FK Clarification

- R1(ABC), R2(EFG). There is a FK: R1.A references R2.G.
- The value of R1.A must be
  - NULL, or
  - A unique value in R2
- That is, R2.G must be unique or NULL.
- R2.G might or might not be a PK!
  - R2.G could be NULL. while the value of PK can never be NULL.

Exercise – FK (Why it is wrong)

STUDENT(SID, Name, Major, Bdate)

COURSE(Course#, Cname, Dept)

ENROLL(SID, Course#, Quarter, Grade)

BOOK_ADOPTION(Course#, Quarter, Book_ISBN)

TEXT(Book_ISBN, Book_Title, Publisher, Author)

FK: BOOK_ADOPTION.(Course#, Quarter)
references ENROLL.(Course#, Quarter)

ENROLL

SID	Course#	Quarter	Grade
12	CS 338	W21	
23	CS 338	W21	
34	CS 136	W21	

BOOK_ADOPTION

Course#	Quarter	Book_ISBN
CS 338	W21	B22
CS 330	W21	B35
CS 136	W21	B99

Exercise – FK (Why it is right)

STUDENT(SID, Name, Major, Bdate)

COURSE(Course#, Cname, Dept)

ENROLL(SID, Course#, Quarter, Grade)

BOOK_ADOPTION(Course#, Quarter, Book_ISBN)

TEXT(Book_ISBN, Book_Title, Publisher, Author)

FK: ENROLL.(Course#, Quarter)
references BOOK_ADOPTION.(Course#, Quarter)

ENROLL

SID	Course#	Quarter	Grade
12	CS 338	W21	
23	CS 338	W21	
34	CS 136	W21	

BOOK_ADOPTION

Course#	Quarter	Book_ISBN
CS 338	W21	B22
CS 330	W21	B35
CS 136	W21	B99

Review of Last Lecture

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	$\pi_{\text{<ATTRIBUTE LIST>}}(R)$
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.	$R_1 - R_2$

Today's Plan

- More advanced RA operators
  - Rename (ρ)
  - Cross Product (X)
  - Join ($\bowtie$)
    - EquiJoin
    - Natural Join (*)
- Exercises

Rename Operation (ρ)

- The rename operator is ρ
  - $\rho_S(R)$ is a renamed relation S based on R
  - $\rho_{S(B_1, B_2, \ldots, B_n)}(R)$ is a renamed relation S based on R with column names $B_1, B_1, \ldots. B_n$.
  - $\rho_{(B_1, B_2, \ldots, B_n)}(R)$ is a renamed relation with column names $B_1, B_1, \ldots.B_n$ which does not specify a new relation name.

Be Aware!

- Rename is in fact a copy!
  - It returns a new relation with the same schema and content of the original, just different name (for the relation, attributes or both)
  - The original relation is unchanged!

- For convenience, when the assignment operator is used, we can do rename like this:

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

$$\text{RESULT}(\text{FN, LN, S}) \leftarrow \pi_{\text{FNAME, LNAME, SALARY}}(\text{DEP5\_EMPS})$$

In-class Exercise #2

1. Retrieve the names of the department managers.

2. Find all the employees (SSN) who do not work on any project.

3. Retrieve the SSN of employees who are not a supervisor nor a manager.

How would you do this manually?

1. Retrieve the names of the department managers.

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

Cartesian (cross product) Operation

- R1 X R2
- Each row of R1 is paired with each row of R2.
- *Result schema* has one field per field of R1 and R2, with field names `inherited' if possible.
- *What about R1 X R1?*

Teacher X Teacher t-num t-name t-num t-name

Conflict!

- Solution: use Rename operation

Rename

Cross Product: Example

X cross product

Teacher

t-num	t-name
101	Smith
105	Jones
110	Fong

Teacher X Course

Course

c-num	c-name
514	Intro to DB
513	Intro to OS

t-num	t-name	c-num	c-name
101	Smith	514	Intro to DB
105	Jones	514	Intro to DB
110	Fong	514	Intro to DB
101	Smith	513	Intro to OS
105	Jones	513	Intro to OS
110	Fong	513	Intro to OS

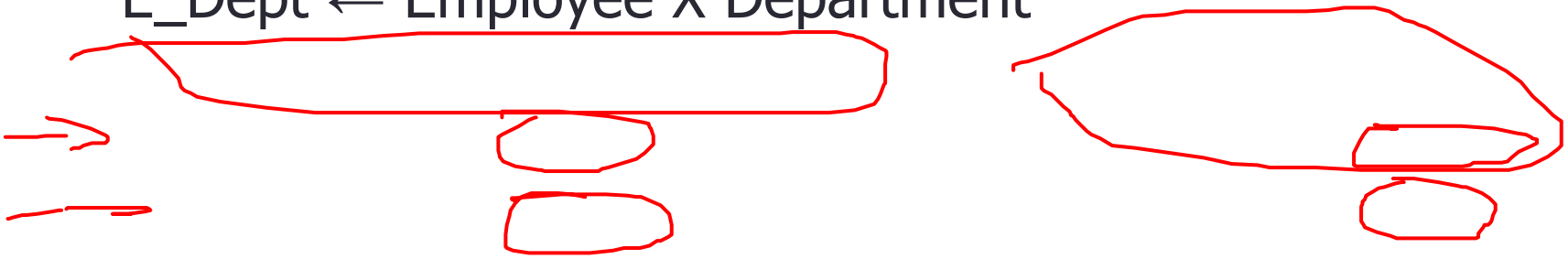
Cross product: combine information from 2 tables

- produces: every possible combination of a teacher and a course

In-class Exercise #2

1. Retrieve the names of the department managers.

E_Dept ← Employee X Department



Fname	Lname	…	SSN	…	DNO	…	DName	DNumber	MGRSSN	…
Franklin	Wong	…	333445555	…		…	Research	5	333445555	…
Jennifer	Wallace	…	987654321	…		…	Administration	4	987654321	…
James	Borg	…	888665555	…		…	Headquarters	1	888665555	…

$$Dept\_MGR \leftarrow \sigma_{MGRSSN = SSN}(E\_Dept)$$

$$Result \leftarrow \pi_{Fname,\ Lname}(Dept\_MGR)$$

Join: Motivation

Company

Example: Suppose that we want 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. We can do this by using the join ⋈ operation.

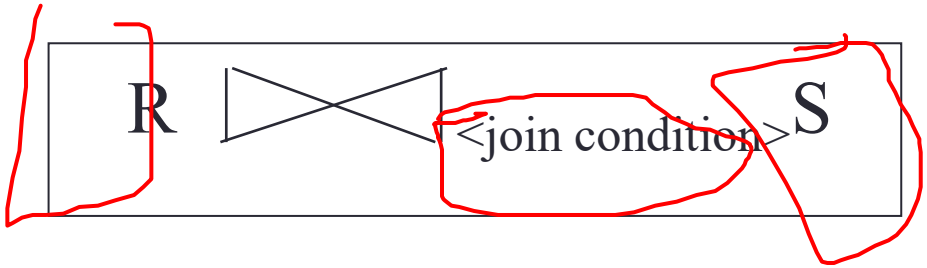
$$E\_Dept \leftarrow Employee \times Department$$
$$Dept\_MGR \leftarrow \sigma_{SSN = MGRSSN}(E\_Dept)$$

$$\mathbf{DEPT\_MGR \leftarrow Employee \bowtie_{MGRSSN=SSN} Department}$$

Fname	Lname	…	SSN	…	DNO	…	DName	DNumber	MGRSSN	…
Franklin	Wong	…	333445555	…		…	Research	5	333445555	…
Jennifer	Wallace	…	987654321	…		…	Administration	4	987654321	…
James	Borg	…	888665555	…		…	Headquarters	1	888665555	…

Join Operation: General

- 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**.



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

Basic Join Operation

- *Condition Join*: $R \bowtie_c S = \sigma_c (R \times S)$

- *Result schema* same as that of cross-product
- Fewer tuples than cross-product, might be able to compute more efficiently
- Sometimes called a *theta-join*.

Join: Example

Div

$$\text{Account} \bowtie_{\text{Number=Account}} \text{Deposit}$$

Account

Number	Owner	Balance	Type
101	J. Smith	1000.00	checking
102	W. Wei	2000.00	checking
103	J. Smith	5000.00	savings
104	M. Jones	1000.00	checking
105	H. Martin	10,000.00	checking

Deposit

Account	Transaction-id	Date	Amount
102	1	10/22/00	500.00
102	2	10/29/00	200.00
104	3	10/29/00	1000.00
105	4	11/2/00	10,000.00

Number	Owner	Balance	Type	Account	Transaction-id	Date	Amount
102	W. Wei	2000.00	checking	102	1	10/22/00	500.00
102	W. Wei	2000.00	checking	102	2	10/29/00	200.00
104	M. Jones	1000.00	checking	104	3	10/29/00	1000.00
105	H. Martin	10,000.00	checking	105	4	11/2/00	10000.00

In-class Exercise #3

- The following query does not produce meaningful information. It is created to test your understanding of the join operator.
- What is its result?

$$\pi_{\text{Number}}(\text{Account} \bowtie_{\text{Number} > \text{Account}} \text{Deposit})$$

Account	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	103	J. Smith	5000.00	savings
	104	M. Jones	1000.00	checking
	105	H. Martin	10,000.00	checking

Deposit	Account	Transaction-id	Date	Amount
	102	1	10/22/00	500.00
	102	2	10/29/00	200.00
	104	3	10/29/00	1000.00
	105	4	11/2/00	10,000.00

In-class Exercise #4

- The following query does not produce meaningful information. It is created to test your understanding of the join operator.
- What is its result?

$\bowtie$ = join $\quad \pi_{\text{Number}}(\text{Account} \bowtie_{\text{Balance=Amount}} \text{Deposit})$

Account	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	103	J. Smith	5000.00	savings
	104	M. Jones	1000.00	checking
	105	H. Martin	10,000.00	checking

Deposit	Account	Transaction-id	Date	Amount
	102	1	10/22/00	500.00
	102	2	10/29/00	200.00
	104	3	10/29/00	1000.00
	105	4	11/2/00	10,000.00

EquiJoin Problem: Example

- Problem: join with equal
  - Some call it EquiJoin

Account $\bowtie_{\text{Number=Account}}$ Deposit

Note that when the join is based on equality, then we have two identical attributes (columns) in the answer.

Number	Owner	Balance	Type	Account	Trans-id	Date	Amount
102	W. Wei	2000.00	checking	102	1	10/22/00	500.00
102	W. Wei	2000.00	checking	102	2	10/29/00	200.00
104	M. Jones	1000.00	checking	104	3	10/29/00	1000.00
105	H. Martin	10,000.00	checking	105	4	11/2/00	10000.00

Natural Join Operation

- Definition: join on the equality of the common attribute(s), only one copy of the common attribute(s) survives
  - "Common": attribute(s) with same name(s)
  - Denoted by "*"
- Assume R(ABC), S(AD)

Result ← R * S

R1 ← R X S
R2 ← $\sigma_{R.A = S.A}$ (R1)
Result: one copy of column A removed

Result(ABCD)

Natural Join: Example

Example: To apply a natural join on the DNUMBER attributes of DEPARTMENT and DEPT_LOCATIONS, it is sufficient to write:

DEPT_LOCS ← DEPARTMENT * DEPT_LOCATIONS

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

In-Class Exercises #5

- What is the result?

1. $AB \bowtie_{A=a1} CD1$
2. $AB \bowtie_{A=a1} CD2$
3. AB*CD1
4. AB*CD2
5. AB*CD3
6. AB*CD4
7. AB*CD5

AB	A	B
	a1	b1
	a2	b2

CD1	C	D	X
	c1	d1	x1
	d2	d2	x2

CD4	AA1	BB1	X
	a1	b1	x1
	a2	b2	x2

CD2	A	D	X
	a2	d2	x1
	a1	d1	x2

CD5	A	B	X
	a1	d1	x1
	a2	d2	x2

CD3	A	B	X
	a1	b1	x1
	a2	b2	x2

Summary of Today's Lecture

- Advanced RA operations:
  - Rename (ρ)
  - Cross Product (X)
  - Join (⋈)
    - EquiJoin
    - Natural Join (*)
- Exercises