Today

Relations

- Binary relations and properties
- Relationship to functions

n-ary relations

- Definitions
- CS application: Relational DBMS

Binary relations establish a relationship between elements of two sets

Definition: Let A and B be two sets. A binary relation from A to B is a subset of A × B.

In other words, a binary relation R is a set of ordered pairs (a_i, b_i) where $a_i \in A$ and $b_i \in B$.

Notation: We say that

- a R b if $(a,b) \in R$
- a $\not R$ b if $(a,b) \notin R$

Example: Course Enrollments

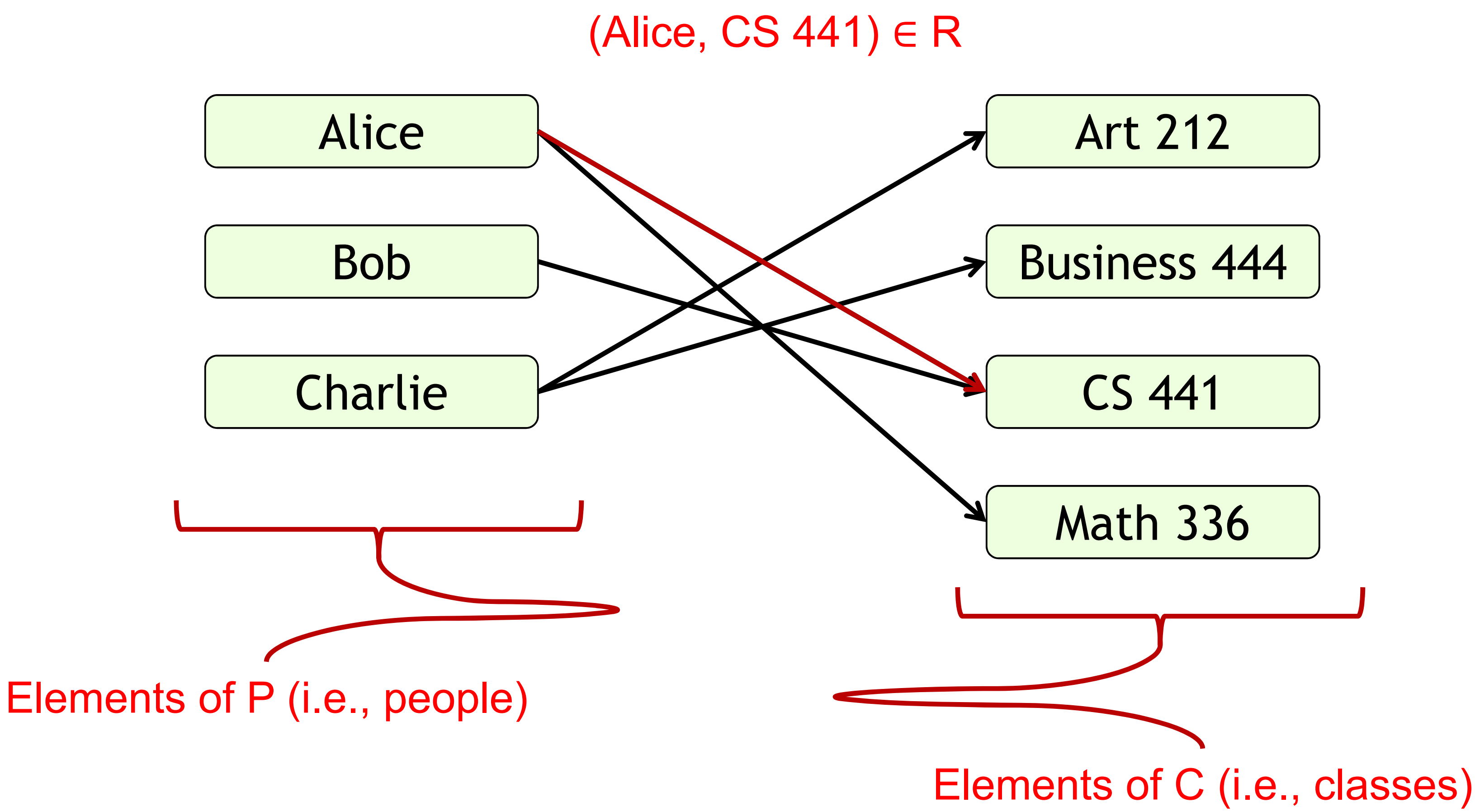
Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.

Solution:

- Let the set P denote people, so P = {Alice, Bob, Charlie}
- Let the set C denote classes, so C = {CS 441, Math 336, Art 212, Business 444)
- By definition $R \subseteq P \times C$
- From the above statement, we know that
  - (Alice, CS 441) ∈ R
  - (Bob, CS 441) ∈ R
  - (Alice, Math 336) ∈ R
  - (Charlie, Art 212) ∈ R
  - (Charlie, Business 444) ∈ R
- So, R = {(Alice, CS 441), (Bob, CS 441), (Alice, Math 336), (Charlie, Art 212), (Charlie, Business 444)}

A relation can also be represented as a graph

Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.



A relation can also be represented as a table

Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.

Name of the relation

(Bob, CS 441) ∈ R

Elements of C (i.e., courses)

R	Art 212	Business 444	CS 441	Math 336
Alice			X	X
Bob			X	
Charlie	X	X		

Elements of P (i.e., people)

Wait, doesn't this mean that relations are the same as functions?

Not quite… Recall the following definition from past Lecture.

Definition: Let A and B be nonempty sets. A function, f, is an assignment of exactly one element of set B to each element of set A.

This would mean that, e.g., a person only be enrolled in one course!

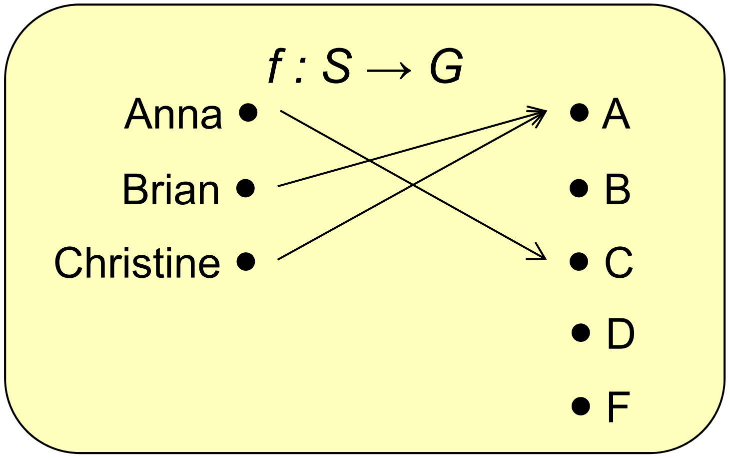
Reconciling this with our definition of a relation, we see that

1. Every function is also a relation
2. Not every relation is a function

Let's see some quick examples…

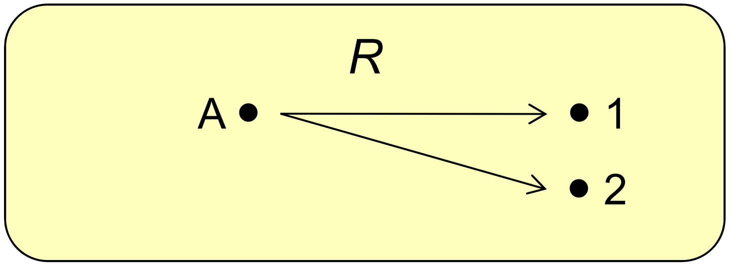
Short and sweet…

1. Consider $f : S \rightarrow G$
   - Clearly a function
   - Can also be represented as the relation R = {(Anna, C), (Brian, A), (Christine A)}



---

1. Consider the set R = {(A, 1), (A, 2)}
   - Clearly a relation
   - Cannot be represented as a function!



We can also define binary relations on a single set

Definition: A relation on the set A is a relation from A to A. That is, a relation on the set A is a subset of A × A.

Example: Let A be the set {1, 2, 3, 4}. Which ordered pairs are in the relation R = {(a, b) | a divides b}?

We can also define binary relations on a single set

Definition: A relation on the set A is a relation from A to A. That is, a relation on the set A is a subset of A × A.

Example: Let A be the set {1, 2, 3, 4}. Which ordered pairs are in the relation R = {(a, b) | a divides b}?

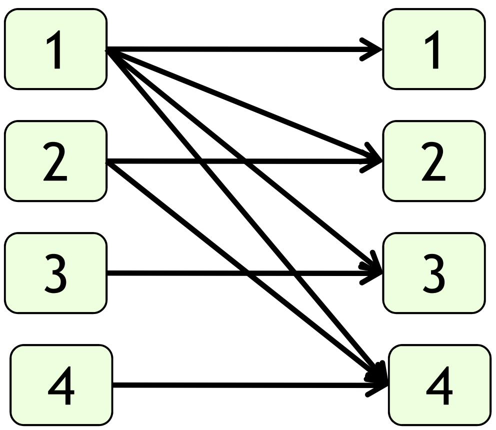
Solution:

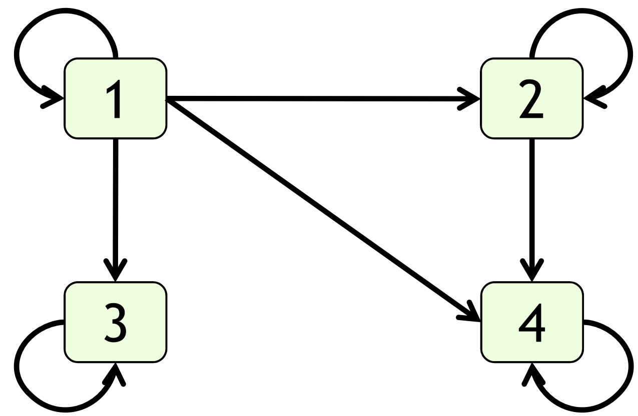
- 1 divides everything (1,1), (1,2), (1,3), (1,4)
- 2 divides itself and 4 (2,2), (2,4)
- 3 divides itself (3,3)
- 4 divides itself (4,4)

- So, R = {(1,1), (1,2), (1,3), (1,4), (2,2), (2,4), (3,3), (4,4)}

Representing the last example as a graph…

Example: Let A be the set {1, 2, 3, 4}. Which ordered pairs are in the relation R = {(a, b) | a divides b}?





Tell me what you know…

Question: Which of the following relations contain each of the pairs (1,1), (1,2), (2,1), (1,-1), and (2,2)?

- $R_1 = \{(a,b) \mid a \leq b\}$
- $R_2 = \{(a,b) \mid a > b\}$
- $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}$
- $R_4 = \{(a,b) \mid a = b\}$
- $R_5 = \{(a,b) \mid a = b + 1\}$
- $R_6 = \{(a,b) \mid a + b \leq 3\}$

These are all relations on an infinite set!

Answer:

	(1,1)	(1,2)	(2,1)	(1,-1)	(2,2)
R_1					
R_2					
R_3					
R_4					
R_5					
R_6					

Tell me what you know…

Question: Which of the following relations contain each of the pairs (1,1), (1,2), (2,1), (1,-1), and (2,2)?

- $R_1 = \{(a,b) \mid a \leq b\}$
- $R_2 = \{(a,b) \mid a > b\}$
- $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}$
- $R_4 = \{(a,b) \mid a = b\}$
- $R_5 = \{(a,b) \mid a = b + 1\}$
- $R_6 = \{(a,b) \mid a + b \leq 3\}$

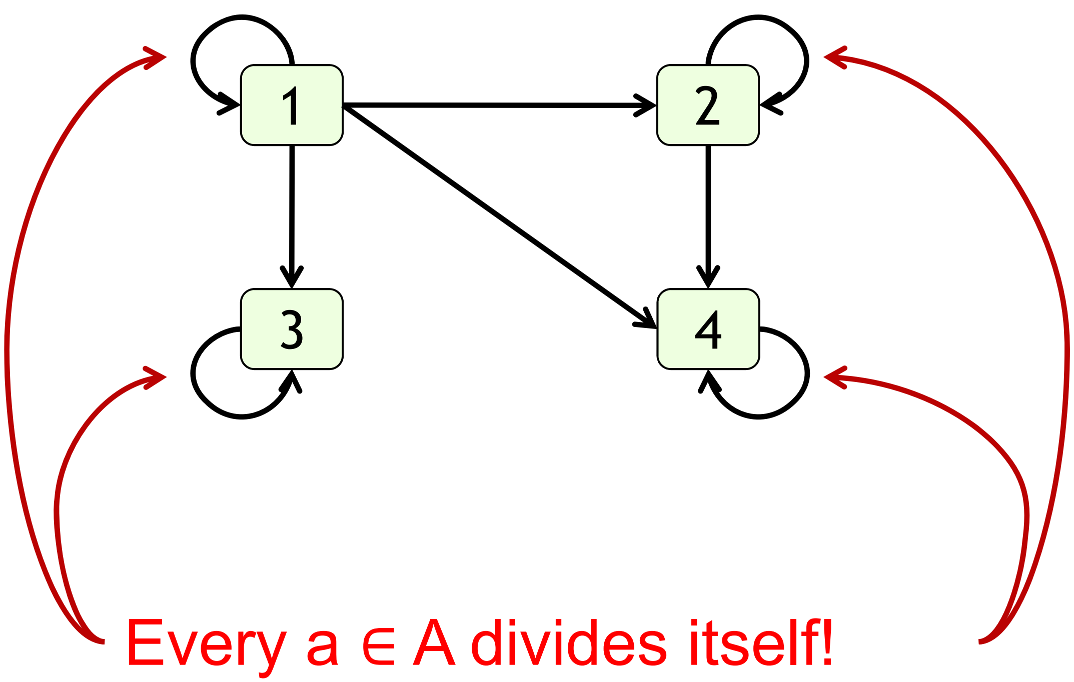
Answer:

	(1,1)	(1,2)	(2,1)	(1,-1)	(2,2)
R_1	Yes	Yes	No	No	Yes
R_2	No	No	Yes	Yes	No
R_3	Yes	No	No	Yes	Yes
R_4	Yes	No	No	No	Yes
R_5	No	No	Yes	No	No
R_6	Yes	Yes	Yes	Yes	No

Properties of Relations

Definition: A relation R on a set A is reflexive if (a,a) ∈ R for every a ∈ A.

Note: Our "divides" relation on the set A = {1,2,3,4} is reflexive.



	1	2	3	4
1	X	X	X	X
2		X		X
3			X	
4				X

Properties of Relations

Definition: A relation R on a set A is symmetric if $(b,a) \in R$ whenever $(a,b) \in R$ for every $a,b \in A$. If R is a relation in which $(a,b) \in R$ and $(b,a) \in R$ implies that $a=b$, we say that R is antisymmetric.

Mathematically:

- Symmetric: $\forall a \forall b((a,b) \in R \rightarrow (b,a) \in R)$
- Antisymmetric: $\forall a \forall b(((a,b) \in R \wedge (b,a) \in R) \rightarrow (a = b))$

Examples:

- Symmetric: R = {(1,1), (1,2), (2,1), (2,3), (3,2), (1,4), (4,1), (4,4)}
- Antisymmetric: R = {(1,1), (1,2), (1,3), (1,4), (2,4), (3,3), (4,4)}

Symmetric and Antisymmetric Relations

R = {(1,1), (1,2), (2,1), (2,3), (3,2), (1,4), (4,1), (4,4)}

	1	2	3	4
1	X	X	X	X
2	X		X	
3	X	X		
4	X			X

Symmetric relation

- Diagonal axis of symmetry
- Not all elements on the axis of symmetry need to be included in the relation

R = {(1,1), (1,2), (1,3), (1,4), (2,4), (3,3), (4,4)}

	1	2	3	4
1	X	X	X	X
2				X
3			X	
4				X

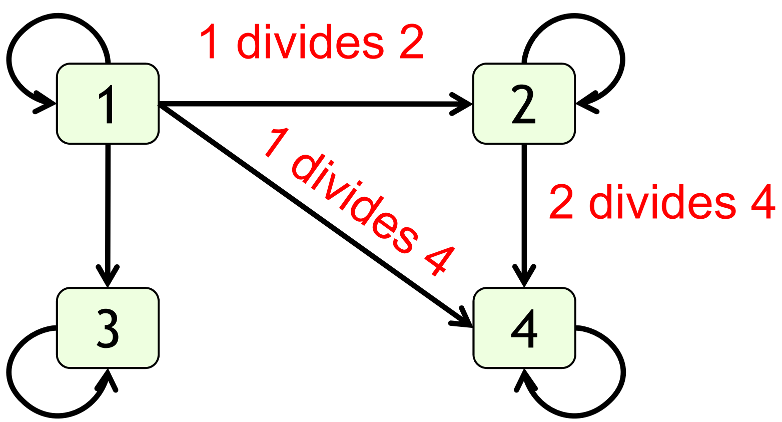
Asymmetric relation

- No axis of symmetry
- Only symmetry occurs on diagonal
- Not all elements on the diagonal need to be included in the relation

Properties of Relations

Definition: A relation R on a set A is transitive if whenver (a,b) ∈ R and (b,c) ∈ R, then (a,c) ∈ R for every a,b,c ∈ A.

Note: Our "divides" relation on the set A = {1,2,3,4} is transitive.



This isn't terribly interesting, but it is transitive nonetheless….

More common transitive relations include equality and comparison operators like <, >, ≤, and ≥.

Examples, redux

Question: Which of the following relations are reflexive, symmetric, antisymmetric, and/or transitive?

- $R_1 = \{(a,b) \mid a \leq b\}$
- $R_2 = \{(a,b) \mid a > b\}$
- $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}$
- $R_4 = \{(a,b) \mid a = b\}$
- $R_5 = \{(a,b) \mid a = b + 1\}$
- $R_6 = \{(a,b) \mid a + b \leq 3\}$

Answer:

Examples, redux

Question: Which of the following relations are reflexive, symmetric, antisymmetric, and/or transitive?

- $R_1 = \{(a,b) \mid a \leq b\}$
- $R_2 = \{(a,b) \mid a > b\}$
- $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}$
- $R_4 = \{(a,b) \mid a = b\}$
- $R_5 = \{(a,b) \mid a = b + 1\}$
- $R_6 = \{(a,b) \mid a + b \leq 3\}$

Answer:

	Reflexive	Symmetric	Antisymmetric	Transitive
R_1	Yes	No	Yes	Yes
R_2	No	No	Yes	Yes
R_3	Yes	Yes	No	Yes
R_4	Yes	Yes	Yes	Yes
R_5	No	No	Yes	No
R_6	No	Yes	No	No

Relations can be combined using set operations

Example: Let R be the relation that pairs students with courses that they have taken. Let S be the relation that pairs students with courses that they need to graduate. What do the relations R ∪ S, R ∩ S, and S – R represent?

Solution:

- R ∪ S = All pairs (a,b) where
  - student a has taken course b OR
  - student a needs to take course b to graduate
- R ∩ S = All pairs (a,b) where
  - Student a has taken course b AND
  - Student a needs course b to graduate
- S - R = All pairs (a,b) where
  - Student a needs to take course b to graduate BUT
  - Student a has not yet taken course b

Relations can be combined using functional composition

Definition: Let R be a relation from the set A to the set B, and S be a relation from the set B to the set C. The composite of R and S is the relation of ordered pairs (a, c), where a ∈ A and c ∈ C for which there exists an element b ∈ B such that (a, b) ∈ R and (b, c) ∈ S. We denote the composite of R and S by R ° S.

Example: What is the composite relation of R and S?

R: {1,2,3} → {1,2,3,4}

- R = {(1,1),(1,4),(2,3),(3,1),(3,4)}

S: {1,2,3,4} → {0,1,2}

- S = {(1,0),(2,0),(3,1),(3,2),(4,1)}

So: R ° S = {(1,0), (3,0), (1,1), (3,1), (2,1), (2,2)}

We can also "relate" elements of more than two sets

Definition: Let $A_1, A_2, \ldots, A_n$ be sets. An n-ary relation on these sets is a subset of $A_1 \times A_2 \times \ldots \times A_n$. The sets $A_1, A_2, \ldots, A_n$ are called the domains of the relation, and n is its degree.

Example: Let R be the relation on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$ consisting of triples (a, b, c) in which a,b,c form an arithmetic progression. That is $(a,b,c) \in R$ iff there exist some k integer such that $b=a+k$ and $c=a+2k$.

- What is the degree of this relation?
- What are the domains of this relation?
- Are the following tuples in this relation?
  - (1,3,5) ??
  - (2,5,9) ??

We can also “relate” elements of more than two sets

Definition: Let $A_1, A_2, \ldots, A_n$ be sets. An n-ary relation on these sets is a subset of $A_1 \times A_2 \times \ldots \times A_n$. The sets $A_1, A_2, \ldots, A_n$ are called the domains of the relation, and n is its degree.

Example: Let R be the relation on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$ consisting of triples (a, b, c) in which a,b,c form an arithmetic progression. That is $(a,b,c) \in R$ iff there exist some k integer such that b=a+k and c=a+2k.

- What is the degree of this relation? 3
- What are the domains of this relation? Ints, Ints, Ints
- Are the following tuples in this relation?
  - (1,3,5) $3=1+2$ and $5= 1+2*2$
  - (2,5,9) $5=2+3$ but $9 \neq 2+2*3$

N-ary relations are the basis of relational database management systems

Data is stored in relations (a.k.a., tables)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Columns of a table represent the attributes of a relation

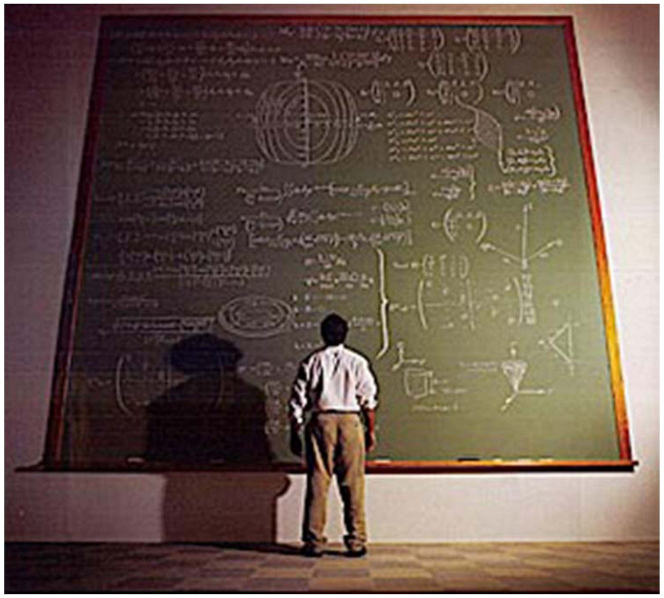
Rows, or records, contain the actual data defining the relation

Operations on an RDBMS are formally defined in terms of a relational algebra

Relational algebra gives a formal semantics to the operations performed on a database by rigorously defining these operations in terms of manipulations on sets of tuples (i.e., records)

Operators in relational algebra include:

- Selection ☆
- Projection ☆
- Rename
- Join
  - Equijoin ☆
  - Left outer join
  - Right outer join
  - …
- Aggregation



The selection operator allows us to filter the rows in a table

Definition: Let R be an n-ary relation and let C be a condition that elements in R must satisfy. The selection s_C maps the n-ary relation R to the n-ary relation of all n-tuples from R that satisfy the condition C.

Example: Consider the Students relation from earlier in lecture. Let the condition C1 be Major="CS" and let C2 be GPA > 2.5. What is the result of $s_{C1 \wedge C2}$(Students)?

Answer:

- (Alice, 334322, CS, 3.45)
- (Charlie, 045628, CS, 2.75)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

The projection operator allows us to consider only a subset of the columns of a table

Definition: The projection $P_{i1,\ldots,in}$ maps the n-tuple $(a_1, a_2, \ldots, a_n)$ to the m-tuple $(a_{i1}, \ldots, a_{im})$ where $m \leq n$

Example: What is the result of applying the projection $P_{1,3}$ to the Students table?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

→

Name	Major
Alice	CS
Bob	Math
Charlie	CS
Denise	Art

The equijoin operator allows us to create a new table based on data from two or more related tables

Definition: Let R be a relation of degree m and S be a relation of degree n. The equijoin $J_{i1=j1,\ldots,ik=jk}$, where $k \leq m$ and $k \leq n$, creates a new relation of degree m+n-k containing the subset of $S \times R$ in which $s_{i1} = r_{j1}, \ldots, s_{ik}=r_{jk}$ and duplicate columns are removed (via projection).

Example: What is the result of the equijoin $J_{2=1}$ on the Students and Enrollment tables?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

What is the result of the equijoin $J_{2=1}$ on the Students and Enrollment tables?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Name	ID	Major	GPA	Course
Alice	334322	CS	3.45	CS 441
Alice	334322	CS	3.45	Math 336
Bob	546346	Math	3.23	Math 422
Denise	964389	Art	4.0	Art 707

SQL queries correspond to statements in relational algebra

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Name	ID
Alice	334322
Charlie	045628

SELECT Name, ID FROM Students WHERE Major = “CS” AND GPA > 2.5

SELECT is actually a projection (in this case, $P_{1,2}$)

The WHERE clause lets us filter (i.e., $S_{major=\text{“CS”} \wedge GPA>2.5}$)

SQL: An Equijoin Example

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

SELECT Name, ID, Major, GPA, Course FROM Students, Enrollment WHERE ID = Stud_ID

Name	ID	Major	GPA	Course
Alice	334322	CS	3.45	CS 441
Alice	334322	CS	3.45	Math 336
Bob	546346	Math	3.23	Math 422
Denise	964389	Art	4.0	Art 707

Group Work!

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Problem 1: What is $P_{1,4}$(Students)?

Problem 2: What relational operators would you use to generate a table containing only the names of Math and CS majors with a GPA > 3.0?

Problem 3: Write an SQL statement corresponding to the solution to problem 2.