    - TEMP will have the same attribute names as EMPLOYEE
  - If we write:
    - $\rho_{R(First\_name, Last\_name, Salary)}$ (TEMP)
    - The 3 attributes of TEMP are renamed to First_name, Last_name and Salary, respectively; and R is the name of the result relation
  - Note: the $\leftarrow$ symbol is an assignment operator

Example of Using Intermediate Relations and Renaming of Attributes

- TEMP $\leftarrow \sigma_{Dno=5}$ (EMPLOYEE)
- $\rho_{R\ (First\_name,\ Last\_name,\ Salary)}$ ($\pi_{Fname,\ Lname,\ Salary}$ (TEMP))

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

TEMP

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
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

R

First_name	Last_name	Salary
John	Smith	30000
Franklin	Wong	40000
Ramesh	Narayan	38000
Joyce	English	25000

Set Theory: UNION Operation (1/2)

- Binary operation, denoted by ∪
- The result of R ∪ S, is a relation that includes all tuples that are either in R or in S or in both R and S
- The two relations R and S must be "type compatible" (or UNION compatible)
  - R and S must have same number of attributes
  - Each pair of corresponding attributes must be type compatible (have same or compatible domains)
- The result of R ∪ S has the same attribute names as the first relation R
- Duplicate tuples are eliminated

Set Theory: UNION Operation (2/2)

- To retrieve the SSN of all employees who either (i) work in department with Dno=5 (RESULT1) or (ii) directly supervise an employee who works in department with Dno=5 (RESULT2)
- We can use the UNION operation as follows:

$$\text{DEP5\_EMPS} \leftarrow \sigma_{\text{Dno=5}}(\text{EMPLOYEE})$$

$$\text{RESULT1} \leftarrow \pi_{\text{Ssn}}(\text{DEP5\_EMPS})$$

$$\text{RESULT2} \leftarrow \rho_{(\text{Ssn})}(\pi_{\text{Super\_ssn}}(\text{DEP5\_EMPS}))$$

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

- The union operation produces the tuples that are in RESULT1, RESULT2 or both (one duplicate Ssn 333445555 is eliminated)

RESULT1

Ssn
123456789
333445555
666884444
453453453

RESULT2

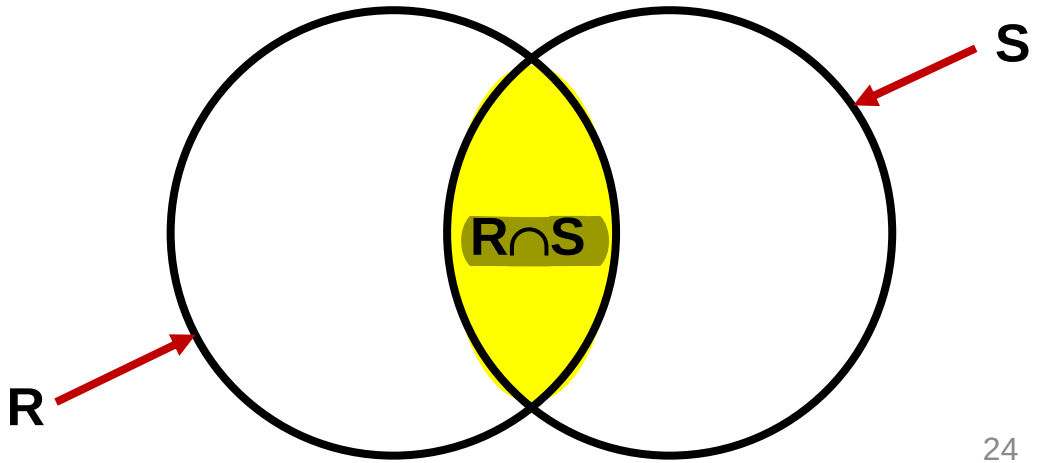
Ssn
333445555
888665555

RESULT

Ssn
123456789
333445555
666884444
453453453
888665555

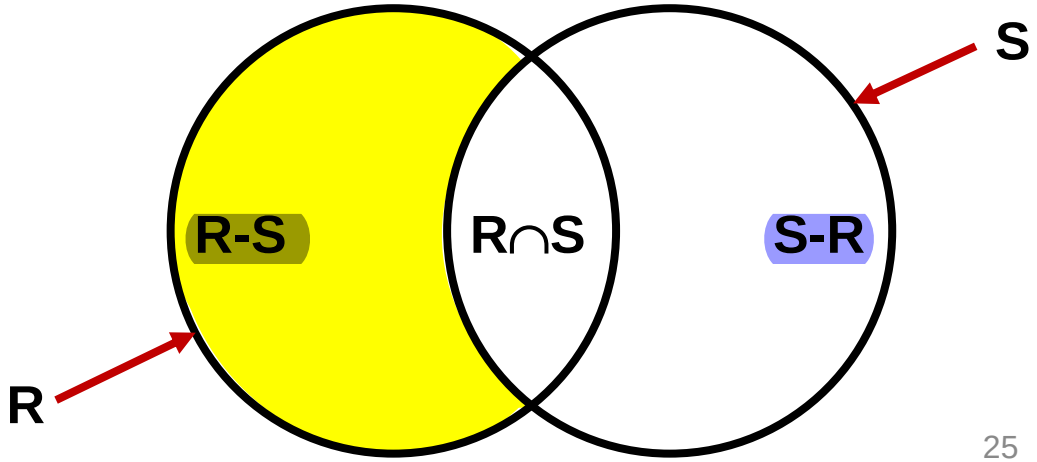
Set Theory: INTERSECTION Operation

- INTERSECTION is denoted by ∩
- The result of the operation R ∩ S, is a relation that includes all tuples that are in both R and S
- The two relations R and S must be “type compatible”
- The result of R ∩ S has the same attribute names as the first relation R
- Duplicate tuples are eliminated



Set Theory: DIFFERENCE Operation

- SET DIFFERENCE (also called MINUS or EXCEPT) is denoted by –
- The result of R – S, is a relation that includes all tuples that are in R but not in S
- The two relations R and S must be “type compatible”
- The result of R – S has the same attribute names as the first relation R



Set Theory: Examples (1/4)

- Given two type compatible relations

STUDENT

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

7 rows

INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

5 rows

- STUDENT ∪ INSTRUCTOR

Eliminate these duplicate tuples

Result

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert
John	Smith
Ricardo	Browne
Francis	Johnson

10 rows

Set Theory: Examples (2/4)

- Given two type compatible relations
- STUDENT ∩ INSTRUCTOR

STUDENT

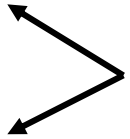
Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

7 rows

INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

5 rows



Keep these common tuples

Result

Fn	Ln
Susan	Yao
Ramesh	Shah

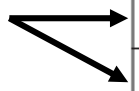
2 rows

Set Theory: Examples (3/4)

- Given two type compatible relations

- STUDENT – INSTRUCTOR

Eliminate these two tuples because they appear in INSTRUCTOR



STUDENT

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

7 rows

INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

5 rows

Result

Fn	Ln
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

5 rows

Set Theory: Examples (4/4)

- Given two type compatible relations

- INSTRUCTOR – STUDENT

STUDENT

Fn	Ln
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

7 rows

INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

5 rows

Eliminate these two tuples because they appear in STUDENT

Result

Fname	Lname
John	Smith
Ricardo	Browne
Francis	Johnson

3 rows

Some Properties of UNION, INTERSECTION, and DIFFERENCE

- Notice that both union and intersection are commutative operations
  - $R \cup S = S \cup R$, and $R \cap S = S \cap R$
- Both union and intersection can be treated as n-ary operations applicable to any number of relations as both are associative operations
  - $R \cup (S \cup T) = (R \cup S) \cup T$
  - $(R \cap S) \cap T = R \cap (S \cap T)$
- The minus operation is not commutative
  - $R - S \neq S - R$

CARTESIAN (or CROSS) PRODUCT (1/7)

- This operation is used to combine tuples from two relations in a combinatorial fashion
- Denoted by $R(A_1, A_2, \ldots, A_n) \times S(B_1, B_2, \ldots, B_m)$, where $A_1, A_2, \ldots, A_n$ are the attributes in R and $B_1, B_2, \ldots, B_m$ are the attributes in S
- Result is a relation Q with degree n + m attributes:
  - $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$, in that order
- The result relation state has one tuple for each combination of tuples - one from R and one from S
- Hence, if R has n_R tuples (denoted as $|R| = n_R$), and S has n_S tuples, then R x S will have $n_R \cdot n_S$ tuples
- The two operands do NOT have to be "type compatible"

CARTESIAN (or CROSS) PRODUCT (2/7)

- Generally, CROSS PRODUCT is not a meaningful operation
  - Some relations do not exist in the world
  - Can become meaningful when followed by other operations
- Example (not meaningful):
  - FEMALE_EMPS ← $\sigma_{Sex='F'}$ (EMPLOYEE)
  - EMPNAMES ← $\pi_{Fname, Lname, Ssn}$ (FEMALE_EMPS)
  - EMP_DEPENDENTS ← EMPNAMES x DEPENDENT
- EMP_DEPENDENTS will contain every combination of EMPNAMES and DEPENDENT no matter whether they are actually related

CARTESIAN (or CROSS) PRODUCT (3/7)

- To keep only combinations where the DEPENDENT is related to the EMPLOYEE, we add a SELECT operation as follows
- Example (meaningful):
  1. FEMALE_EMPS ← $\sigma_{\text{Sex='F'}}$ (EMPLOYEE)
  2. EMPNAMES ← $\pi_{\text{Fname, Lname, Ssn}}$ (FEMALE_EMPS)
  3. EMP_DEPENDENTS ← EMPNAMES x DEPENDENT
  4. ACTUAL_DEPS ← $\sigma_{\text{Ssn=Essn}}$ (EMP_DEPENDENTS)
  5. RESULT ← $\pi_{\text{Fname, Lname, Dependent_name}}$ (ACTUAL_DEPS)
- RESULT will now contain the name of female employees and their dependents

CARTESIAN (or CROSS) PRODUCT (4/7)

1

FEMALE_EMPS

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Alicia	J	Zelaya	999887777	1968-07-19	3321Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291Berry, Bellaire, TX	F	43000	888665555	4
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5

2

EMPNAMES

Fname	Lname	Ssn
Alicia	Zelaya	999887777
Jennifer	Wallace	987654321
Joyce	English	453453453

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

3

EMP_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	...
Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	...
Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	...
Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	...
Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	...
Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	...
Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	...
Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	...
Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	...
Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...
Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	...
Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	...
Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	...
Joyce	English	453453453	333445555	Alice	F	1986-04-05	...
Joyce	English	453453453	333445555	Theodore	M	1983-10-25	...
Joyce	English	453453453	333445555	Joy	F	1958-05-03	...
Joyce	English	453453453	987654321	Abner	M	1942-02-28	...
Joyce	English	453453453	123456789	Michael	M	1988-01-04	...
Joyce	English	453453453	123456789	Alice	F	1988-12-30	...
Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	...

CARTESIAN (or CROSS) PRODUCT (5/7)

1

FEMALE_EMPS

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Alicia	J	Zelaya	999887777	1968-07-19	3321Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291Berry, Bellaire, TX	F	43000	888665555	4
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5

2

EMPNAMES

Fname	Lname	Ssn
Alicia	Zelaya	999887777
Jennifer	Wallace	987654321
Joyce	English	453453453

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

3

EMP_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	...
Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	...
Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	...
Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	...
Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	...
Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	...
Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	...
Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	...
Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	...
Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...
Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	...
Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	...
Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	...
Joyce	English	453453453	333445555	Alice	F	1986-04-05	...
Joyce	English	453453453	333445555	Theodore	M	1983-10-25	...
Joyce	English	453453453	333445555	Joy	F	1958-05-03	...
Joyce	English	453453453	987654321	Abner	M	1942-02-28	...
Joyce	English	453453453	123456789	Michael	M	1988-01-04	...
Joyce	English	453453453	123456789	Alice	F	1988-12-30	...
Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	...

CARTESIAN (or CROSS) PRODUCT (6/7)

1

FEMALE_EMPS

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Alicia	J	Zelaya	999887777	1968-07-19	3321Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291Berry, Bellaire, TX	F	43000	888665555	4
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5

2

EMPNAMES

Fname	Lname	Ssn
Alicia	Zelaya	999887777
Jennifer	Wallace	987654321
Joyce	English	453453453

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

3

EMP_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	...
Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	...
Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	...
Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	...
Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	...
Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	...
Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	...
Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	...
Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	...
Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...
Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	...
Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	...
Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	...
Joyce	English	453453453	333445555	Alice	F	1986-04-05	...
Joyce	English	453453453	333445555	Theodore	M	1983-10-25	...
Joyce	English	453453453	333445555	Joy	F	1958-05-03	...
Joyce	English	453453453	987654321	Abner	M	1942-02-28	...
Joyce	English	453453453	123456789	Michael	M	1988-01-04	...
Joyce	English	453453453	123456789	Alice	F	1988-12-30	...
Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	...

CARTESIAN (or CROSS) PRODUCT (7/7)

1

FEMALE_EMPS

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Alicia	J	Zelaya	999887777	1968-07-19	3321Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291Berry, Bellaire, TX	F	43000	888665555	4
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5

2

EMPNAMES

Fname	Lname	Ssn
Alicia	Zelaya	999887777
Jennifer	Wallace	987654321
Joyce	English	453453453

3

EMP_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	...
Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	...
Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	...
Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	...
Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	...
Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	...
Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	...
Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	...
Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	...
Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...
Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	...
Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	...
Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	...
Joyce	English	453453453	333445555	Alice	F	1986-04-05	...
Joyce	English	453453453	333445555	Theodore	M	1983-10-25	...
Joyce	English	453453453	333445555	Joy	F	1958-05-03	...
Joyce	English	453453453	987654321	Abner	M	1942-02-28	...
Joyce	English	453453453	123456789	Michael	M	1988-01-04	...
Joyce	English	453453453	123456789	Alice	F	1988-12-30	...
Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	...

4

ACTUAL_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...

RESULT

Fname	Lname	Dependent_name
Jennifer	Wallace	Abner

5

JOIN Operation (1/2)

- JOIN operation (denoted by ⋈)
- The sequence of CARTESIAN PRODUCT followed by SELECT is used quite commonly to identify and select related tuples from two relations
- A special operation, called JOIN combines this sequence into a single operation
- The general form of a join operation on two relations $R(A_1, A_2, . . ., A_n)$ and $S(B_1, B_2, . . ., B_m)$ is:

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

- R and S can be any relations that result from general relational algebra expressions

JOIN Operation (2/2)

- Example: 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 Mgr_ssn value in the department tuple.
  - DEPT_MGR $\leftarrow$ DEPARTMENT $\bowtie_{\text{Mgr\_ssn=Ssn}}$ EMPLOYEE
- Mgr_ssn=Ssn is the join condition that combine each department record with the employee who manages the department

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	...

Some Properties of JOIN Operation

- Consider the following JOIN operation:
  - $Q \leftarrow R(A_1, A_2, \ldots, A_n) \bowtie_{R.A_i=S.B_j} S(B_1, B_2, \ldots, B_m)$
  - Result is a relation Q with degree n + m attributes, i.e., $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$
  - The result relation has one tuple for each combination of tuples: r from R and s from S, but only if they satisfy the join condition $r[A_i]=s[B_j]$
  - Hence, if R has n_R tuples, and S has n_S tuples, then the join result will generally have less than $n_R \times n_S$ tuples

THETA JOIN Operation

- A general join condition is of the form

  $R \bowtie_{\text{<condition> AND <condition> AND ... AND <condition>}} S$

  where each <condition> is of the form $A_i \; \theta \; B_j$, A_i is an attribute of R, B_j is an attribute of S, A_i and B_j have the same domain, and θ (theta) is one of the comparison operators {=, <, ≤, >, ≥, ≠}.
- A JOIN operation with such a general join condition is called a THETA JOIN.
- Tuples whose join attributes are NULL or for which the join condition is FALSE do not appear in the result.

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. For example,
  - $\text{DEPT\_MGR} \leftarrow \text{DEPARTMENT} \bowtie_{\text{Mgr\_ssn=Ssn}} \text{EMPLOYEE}$
- The result of an EQUIJOIN we always have one or more pairs of attributes that have identical values in every tuple.

NATURAL JOIN Operation (1/3)

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

NATURAL JOIN Operation (2/3)

- Example: Suppose we want to combine each PROJECT tuple with the DEPARTMENT tuple that controls the project.
- We first rename the Dnumber attribute of DEPARTMENT to Dnum, so that it has the same name as the Dnum attribute in PROJECT, and then we apply NATURAL JOIN.
  - DEPT ← ρ(Dname, Dnum, Mgr_ssn, Mgr_start_date)(DEPARTMENT)
  - PROJ_DEPT ← PROJECT * DEPT
- The attribute Dnum is called the join attribute for NATURAL JOIN, because it is the only attribute with the same name in both relations.
- In the PROJ_DEPT relation, each tuple combines a PROJECT tuple with the DEPARTMENT tuple for the department that controls the project, **but only one join attribute value is kept**.

NATURAL JOIN Operation (3/3)

DEPT ← ρ(Dname, Dnum, Mgr_ssn, Mgr_start_date)(DEPARTMENT)

PROJ_DEPT ← PROJECT * DEPT

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

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

*

DEPT

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

PROJ_DEPT

Pname	Pnumber	Plocation	Dnum	Dname	Mgr_ssn	Mgr_start_date
ProductX	1	Bellaire	5	Research	333445555	1988-05-22
ProductY	2	Sugarland	5	Research	333445555	1988-05-22
ProductZ	3	Houston	5	Research	333445555	1988-05-22
Computerization	10	Stafford	4	Administration	987654321	1995-01-01
Reorganization	20	Houston	1	Headquarters	888665555	1981-06-19
Newbenefits	30	Stafford	4	Administration	987654321	1995-01-01

DIVISION Operation (1/3)

- The DIVISION operation, denoted by ÷
- In general, the DIVISION operation is applied to two relations R(Z) ÷ S(X), where the attributes of S are a subset of the attributes of R; that is, $X \subseteq Z$.
- Let Y be the set of attributes of R that are not attributes of S; that is, $Y = Z - X$ (and hence $Z = X \cup Y$).
- The result of DIVISION is a relation T(Y) that includes a tuple t if tuples t_R appear in R with $t_R[Y] = t$, and with $t_R[X] = t_S$ for every tuple t_S in S. This means that, for a tuple t to appear in the result T of the DIVISION, the values in t must appear in R in combination with every tuple in S.

DIVISION Operation (2/3)

- Example, retrieve the Social Security numbers of employees who work on all the projects that ‘John Smith’ works on
- First, retrieve the list of project numbers that ‘John Smith’ works on in the intermediate relation SMITH_PNOS:
  - SMITH $\leftarrow \sigma_{\text{Fname='John' AND Lname='Smith'}}$ (EMPLOYEE)
  - SMITH_PNOS $\leftarrow \pi_{\text{Pno}}$ (WORKS_ON $\bowtie_{\text{Essn=Ssn}}$ SMITH)
- Next, create a relation that includes a tuple <Essn, Pno> whenever the employee whose Ssn is Essn works on the project whose number is Pno in the intermediate relation SSN_PNOS:
  - SSN_PNOS $\leftarrow \pi_{\text{Essn, Pno}}$ (WORKS_ON)

DIVISION Operation (3/3)

- Finally, apply the DIVISION operation to the two relations, which gives the desired employees' Social Security numbers:
  - SSNS(Ssn) ← SSN_PNOS ÷ SMITH_PNOS
- Y = {Essn}, X ={Pno}
- E.g., 123456789 is in SSNS because tuples t_R appear in R with $t_R[Y]$=123456789, and with $t_R[X]$=tS for every tuple t_S (i.e., 1 and 2) in S

R{Essn , Pno}

SSN_PNOS

Essn	Pno
123456789	1
123456789	2
666884444	3
453453453	1
453453453	2
333445555	2
333445555	3
333445555	10
333445555	20
999887777	30
999887777	10
987987987	10
987987987	30
987654321	30
987654321	20
888665555	20

S{Pno}

SMITH_PNOS

Pno
1
2

SSNS

Ssn
123456789
453453453

Operations of Relational Algebra (1/2)

OPERATION	PURPOSE	NOTATION
SELECT	Selects all tuples that satisfy the selection condition from a relation R.	$\sigma_{\text{<selection condition>}}(R)$
PROJECT	Produces a new relation with only some of the attributes of R, and removes duplicate tuples.	$\pi_{\text{<attribute list>}}(R)$
THETA JOIN	Produces all combinations of tuples from R_1 and R_2 that satisfy the join condition.	$R_1 \bowtie_{\text{<join condition>}} R_2$
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 $R_1 * R_2$

Operations of Relational Algebra (2/2)

OPERATION	PURPOSE	NOTATION
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.	$R_1 \cup R_2$
INTERSECTION	Produces a relation that includes all the tuples in both R_1 and R_2; R_1 and R_2 must be union compatible.	$R_1 \cap R_2$
DIFFERENCE	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.	$R_1 - R_2$
CARTESIAN 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.	$R_1 \times R_2$
DIVISION	Produces a relation $R(X)$ that includes all tuples $t[X]$ in $R_1(Z)$ that appear in R_1 in combination with every tuple from $R_2(Y)$, where $Z = X \cup Y$.	$R_1(Z) \div R_2(Y)$