The Relational Algebra and Relational Calculus

Logical Equivalence of RA Plans

- Given relations R(A,B) and S(B,C):
  - Here, projection & selection commute:
    - $\sigma_{A=5}(\Pi_A(R)) \stackrel{?}{=} \Pi_A(\sigma_{A=5}(R))$
  - What about here?
    - $\sigma_{A=5}(\Pi_B(R)) \stackrel{?}{=} \Pi_B(\sigma_{A=5}(R))$

Renaming (ρ)

- For most queries, need to apply several relational algebra operations one after the other
- **Intermediate result relations** exist. Must give names to the intermediate results.
- Example:

$$\pi_{\text{Fname, Lname, Salary}}(\sigma_{\text{Dno=5}}(\text{EMPLOYEE}))$$

single relational algebra expression

$$\text{DEP5\_EMPS} \leftarrow \sigma_{\text{Dno=5}}(\text{EMPLOYEE})$$
$$\text{RESULT} \leftarrow \pi_{\text{Fname, Lname, Salary}}(\text{DEP5\_EMPS})$$

Renaming (ρ) (cont.)

- To **rename** the attributes in the intermediate and result relations.
- Useful in connection with more complex operations such as **Union** and **Join**

- Example:

Renaming (ρ) (Example)

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

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

$$R(\text{First\_name}, \text{Last\_name}, \text{Salary}) \leftarrow \pi_{\text{Fname, Lname, Salary}}(\text{TEMP})$$

- Can also define a formal **Rename** operation:
  - Rename either the **relation name**
  - or rename the attribute **name**
  - or **both**

- Three forms

$\rho_{S(B1, B2, \ldots, Bn)}(R)$

$\rho_{S}(R)$

$\rho_{(B1, B2, \ldots, Bn)}(R)$

Another example:

Students

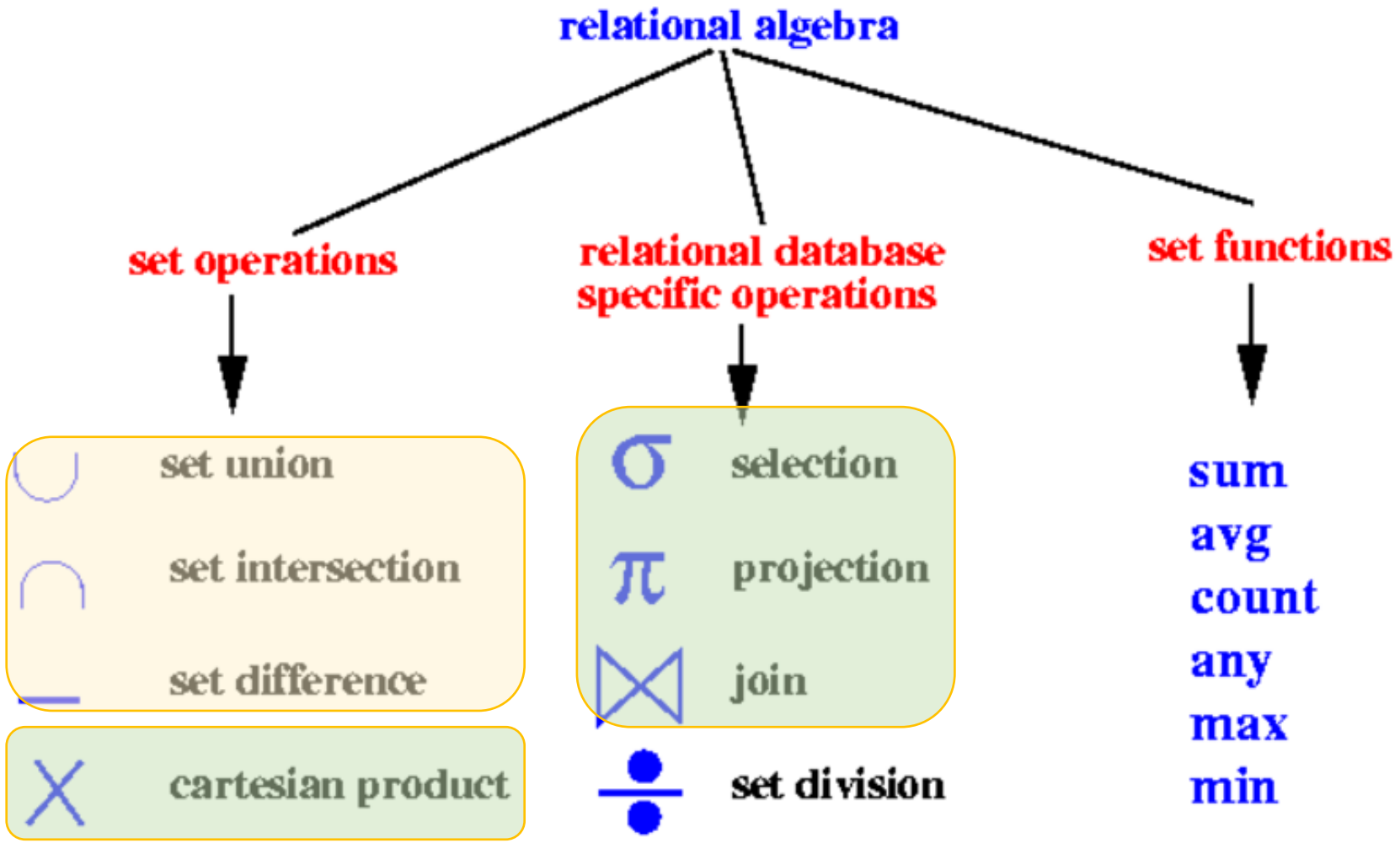
sid	sname	gpa
001	John	3.4
002	Bob	1.3

$\rho_{studId,name,gradePtAvg}(Students)$

Students

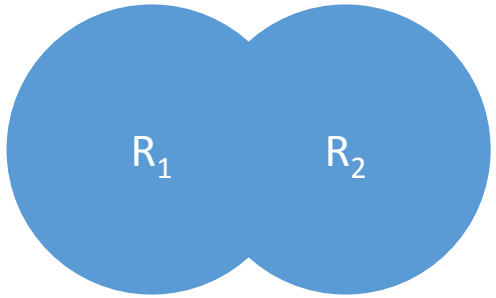
studId	name	gradePtAvg
001	John	3.4
002	Bob	1.3

Relational Algebra (RA) Overview



1. Union (∪)

- R1 ∪ R2 is a relation that includes all tuples that are either in R1 or in R2 or in both R1 and R2
- Example:
  - Retrieve the SSNs of all employees who either work in department 5 or directly supervise an employee who works in department 5



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

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

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

$$RESULT \leftarrow RESULT1 \cup RESULT2$$

Union: produces the tuples that are in either RESULT1 or RESULT2 or both

1. Union (∪) (example)

RESULT1

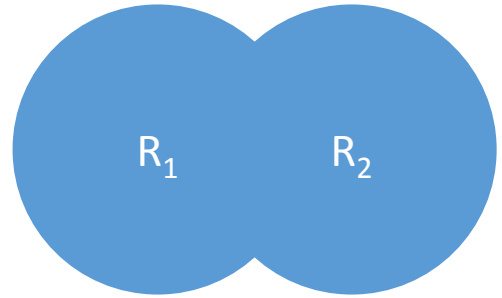
Ssn
123456789
333445555
666884444
453453453

RESULT2

Ssn
333445555
888665555

RESULT

Ssn
123456789
333445555
666884444
453453453
888665555



Union: produces the tuples that are in either RESULT1 or RESULT2 or both

2. Difference (–)

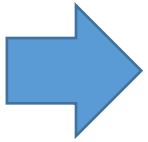
- R1 – R2 is a relation that includes all tuples that are in R1 but not in R2
- Example:

STUDENT

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

INSTRUCTOR

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



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

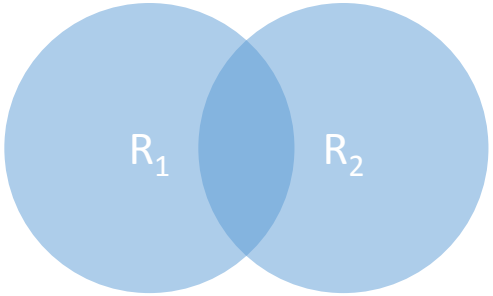
STUDENT – INSTRUCTOR

Fname	Lname
John	Smith
Ricardo	Browne
Francis	Johnson

INSTRUCTOR – STUDENT

3. Intersection (∩) ?

- It is a derived operator
- R1 ∩ R2 = R1 – (R1 – R2)
- is a relation that includes all tuples that are in both R1 and R2
- Example



STUDENT

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

INSTRUCTOR

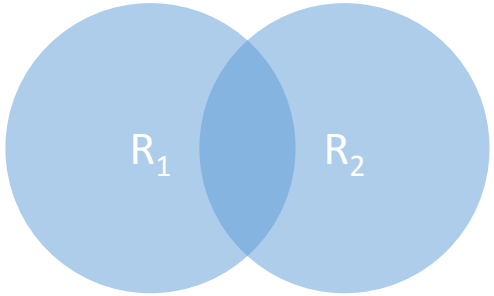
Fname	Lname
John	Smith
Ricardo	Browne
Susan	Yao
Francis	Johnson
Ramesh	Shah

Fn	Ln
Susan	Yao
Ramesh	Shah

STUDENT ∩ INSTRUCTOR

Property of Union, Intersection, Difference

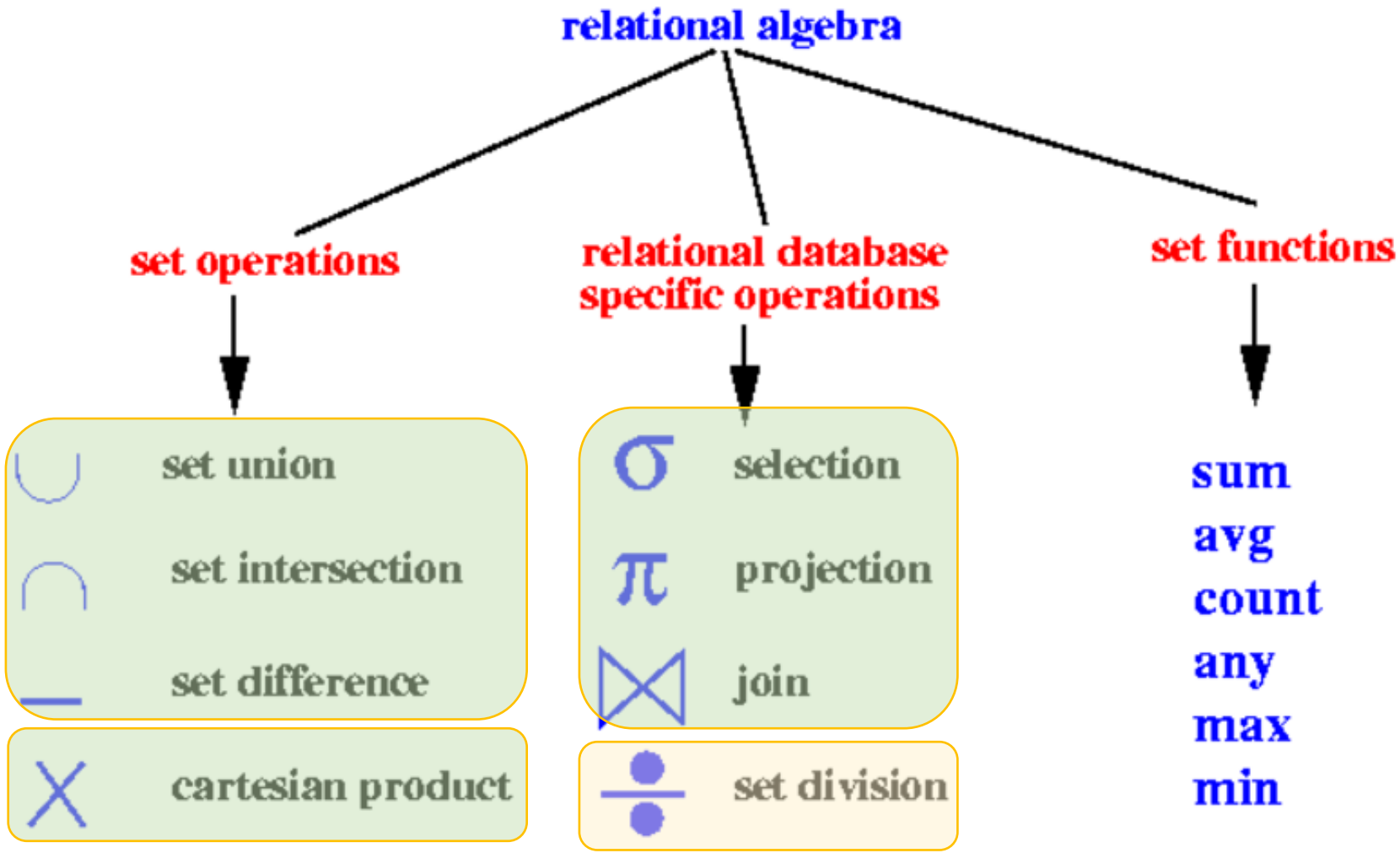
- Commutative
  - $R \cup S = S \cup R$
  - $R \cap S = S \cap R$
  - $R - S \neq S - R$



- Associative
  - $R \cup (S \cup T) = (S \cup R) \cup T$
  - $(R \cap S) \cap T = R \cap (S \cap T)$

- $R \cap S = ((R \cup S) - (R - S)) - (S - R)$

Relational Algebra (RA) Overview



Set Division

- Notation: ÷
- Applied to two relations $R(Z) \div S(X)$, 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 $Z = X \cup Y$)

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

Compute T ← R ÷ S?

$X = \{A\}$, $Y = \{B\}$, and $Z = \{A, B\}$

Set Division (cont.)

- 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

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

$X = \{A\}$, $Y = \{B\}$, and $Z = \{A, B\}$

Set Division (Example)

- Example:
  - *Retrieve the names of employees* ***who*** *work on all the projects* ***that*** *'John Smith' works on*

1) Retrieve the list of project numbers that 'John Smith' works on

$$SMITH \leftarrow \sigma_{Fname='John' \text{ AND } Lname='Smith'}(EMPLOYEE)$$
$$SMITH\_PNOS \leftarrow \pi_{Pno}(WORKS\_ON \bowtie_{Essn=Ssn} SMITH)$$

AMITH_PNOS

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno	Pno	Hours
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5	1	32.5
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5	2	7.5

$$SMITH \leftarrow \sigma_{Fname='John' \text{ AND } Lname='Smith'}(EMPLOYEE)$$
$$SMITH\_PNOS \leftarrow \pi_{Pno}(WORKS\_ON \bowtie_{Essn=Ssn} SMITH)$$

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

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

Set Division (Example) (cont.)

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

SSN_PNOS

e the names of employees ***who*** *work on all the*
s ***that*** *'John Smith' works on*

a relation whenever the employee whose Ssn is
ks on the project whose number is Pno

$$SSN\_PNOS \leftarrow \pi_{Essn,\ Pno}(WORKS\_ON)$$

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

Set Division (Example) (cont.)

- Apply the Division operation to the two relations
  - *Retrieve the names of employees* ***who*** *work on all the projects* ***that*** *'John Smith' works on*

3) Obtain the desired employees' SSNs

$$SSNS(Ssn) \leftarrow SSN\_PNOS \div SMITH\_PNOS$$
$$RESULT \leftarrow \pi_{Fname,\ Lname}(SSNS * EMPLOYEE)$$

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

SMITH_PNOS

Pno
1
2

SSNS

Ssn
123456789
453453453

$$SSNS(Ssn) \leftarrow SSN\_PNOS \div SMITH\_PNOS$$

Set Division (cont.)

- Question: how to express Division as a sequence of other operations?

- The Division operation can be expressed as a sequence of **π**, **×**, and **–** operations as follows:

$$T1 \leftarrow \pi_Y(R)$$
$$T2 \leftarrow \pi_Y((S \times T1) - R)$$
$$T \leftarrow T1 - T2$$

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

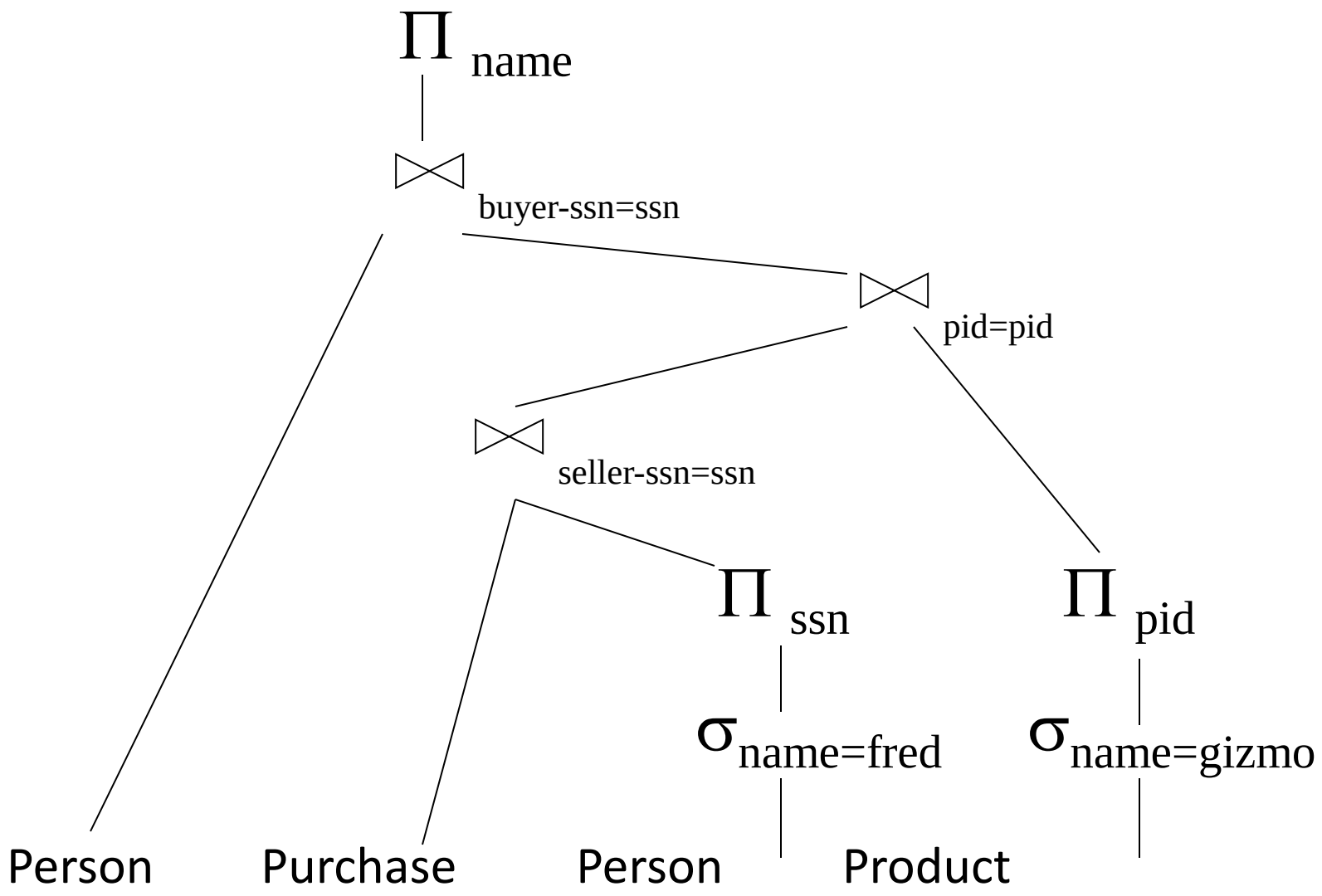
$X = \{A\}$, $Y = \{B\}$, and $Z = \{A, B\}$

Query Tree

- Used as a possible **data structure** for the internal representation of the **query** in an RDBMS
- A tree data structure that corresponds to **a relational algebra expression**
- Represents the **input relations** of the query as *leaf nodes* of the tree
- Represents the **relational algebra operations** as *internal nodes*

RA Expressions Can Get Complex!

Query tree



- Draw the **query tree** for the following query:
  - For every **project** located in 'Stafford', list the project number, the controlling department number, and the department manager's last name, address, and birth date.

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

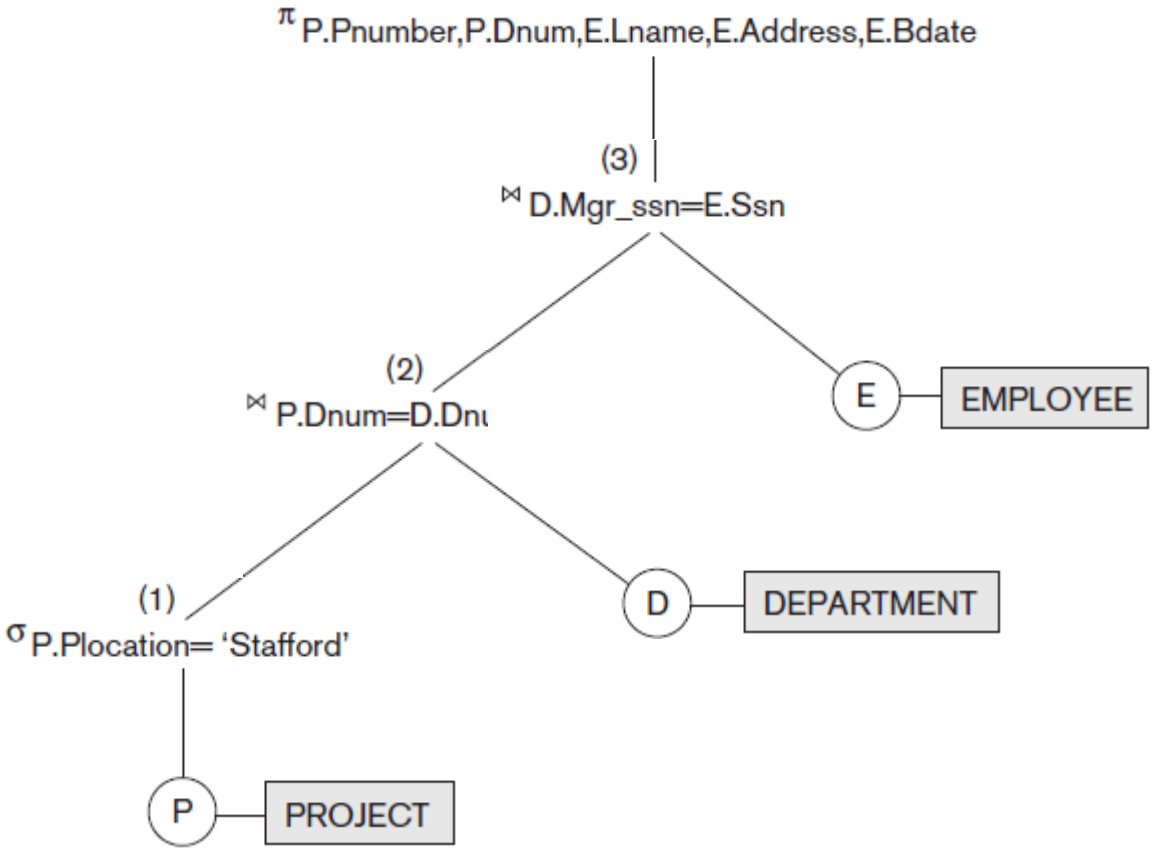
WORKS_ON

Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

Figure 5.5
Schema diagram for the COMPANY relational database schema.



$$\pi_{\text{Pnumber, Dnum, Lname, Address, Bdate}}(((\sigma_{\text{Plocation='Stafford'}}(\text{PROJECT})) \bowtie_{\text{Dnum=Dnumber}} (\text{DEPARTMENT})) \bowtie_{\text{Mgr\_ssn=Ssn}} (\text{EMPLOYEE}))$$

Query Tree (cont.)

- The query tree signifies an explicit order of execution
- A **query tree** gives a good visual representation and understanding of the query in terms of the relational operations it uses
- It is recommended as an additional means for expressing queries in relational algebra

Additional Relational Operations

To enhance the express power of the original relational algebra

Generalized Projection

- Extends the projection operation by allowing **functions of attributes** to be included in the **projection list**
  - $\pi_{F1,\, F2,\, \ldots,\, Fn}\,(R)$
  - where F_1, F_2, … , F_n are functions over the attributes in relation R

- It is helpful when developing reports where computed values have to be produced in the columns of a query result.

- See an example:

Generalized Projection (Example)

- Consider the relation
  - EMPLOYEE (Ssn, Salary, Deduction, Years_service)

- A report may be required to show:

  Net Salary = Salary – Deduction,

  Bonus = 2000 * Years_service, and

  Tax = 0.25 * Salary

$$\text{REPORT} \leftarrow \rho_{(\text{Ssn, Net\_salary, Bonus, Tax})} (\pi_{\text{Ssn, Salary} - \text{Deduction, } 2000 * \text{Years\_service, } 0.25 * \text{Salary}}(\text{EMPLOYEE}))$$

Aggregate Functions and Grouping

- How to handle the following queries?
  - Retrieve the average salary of all employee
  - The total salary
  - The total number of employee tuples

- Aggregate function
  - Used in simple statistical queries that summarize information from the database tuples
  - Sum, Average, Maximum, Minimum, and Counting

Aggregate Functions and Grouping (cont.)

- How to handle the following queries?
  - Retrieve each department number
  - Retrieve the average salary of employees within one department
  - The number of employees who work in the department

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

Aggregate Functions and Grouping (cont.)

- How to handle the following queries?
  - Retrieve each department number
  - Retrieve the average salary of employees within one department
  - The number of employees who work in the department

- Define an Aggregate Function operation

$$\langle \text{grouping attributes} \rangle \, \Im \, \langle \text{function list} \rangle \, (R)$$

Example: $\text{Dno} \, \Im \, \text{COUNT Ssn, AVERAGE Salary}(\text{EMPLOYEE})$.

Aggregate Functions and Grouping (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

Dno	Count_ssn	Average_salary
5	4	33250
4	3	31000
1	1	55000

$$\text{Dno}\ \mathfrak{I}\ \text{COUNT Ssn, AVERAGE Salary}(\text{EMPLOYEE}).$$

Aggregate Functions and Grouping (Example) (cont.)

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

Count_ssn	Average_salary
8	35125

$\Im$ COUNT Ssn, AVERAGE Salary(EMPLOYEE)

Recursive Closure Operations

- How to handle the following queries?
  - Retrieve **all supervisees** of an **employee *e*** at all levels—, that is, all employees **e' directly supervised by e**, and
  - all employees **e'' directly supervised by each employee e',** and
  - all employees **e''' directly supervised by each employee e'',** and so on.

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

Recursive Closure Operations (cont.)

- Applied to a **recursive relationship** between tuples of the same type, such as the relationship between an employee and a supervisor

- Example: specify the Ssns of all employees *e'* directly supervised—at level one—by the employee e whose name is 'James Borg'

$$\text{BORG\_SSN} \leftarrow \pi_{\text{Ssn}}(\sigma_{\text{Fname='James' AND Lname='Borg'}}(\text{EMPLOYEE}))$$
$$\text{SUPERVISION}(\text{Ssn1}, \text{Ssn2}) \leftarrow \pi_{\text{Ssn,Super\_ssn}}(\text{EMPLOYEE})$$
$$\text{RESULT1}(\text{Ssn}) \leftarrow \pi_{\text{Ssn1}}(\text{SUPERVISION} \bowtie_{\text{Ssn2}=\text{Ssn}} \text{BORG\_SSN})$$

$$BORG\_SSN \leftarrow \pi_{Ssn}(\sigma_{Fname='James' \text{ AND } Lname='Borg'}(EMPLOYEE))$$

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

BROG_SSN

Ssn
888665555

SUPERVISION

(Borg's Ssn is 888665555)

(Ssn)	(Super_ssn)
Ssn1	Ssn2
123456789	333445555
333445555	888665555
999887777	987654321
987654321	888665555
666884444	333445555
453453453	333445555
987987987	987654321
888665555	null

$$SUPERVISION(Ssn1, Ssn2) \leftarrow \pi_{Ssn,Super\_ssn}(EMPLOYEE)$$

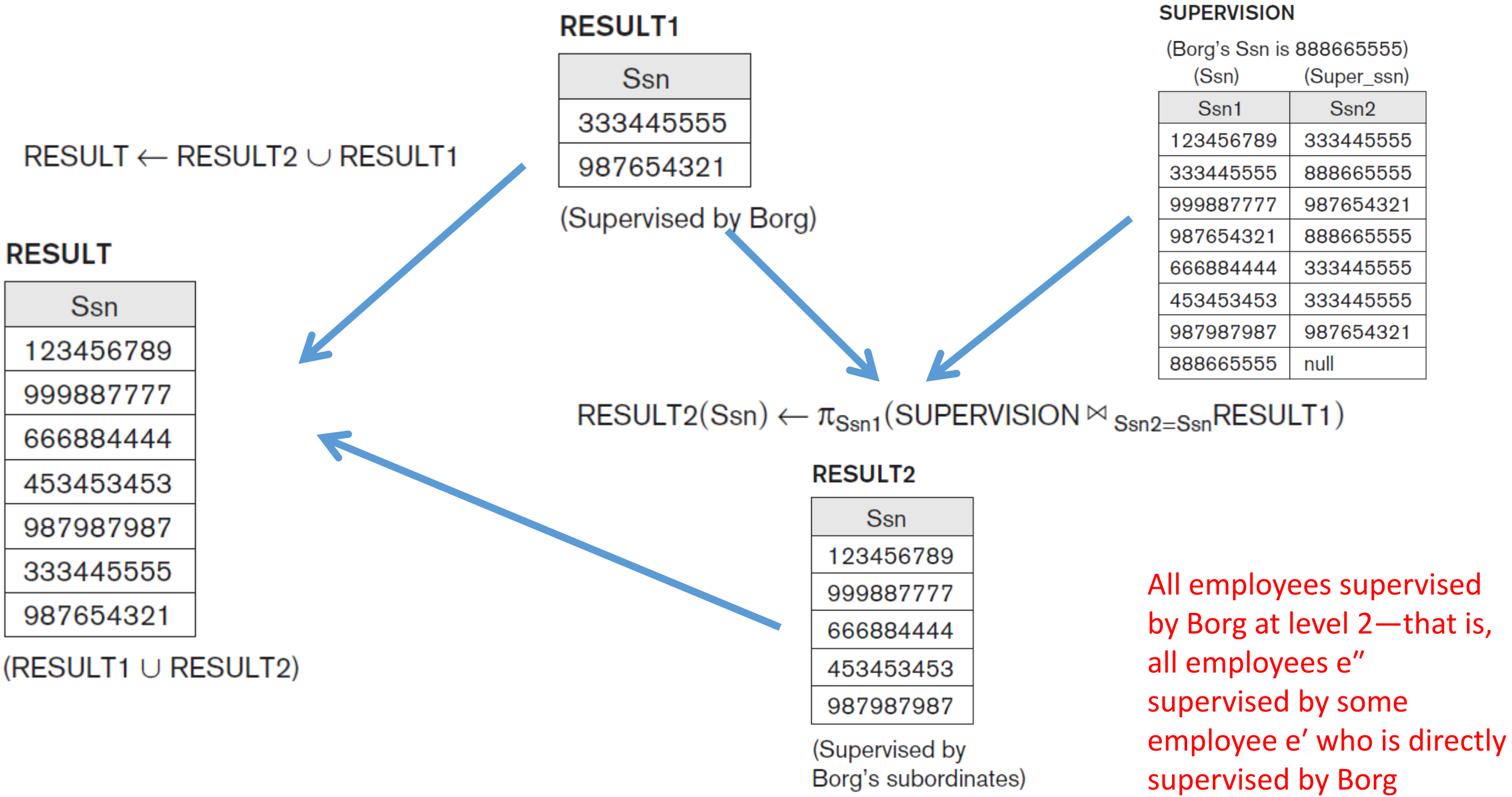
$$RESULT1(Ssn) \leftarrow \pi_{Ssn1}(SUPERVISION \bowtie_{Ssn2=Ssn} BORG\_SSN)$$

RESULT1

Ssn
333445555
987654321

(Supervised by Borg)

All employees e' directly supervised—at level one—by James Borg'


RESULT ← RESULT2 ∪ RESULT1
RESULT1
Ssn
333445555
987654321
(Supervised by Borg)
SUPERVISION
(Borg's Ssn is 888665555)
(Ssn) (Super_ssn)
Ssn1 Ssn2
123456789 333445555
333445555 888665555
999887777 987654321
987654321 888665555
666884444 333445555
453453453 333445555
987987987 987654321
888665555 null
RESULT
Ssn
123456789
999887777
666884444
453453453
987987987
333445555
987654321
(RESULT1 ∪ RESULT2)
RESULT2(Ssn) ← πSsn1(SUPERVISION ⋈ Ssn2=SsnRESULT1)
RESULT2
Ssn
123456789
999887777
666884444
453453453
987987987
(Supervised by
Borg's subordinates)
All employees supervised by Borg at level 2—that is, all employees e" supervised by some employee e' who is directly supervised by Borg