# Lecture 3 Relational Algebra

CS211 - Introduction to Database

# Relational Algebra

CS211 Introduction to Database – Relational Algebra

## Relational Algebra

A formal language

Queries in algebra are based on sets, set operators

• Every operator in the algebra accepts(one or two) relation instances as operands, and return a relation instance as the result

• Schemas of input relations for a query are fixed, schema of output relation also fixed, but query will run regardless of instance.

## Relational Algebra and SQL

select name, salary from employee where salary < 50000

 $\pi$ name, salary ( $\sigma$ salary < 50000 (employee))

Osalary <  $50000(\Pi name, salary(employee))$ 

### Operations

- Usual set operations (relations are sets)
  - union  $R \cup S$
  - intersection  $R \cap S$
  - Set-difference *R-S*
  - Cartesian Product  $R \times S$

- Division
- Outer join

- Operations on relations
  - selection:  $\sigma_{condition}(R)$
  - projection:  $\pi_A(R)$
  - join  $R \bowtie_{A\Theta B} S$

 $\pi_{\text{name, salary}}(\sigma_{\text{salary}} < 50000(\text{employee}))$ 



#### Union - U

- *R* U *S* returns a relation instance containing all tuples that occur in either *R* or *S* or both
- R and S must be union-compatible
  - Same number of attributes
  - Corresponding attributes have the same domains

R			
A1	A2	А3	
а	b	С	
а	d	g	
f	b	е	

S			
B1	B2	В3	
а	b	С	
а	b	е	
а	d	g	
h	d	g	

How many tuples in  $R \cup S$ ?

Union: Remove duplicate tuples

### Union – example

Girls Boys sID sID **sName** sName 4003 4001 Bob Amy 4005 John 4002 Alice 4004 Cathy

Girls  $\cup$  *Boys* 

List of all students.

#### intersection

#### $R \cap S$

- $R \cap S$  returns a relation instance containing all tuples that occur in both R and S
- R and S must be union-compatible

R			
A1	A2	A3	
а	b	С	
а	d	g	
f	b	е	

S			
B1	B2	В3	
а	b	С	
а	b	е	
а	d	g	
h	d	g	

$R \cap S$			
C1	C2	C3	
а	b	С	
а	d	g	

### Intersection - example

RegDB: Students take Database

RegCPP: Students take C++

 $RegDB \cap RegCPP$ 

List of students who registered both Database and C++

### Difference

- R-S returns a relation instance containing all tuples that occur in R but not in S
- R and S must be union-compatible

R			
A1	A2	A3	
а	b	С	
а	d	g	
f	b	е	

S			
B1	B2	В3	
а	b	С	
а	b	е	
а	d	g	
h	d	g	

R-S		
C1	C2	C3
f	b	е

S-R			
D1	D2	D3	
а	b	е	
h	d	g	

### Difference – example

RegDB: Students take Database

RegCPP: Students take C++



RegDB - RegCPP

List of students who takes course Database but not C++

#### Cartesian Product

 R×S returns a relation instance whose schema contains all the fields of R followed by all the fields of S

• Result of  $R \times S$  $\{(r,s) | r \in R, s \in S\}$ 

R	
Α	В
а	1
b	2

S		
С	D	E
а	10	α
b	10	α
b	20	β
С	10	β

$R \times S / S \times R$				
Α	В	С	D	Е
а	1	а	10	α
а	1	b	10	α
а	1	b	20	β
а	1	С	10	β
b	2	а	10	α
b	2	b	10	α
b	2	b	20	β
b	2	С	10	β



## Cartesian Product – example

Student				Course	
sID	sName	grade	gender	cID	cName
4001	Amy	4	0	11	Database
4002	Alice	4	0	25	C++
4003	Bob	3	1	28	OS
4004	Cathy	3	0		
4005	John	3	1		

How many tuples in *Student*×*Course*?

### Selection

- Selection selects subset of rows from relation
- Expression

$$\sigma_{condition}(R)$$

 Schema of the result is the same as the schema of input

R			
A1	A2	A3	
а	а	10	
а	d	-4	
f	b	5	

$\sigma_{A3>0}(R)$				
A1	A2	A3		
а	а	10		
f	b	5		

$\sigma_{A2='a'orA2='b'}(R)$				
A1	A2	А3		
а	а	10		
f	b	5		

#### Selection $\sigma_{condition}(R)$

#### Student

sID	sName	grade	gender
4001	Amy	4	0
4002	Alice	4	0
4003	Bob	3	1
4004	Cathy	3	0
4005	John	3	1

Return students whose sID is greater than or equal to 4003

$$\sigma_{SID \geq 4003}(Student)$$

Return students who are not in the above relation (sID < 4003)

$$\sigma_{\neg(SID \geq 4003)}(Student)$$

#### Selection condition

- Selection condition is a Boolean combination of terms that have the forms
  - attribute op constant
  - attribute1 op attributes2
    - where op is a comparison operator:  $\geq$ ,  $\leq$ , >, <,=,!=
  - Terms are combined with: and  $\Lambda$ , or V

	operator
1	()
2	≥,≤,>,<,=,!=
3	7
4	Λ
5	V

#### Student

sID	sName	grade	gender	GPA
4001	Amy	4	0	3.6
4002	Alice	4	0	3
4003	Bob	3	1	3.3
4004	Cathy	3	0	3.5
4005	John	3	1	3.5

### Projection

- Projection deletes unwanted columns from relation
- Expression (unary operator)  $\pi_{Attributes}(R)$
- Schema of the result is determined by the attributes that are projected

R		
A1	A2	А3
а	b	С
е	d	С
f	b	С

Duplicates are eliminated, however, real systems often omit this step

$\pi_{A3,A2}(R)$		
A3	A2	
С	b	
С	d	

### Example – selection and projection

#### Student

sID	sName	grade	gender
4001	Amy	4	0
4002	Alice	4	0
4003	Bob	3	1
4004	Cathy	3	0
4005	John	3	1

Return the ID and name of those female students who are junior or senior

$$\pi_{SID,SName}(\sigma_{gender=0 \ and \ grade \geq 3}(Student))$$

## Renaming the relation

- Renaming the relation
- Expression

 $\rho_{Rnew}(Rold)$ 

• E.g.  $\rho_{stu}(student)$ 

### Join

- To combine information from 2 or more relations
- Expression

 $R\bowtie_{A\Theta B} S$ 

- where  $A\Theta B$  is the condition
- Step1: Cartesian product
- Step2: selection

R	
Α	В
а	1
b	2

S	
Н	С
1	X
1	У
3	Z

$R \times S$			
Α	В	Н	D
а	1	1	X
а	1	1	У
а	1	3	Z
b	2	1	X
b	2	1	У
b	2	3	Z

R ⋈ <sub>B≤H</sub> S			
Α	В	Н	D
а	1	1	X
а	1	1	У
а	1	3	Z
b	2	3	Z

## join-example1

R

sID	sName	grade	gender	GPA
4001	Amy	4	0	3.6
4002	Alice	4	0	3
4003	Bob	3	1	3.3
4004	Cathy	3	0	3.5
4005	John	3	1	3.5

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

 $\pi_{R.SID,R.SName,S.cName}(R \bowtie_{R.GPA \geq S.minGPA} S)$ 

List all colleges a student (ID and name) can apply

## join-example2

R

sID	cID
4001	11
4001	25
4001	28
4002	11
4003	25
4003	28
4004	11
4004	25
4005	25
4005	28

Return the courses(ID) that both student 4001 and 4005 registered

l	?	R	1
sID	cID	sID	cID
4001	11	4001	11
4001	25	4001	25
4001	28	4001	28
4002	11	4002	11
4003	25	4003	25
4003	28	4003	28
4004	11	4004	11
4004	25	4004	25
4005	25	4005	25
4005	28	4005	28

## join-example2-step1

Return the courses that both student 4001 and 4005 registered

$$R \bowtie_{R.cID=R1.cID} \rho_{R1}(R)$$

	sID	cID	sID	cID	
	4001	11	4001	11	
	4002	11	4001	11	
	4004	11	4001	11	
	4001	25	4001	25	
	4003	25	4001	25	
	4004	25	4001	25	
	4005	25	4001	25	
	4001	28	4001	28	
	4003	28	4001	28	
	4005	28	4001	28	
	4001	11	4002	11	
	4002	11	4002	11	
	4004	11	4002	11	
	4001	25	4003	25	
	4003	25	4003	25	
	4004	25	4003	25	
	4005	25	4003	25	
	4001	28	4003	28	
	4003	28	4003	28	
	4005	28	4003	28	
	4001	11	4004	11	
	4002	11	4004	11	
	4004	11	4004	11	
	4001	25	4004	25	
	4003	25	4004	25	
	4004	25	4004	25	
_	4005	25	4004	25	
$\subset$	4001	25	4005	25	
	4003	25	4005	25	
	4004	25	4005	25	
	4005	25	4005	25	
	4001	28	4005	28	)
	4003	28	4005	28	
	4005	28	4005	28	

## join-example2-step2

Return the courses that both student 4001 and 4005 registered

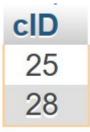
$$\sigma_{R.SID=4001 \ and \ R1.SID=4005}(R \bowtie_{R.cID=R1.cID} \rho_{R1}(R))$$

sID	cID	sID	cID
4001	25	4005	25
4001	28	4005	28

## join-example2-step3

Return the courses that both student 4001 and 4005 registered

$$\pi_{R.cID}(\sigma_{R.sID=4001 \ and \ R1.sID=4005}(R \bowtie_{R.cID=R1.cID} \rho_{R1}(R)))$$



## Natural join

 $R\bowtie S$ 

- Enforce equality on all attributes with same name
- Eliminate one copy of duplicate attributes

R	
Α	В
а	1
b	2

S	
В	С
1	X
1	У
3	Z

R⋈S				
Α	В	С		
а	1	X		
а	1	У		

## Natural join

#### instructor

ilD	iName
101	Edward
102	Vadim
103	Prabhu
104	Liu Fang

#### lecture

ilD	cID
104	11
103	28
102	25

#### instructor∞lecture

iID	iName	cID
102	Vadim	25
103	Prabhu	28
104	Liu Fang	11

### Composing larger expressions

- Expressions can be composed recursively.
- Parentheses and precedence rules define the order of evaluation.
- Precedence, from highest to lowest, is:
  - σ, π, ρ
  - ×, ⋈
  - \(\)
  - ∪, –
- Unless very sure, use brackets

# Relational Algebra - Practice

CS211 Introduction to Database – Relational Algebra

**Apply** 

sID	cName	major	decision
123	Stanford	CS	N
123	Stanford	EE	N
123	Berkeley	CS	Υ
123	Cornell	EE	N
234	Berkeley	biology	Υ
345	MIT	bioengineering	Υ
345	Cornell	bioengineering	N
345	Cornell	CS	Υ
345	Cornell	EE	N
678	Stanford	history	Y
987	Stanford	CS	Υ
987	Berkeley	CS	Υ
876	Stanford	CS	N
876	MIT	biology	Υ
876	MIT	marine biology	N
765	Stanford	history	Υ
765	Cornell	history	N
765	Cornell	psychology	Υ
543	MIT	CS	N

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

#### College

sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

#### Student

#### **Apply** cName major decision Stanford CS Ν Stanford EE Berkeley CS EE Cornell Berkeley biology 345 MIT bioengineering Y 345 Cornell bioengineering N 345 Cornell CS EE Cornell Ν Stanford history Stanford CS Berkeley CS Stanford CS 876 MIT biology 876 MIT marine biology N 765 Stanford history 765 Cornell history Ν Cornell psychology 543 MIT CS Ν

#### Student

sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

#### College

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• List ID of the students whose GPA=3.5

#### Apply cName major decision Stanford CS Ν Stanford EE Berkeley CS EE Cornell Berkeley biology 345 MIT bioengineering Y 345 Cornell bioengineering N 345 Cornell CS Cornell EE Ν Stanford history Stanford CS Berkeley CS Stanford CS 876 MIT biology 876 MIT marine biology N Stanford history 765 765 Cornell history Ν psychology Cornell 543 MIT CS Ν

#### Student

sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

#### College

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• List ID of the students whose GPA=3.5

$$\pi_{SID}(\sigma_{GPA=3.5}Student)$$

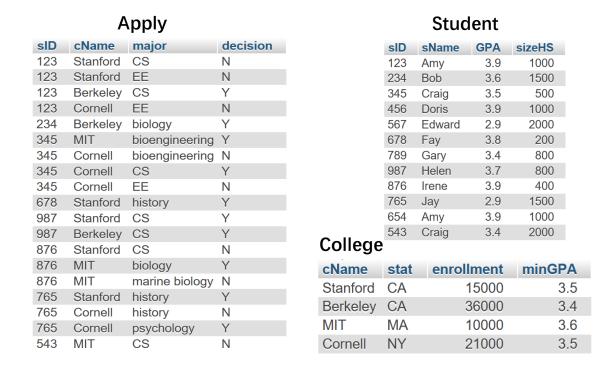
	Apply						Stuc	lent	
sID	cName	major	decision			sID	sName	GPA	sizeHS
123	Stanford	CS	N			123	Amy	3.9	1000
123	Stanford	EE	N			234	Bob	3.6	1500
123	Berkeley	CS	Υ			345	Craig	3.5	500
123	Cornell	EE	N			456	Doris	3.9	1000
234	Berkeley	biology	Υ			567	Edward	2.9	2000
345	MIT	bioengineering	Υ			678	Fay	3.8	200
345	Cornell	bioengineering	N			789	Gary	3.4	800
345	Cornell	CS	Υ			987	Helen	3.7	800
345	Cornell	EE	N			876	Irene	3.9	400
678	Stanford	history	Υ			765	Jay	2.9	1500
987	Stanford	CS	Υ			654	Amy	3.9	1000
987	Berkeley	CS	Υ		<b>.</b>	543	Craig	3.4	2000
876	Stanford	CS	N		College	<u> </u>			
876	MIT	biology	Υ		cName	stat	enrol	lment	minGF
876	MIT	marine biology	N		Stanford	CA		15000	3
765	Stanford	history	Υ						_
765	Cornell	history	N		Berkeley	CA		36000	3
765	Cornell	psychology	Υ		MIT	MA		10000	3
543	MIT	CS	N		Cornell	NY		21000	3

• List ID and name of the students who is studying at a large high school (sizeHS >1000)

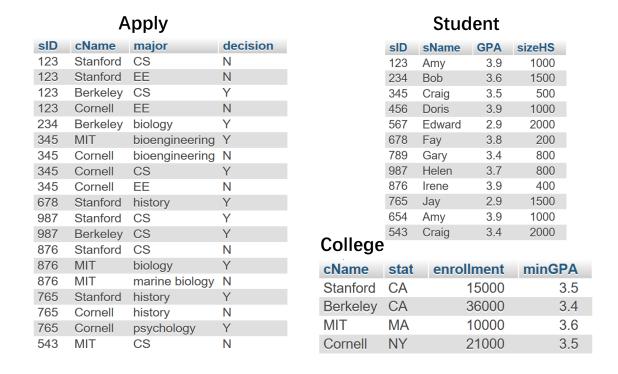
Apply						Stuc	dent	
sID	cName	major	decision		sID	sName	GPA	sizeHS
123	Stanford	CS	N		123	Amy	3.9	1000
123	Stanford	EE	N		234	Bob	3.6	1500
123	Berkeley	CS	Υ		345	Craig	3.5	500
123	Cornell	EE	N		456	Doris	3.9	1000
234	Berkeley	biology	Υ		567	Edward	2.9	2000
345	MIT	bioengineering	Υ		678	Fay	3.8	200
345	Cornell	bioengineering	N		789	Gary	3.4	800
345	Cornell	CS	Υ		987	Helen	3.7	800
345	Cornell	EE	N		876	Irene	3.9	400
678	Stanford	history	Υ		765	Jay	2.9	1500
987	Stanford	CS	Υ		654	Amy	3.9	1000
987	Berkeley	CS	Υ	0 - 11	543	Craig	3.4	2000
876	Stanford	CS	N	College	)			
876	MIT	biology	Υ	cName	stat	enrol	lment	minGP
876	MIT	marine biology	N	Stanford	CA		15000	3
765	Stanford	history	Υ					
765	Cornell	history	N	Berkeley	CA		36000	3
765	Cornell	psychology	Υ	MIT	MA		10000	3
543	MIT	CS	N	Cornell	NY		21000	3

• List ID and name of the students who is studying at a large high school (sizeHS >1000)

 $\pi_{SID,SName}(\sigma_{sizeHS>1000}Student)$ 

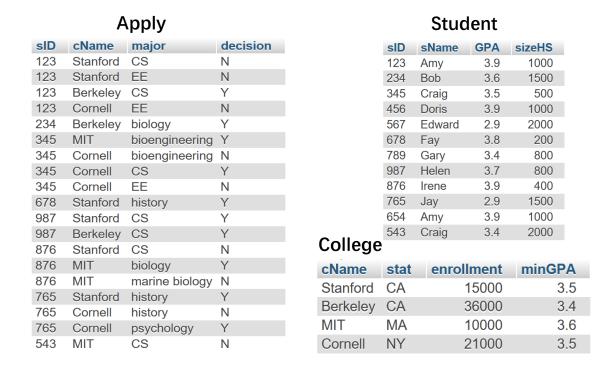


• List student ID, student name, the applied college name, and the state each college locates in.



• List student ID, student name, the applied college name, and the state each college locates in.

 $\pi_{Student.sID,SName,Apply.cName,stat}(Student \bowtie Apply \bowtie College)$ 



• Find students who applies colleges located in CA, list student ID, student name, college name and enrollment.

	A	pply				Stuc	lent	
sID	cName	major	decision		sID	sName	GPA	sizeHS
123	Stanford	CS	N		123	Amy	3.9	1000
123	Stanford	EE	N		234	Bob	3.6	1500
123	Berkeley	CS	Υ		345	Craig	3.5	500
123	Cornell	EE	N		456	Doris	3.9	1000
234	Berkeley	biology	Υ		567	Edward	2.9	2000
345	MIT	bioengineering	Υ		678	Fay	3.8	200
345	Cornell	bioengineering	N		789	Gary	3.4	800
345	Cornell	CS	Υ		987	Helen	3.7	800
345	Cornell	EE	N		876	Irene	3.9	400
678	Stanford	history	Υ		765	Jay	2.9	1500
987	Stanford	CS	Υ		654	Amy	3.9	1000
987	Berkeley	CS	Υ	0.11	543	Craig	3.4	2000
876	Stanford	CS	N	College	!			
876	MIT	biology	Υ	cName	stat	enrol	lment	minGF
876	MIT	marine biology	N	Stanford	CA		15000	3
765	Stanford	history	Υ					
765	Cornell	history	N	Berkeley	CA		36000	3
765	Cornell	psychology	Υ	MIT	MA	,	10000	3
543	MIT	CS	N	Cornell	NY		21000	3

• Find students who applies colleges located in CA, list student ID, student name, college name and enrollment.

 $\pi_{Student.sID,SName,Apply.cName,enrollment}(\sigma_{stat="CA"}(Student \bowtie Apply \bowtie College))$ 

#### **Apply** cName major decision Stanford CS Ν Stanford EE Ν 123 Berkeley CS EE Cornell Ν Berkeley biology 345 MIT bioengineering Y 345 bioengineering N Cornell 345 CS Cornell EE Cornell Ν Stanford history Stanford CS Berkeley CS 876 Stanford CS 876 MIT biology 876 MIT marine biology N 765 Stanford history 765 Cornell history Ν Cornell psychology 543 MIT CS Ν

0.0.0.0					
sID	sName	GPA	sizeHS		
123	Amy	3.9	1000		
234	Bob	3.6	1500		
345	Craig	3.5	500		
456	Doris	3.9	1000		
567	Edward	2.9	2000		
678	Fay	3.8	200		
789	Gary	3.4	800		
987	Helen	3.7	800		
876	Irene	3.9	400		
765	Jay	2.9	1500		
654	Amy	3.9	1000		

3.4

2000

Student

#### College 543 Craig

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• Find students(ID) who applies MIT or Cornell

Apply				
sID	cName	major	decision	
123	Stanford	CS	N	
123	Stanford	EE	N	
123	Berkeley	CS	Υ	
123	Cornell	EE	N	
234	Berkeley	biology	Υ	
345	MIT	bioengineering	Υ	
345	Cornell	bioengineering	N	
345	Cornell	CS	Υ	
345	Cornell	EE	N	
678	Stanford	history	Υ	
987	Stanford	CS	Υ	
987	Berkeley	CS	Υ	
876	Stanford	CS	N	
876	MIT	biology	Υ	
876	MIT	marine biology	N	
765	Stanford	history	Υ	
765	Cornell	history	N	
765	Cornell	psychology	Υ	
543	MIT	CS	N	

sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
			4500

3.4

2000

Student

#### College 543 Craig

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• Find students(ID) who applies MIT or Cornell

$$\pi_{SID} \left( \sigma_{cName="MIT" \ or \ cName="Cornell"} \left( Apply \right) \right)$$

#### **Apply** cName major decision Stanford CS Ν Stanford EE Ν 123 Berkeley CS Υ EE Cornell Ν Berkeley biology Υ 345 MIT bioengineering Y bioengineering N 345 Cornell 345 CS Cornell Cornell EE Ν Stanford history Stanford CS Berkeley CS 876 Stanford CS 876 MIT biology 876 MIT marine biology N 765 Stanford history 765 Cornell history Ν Cornell psychology 543 MIT CS Ν

		Student					
	sID	sName	GPA	sizeHS			
	123	Amy	3.9	1000			
	234	Bob	3.6	1500			
	345	Craig	3.5	500			
	456	Doris	3.9	1000			
	567	Edward	2.9	2000			
	678	Fay	3.8	200			
	789	Gary	3.4	800			
	987	Helen	3.7	800			
	876	Irene	3.9	400			
	765	Jay	2.9	1500			
	654	Amy	3.9	1000			
• "	543	Craig	3.4	2000			
College							
cName	stat	enrol	lment	minGPA	1		
Stanford	CA	•	15000	3.5	5		

36000

10000

21000

3.4

3.6

3.5

Berkeley CA

MA

NY

MIT

Cornell

• Find students(ID) who applies both MIT and Cornell

	Apply				
sID	cName	major	decision		
123	Stanford	CS	N		
123	Stanford	EE	N		
123	Berkeley	CS	Υ		
123	Cornell	EE	N		
234	Berkeley	biology	Υ		
345	MIT	bioengineering	Υ		
345	Cornell	bioengineering	N		
345	Cornell	CS	Υ		
345	Cornell	EE	N		
678	Stanford	history	Υ		
987	Stanford	CS	Υ		
987	Berkeley	CS	Υ		
876	Stanford	CS	N		
876	MIT	biology	Υ		
876	MIT	marine biology	N		
765	Stanford	history	Υ		
765	Cornell	history	N		
765	Cornell	psychology	Υ		
543	MIT	CS	N		

sID	sName	GPA	sizeHS
	0		0
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

Student

#### College

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• Find students(ID) who applies both MIT and Cornell

$$\pi_{SID} \left( \sigma_{cName="MIT" \ and \ cName="Cornell"} \left( Apply \right) \right) \times$$

$$\pi_{SID} \left( \sigma_{cName="MIT"} (Apply) \right) \cap \pi_{SID} \left( \sigma_{cName="Cornell"} (Apply) \right)$$

#### **Apply** cName major decision Stanford CS Ν Stanford EE Ν Berkeley CS EE Cornell Ν Berkeley biology 345 MIT bioengineering Y 345 bioengineering N Cornell 345 CS Cornell EE Cornell Ν Stanford history Stanford CS Berkeley CS 876 Stanford CS 876 MIT biology 876 MIT marine biology N 765 Stanford history 765 Cornell history Ν Cornell psychology 543 MIT CS Ν

	Otaaciit					
sID	sName	GPA	sizeHS			
123	Amy	3.9	1000			
234	Bob	3.6	1500			
345	Craig	3.5	500			
456	Doris	3.9	1000			
567	Edward	2.9	2000			
678	Fay	3.8	200			
789	Gary	3.4	800			
987	Helen	3.7	800			
876	Irene	3.9	400			
765	Jay	2.9	1500			
654	Amy	3.9	1000			

3.4

2000

Student

#### College 543 Craig

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• Find students(ID) who does not apply Stanford

Apply			
sID	cName	major	decision
123	Stanford	CS	N
123	Stanford	EE	N
123	Berkeley	CS	Υ
123	Cornell	EE	N
234	Berkeley	biology	Υ
345	MIT	bioengineering	Υ
345	Cornell	bioengineering	N
345	Cornell	CS	Υ
345	Cornell	EE	N
678	Stanford	history	Υ
987	Stanford	CS	Υ
987	Berkeley	CS	Υ
876	Stanford	CS	N
876	MIT	biology	Υ
876	MIT	marine biology	N
765	Stanford	history	Υ
765	Cornell	history	N
765	Cornell	psychology	Υ
543	MIT	CS	N

sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

Student

College

cName	stat	enrollment	minGPA
Stanford	CA	15000	3.5
Berkeley	CA	36000	3.4
MIT	MA	10000	3.6
Cornell	NY	21000	3.5

• Find students(ID) who does not apply Stanford

$$\pi_{SID} \left( \sigma_{cName!="Standford"} \left( Apply \right) \right) \times$$

$$\pi_{SID}(Student) - \pi_{SID} \left(\sigma_{cName="Standford"} (Apply)\right)$$

# Relational Algebra – division, outer join

CS211 Introduction to Database – Relational Algebra

- R has 2 attributes, A1, and A2
- S has only attribute A2
- R÷S
  - If the set of A2 values associated with an A1 value in R contains all A2 values in S, the A1 value is in  $R \div S$
  - Schema: attributes of R attributes of S
  - $\{A1 A2\} \{A2\} = \{A1\}$

R		
A1	A2	
а	X	
а	Υ	
b	X	

S	
A2	
Χ	
Υ	



R			
A1	A2	A3	
а	b	С	
d	b	С	
а	е	С	
а	е	f	
d	b	f	
а	е	g	
а	е	h	
а	b	I	

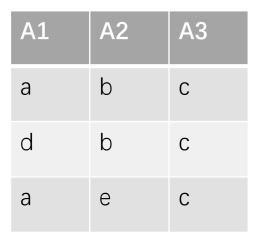
**A3** 

A1	A2	A3
а	b	С
d	b	С
а	е	С

R÷S			
A1	A2		
а	b		
d	b		
а	е		

	R	
A1	A2	A3
а	b	С
d	b	С
а	е	С
а	е	f
d	b	f
а	е	g
а	е	h
а	b	1

S'
A3
C



A1	A2	A3
а	е	f
d	b	f

R <b>÷S′</b>		
A1	A2	
d	b	
а	е	

R			
A1	A2	A3	
а	b	С	
d	b	С	
а	е	С	
а	е	f	
d	b	f	
а	е	g	
а	е	h	
а	b	1	

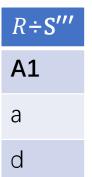
S"
A3
С
f
g
h

A1	A2	A3
а	b	С
d	b	С
а	е	С
A1	A2	A3
а	е	f
d	b	f
A1	A2	A3
а	е	g
A1	A2	A3
а	е	h

	, <u>, , , , , , , , , , , , , , , , , , </u>					
	b	С				
	b	С				
	е	С				
1	A2	A3				
	е	f			R	e÷S"
	b	f		$\cap$	A1	A2
1	A2	A3			а	е
	е	9				
1	A2	A3				
	е	h	 J			

R					
A1	A2	A3			
а	b	С			
d	b	С			
а	е	С			
а	е	f			
d	b	f			
а	е	g			
а	е	h			
а	b	1			

S'''				
A2	A3			
b	С			



#### Use of Division

R				
sID	cID			
4001	11			
4001	25			
4001	28			
4002	11			
4003	25			
4003	28			
4004	11			
4004	25			
4005	25			
4005	28			

S

cID	cName
11	Database
25	C++
28	OS

Return students(sID) who registered all the courses in S

$$\pi_{SID,CID}(R) \div \pi_{CID}(S)$$

#### Outer join - Motivation

Return all instructors' ID, name, lectured course id and name

Inst	tructor		Le	cture		Co	ourse
ilD	iName		iID	cID		cID	cName
101	Edward	$\bowtie$	104	11	$\bowtie$	11	Database
102	Vadim		103	28		25	C++
103	Prabhu		102	25		28	OS
104	Liu Fang						

ilD	iName	cID	cName
102	Vadim	25	C++
103	Prabhu	28	OS
104	Liu Fang	11	Database

 $\pi_{iID,iName,cID,cName}(Instructor \bowtie Lecture \bowtie Course)$ 

# Outer join >

R			
A1	A2		
1	а		
2	b		
3	С		

S			
A2	А3		
С	11		
d	44		
е	55		

$R \bowtie \mathbf{S}$						
R.A2=S.A2						
A1	R.A2	S.A2	A3			
1	а	Null	Null			
2	b	Null	Null			
3	С	С	11			
Null	Null	d	44			
Null	Null	е	55			

# Left/Right Outer join ≥ ✓/✓

R					
A1	A2				
1	а				
2	b				
3	С				

S						
A2	A3					
С	11					
d	44					
е	55					

<i>R</i> ⋈ <b><i>S</i></b> R.A2=S.A2						
A1	R.A2	S.A2	A3			
1	а	Null	Null			
2	b	Null	Null			
3	С	С	11			

<i>R</i> ⋈ <b><i>S</i></b> R.A2=S.A2						
A1	R.A2	S.A2	A3			
3	С	С	11			
Null	Null	d	44			
Null	Null	е	55			

### Outer join

Return all instructor's ID, name, lectured course id and name

Instructor			Lecture			Course					
iID	iName		:ID	٥ID		cID	cName	iID	iName	cID	cName
	IIVallie		ilD	cID		CID	104	104	Liu Fang	11	Database
101	Edward	$\bowtie$	104	11	$\bowtie$	11	Database	102	Vadim	25	C++
102	Vadim		103	28		25	C++				
102		102			28	28	28 OS	103	Prabhu	28	OS
103	Prabhu		25	20	03	101	Edward	NULL	NULL		
104	Liu Fang										

 $\pi_{iID,iName,cID,cName}(Instructor \bowtie Lecture \bowtie Course)$