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
```

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

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

Operations

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

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

- Division
- Outer join

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

Union - ∪

- $R \cup S$ 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	A3
a	b	c
a	d	g
f	b	e

S		
B1	B2	B3
a	b	c
a	b	e
a	d	g
h	d	g

How many tuples in $R \cup S$?

Union: Remove duplicate tuples

Union – example

Girls

sID	sName
4001	Amy
4002	Alice
4004	Cathy

Boys

sID	sName
4003	Bob
4005	John

$\text{Girls} \cup Boys$

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**
a	b	c
a	d	g
f	b	e

S		
B1	**B2**	**B3**
a	b	c
a	b	e
a	d	g
h	d	g

$R \cap S$		
C1	**C2**	**C3**
a	b	c
a	d	g

Intersection - example

RegDB: Students take Database

sID
4001
4002
4004

RegCPP: Students take C++

sID
4001
4003
4004
4005

$RegDB \cap RegCPP$

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**
a	b	c
a	d	g
f	b	e

S		
B1	**B2**	**B3**
a	b	c
a	b	e
a	d	g
h	d	g

R-S		
C1	**C2**	**C3**
f	b	e

S-R		
D1	**D2**	**D3**
a	b	e
h	d	g

Difference – example

RegDB: Students take Database

sID
4001
4002
4004

RegCPP: Students take C++

sID
4001
4003
4004
4005

$$RegDB - RegCPP$$

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

Cartesian Product

- $R \times S$ 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	
A	**B**
a	1
b	2

S		
C	**D**	**E**
a	10	α
b	10	α
b	20	β
c	10	β

$R \times S$ / $S \times R$				
A	**B**	**C**	**D**	**E**
a	1	a	10	α
a	1	b	10	α
a	1	b	20	β
a	1	c	10	β
b	2	a	10	α
b	2	b	10	α
b	2	b	20	β
b	2	c	10	β

Cartesian Product – example

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

Course

cID	cName
11	Database
25	C++
28	OS

How many tuples in $Student \times Course$?

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**
a	a	10
a	d	-4
f	b	5

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

$\sigma_{A2='a'orA2='b'}(R)$		
A1	**A2**	**A3**
a	a	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: ≥ , ≤, >, <,=,!=
  - Terms are combined with: and Λ, or V

	operator
1	()
2	≥ , ≤, >, <,=,!=
3	¬
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

$$sID < 4004 \text{ V } grade < 4 \text{ Λ } gender < 1$$

$$(sID < 4004 \text{ V } grade < 4) \text{ Λ } gender < 1$$

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**	**A3**
a	b	c
e	d	c
f	b	c

$\pi_{A1}(R)$
A1
a
e
f

$\pi_{A3,A2}(R)$	
A3	**A2**
c	b
c	d

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

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	
A	**B**
a	1
b	2

S	
H	**C**
1	x
1	y
3	z

$R \times S$			
A	**B**	**H**	**D**
a	1	1	x
a	1	1	y
a	1	3	z
b	2	1	x
b	2	1	y
b	2	3	z

$R \bowtie_{B \leq H} S$			
A	**B**	**H**	**D**
a	1	1	x
a	1	1	y
a	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

S

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

R

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

$R1$

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

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

cID
25
28

Natural join $R \bowtie S$

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

R	
A	**B**
a	1
b	2

S	
B	**C**
1	x
1	y
3	z

$R \bowtie S$		
A	**B**	**C**
a	1	x
a	1	y

Natural join

instructor

iID	iName
101	Edward
102	Vadim
103	Prabhu
104	Liu Fang

lecture

iID	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
  - $\times$, $\bowtie$
  - $\cap$
  - $\cup$, $-$
- 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	Y
123	Cornell	EE	N
234	Berkeley	biology	Y
345	MIT	bioengineering	Y
345	Cornell	bioengineering	N
345	Cornell	CS	Y
345	Cornell	EE	N
678	Stanford	history	Y
987	Stanford	CS	Y
987	Berkeley	CS	Y
876	Stanford	CS	N
876	MIT	biology	Y
876	MIT	marine biology	N
765	Stanford	history	Y
765	Cornell	history	N
765	Cornell	psychology	Y
543	MIT	CS	N

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

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

Apply

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

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

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

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

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

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

Apply

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

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

$$\pi_{sID,sName}(\sigma_{sizeHS>1000}Student)$$

Apply

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

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

Apply

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

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

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

Apply

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

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

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

Apply

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

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

- 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

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

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

- Find students(ID) who applies MIT or Cornell

Apply

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

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

- Find students(ID) who applies MIT or Cornell

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

Apply

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

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

- 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	Y
123	Cornell	EE	N
234	Berkeley	biology	Y
345	MIT	bioengineering	Y
345	Cornell	bioengineering	N
345	Cornell	CS	Y
345	Cornell	EE	N
678	Stanford	history	Y
987	Stanford	CS	Y
987	Berkeley	CS	Y
876	Stanford	CS	N
876	MIT	biology	Y
876	MIT	marine biology	N
765	Stanford	history	Y
765	Cornell	history	N
765	Cornell	psychology	Y
543	MIT	CS	N

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

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

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

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

Apply

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

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

- Find students(ID) who does not apply Stanford

Apply

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

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

- Find students(ID) who does not apply Stanford

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

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

Relational Algebra – division, outer join

CS211 Introduction to Database – Relational Algebra

Division

- *R* has 2 attributes, A1, and A2
- *S* has only attribute A2
- $R \div S$
  - 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**
a	X
a	Y
b	X

S
A2
X
Y

$R \div S$
A1
a

Division

R		
A1	A2	A3
a	b	c
d	b	c
a	e	c
a	e	f
d	b	f
a	e	g
a	e	h
a	b	I

S
A3
c

A1	A2	A3
a	b	c
d	b	c
a	e	c

$R \div S$	
A1	A2
a	b
d	b
a	e

Division

R		
A1	**A2**	**A3**
a	b	c
d	b	c
a	e	c
a	e	f
d	b	f
a	e	g
a	e	h
a	b	I

S’
A3
c
f

A1	A2	A3
a	b	c
d	b	c
a	e	c

A1	A2	A3
a	e	f
d	b	f

∩

$R \div S'$	
A1	**A2**
d	b
a	e

Division

R		
A1	**A2**	**A3**
a	b	c
d	b	c
a	e	c
a	e	f
d	b	f
a	e	g
a	e	h
a	b	I

S''
A3
c
f
g
h

A1	A2	A3
a	b	c
d	b	c
a	e	c

A1	A2	A3
a	e	f
d	b	f

A1	A2	A3
a	e	g

A1	A2	A3
a	e	h

∩

$R \div S''$	
A1	**A2**
a	e

Division

R		
A1	A2	A3
a	b	c
d	b	c
a	e	c
a	e	f
d	b	f
a	e	g
a	e	h
a	b	I

S'''	
A2	A3
b	c

$R \div S'''$
A1
a
d

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

Instructor

iID	iName
101	Edward
102	Vadim
103	Prabhu
104	Liu Fang

⋈

Lecture

iID	cID
104	11
103	28
102	25

⋈

Course

cID	cName
11	Database
25	C++
28	OS

iID	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	a
2	b
3	c

S	
A2	**A3**
c	11
d	44
e	55

$R \text{⟗} S$ R.A2=S.A2			
A1	**R.A2**	**S.A2**	**A3**
1	a	Null	Null
2	b	Null	Null
3	c	c	11
Null	Null	d	44
Null	Null	e	55

Left/Right Outer join ⟕/⟖

R	
A1	**A2**
1	a
2	b
3	c

S	
A2	**A3**
c	11
d	44
e	55

$R \overset{\text{R.A2=S.A2}}{\bowtie} S$			
A1	**R.A2**	**S.A2**	**A3**
1	a	Null	Null
2	b	Null	Null
3	c	c	11

$R \overset{\text{R.A2=S.A2}}{\bowtie} S$			
A1	**R.A2**	**S.A2**	**A3**
3	c	c	11
Null	Null	d	44
Null	Null	e	55

Outer join

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

Instructor

iID	iName
101	Edward
102	Vadim
103	Prabhu
104	Liu Fang

⋈

Lecture

iID	cID
104	11
103	28
102	25

⋈

Course

cID	cName
11	Database
25	C++
28	OS

iID	iName	cID	cName
104	Liu Fang	11	Database
102	Vadim	25	C++
103	Prabhu	28	OS
101	Edward	*NULL*	*NULL*

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