



Chapter 2: Intro to Relational Model

Database System Concepts, 6th Ed.

- Chapter 1: Introduction
- **Part 1: Relational databases**
 - [Chapter 2: Introduction to the Relational Model](#)
 - [Chapter 3: Introduction to SQL](#)
 - [Chapter 4: Intermediate SQL](#)
 - [Chapter 5: Advanced SQL](#)
 - [Chapter 6: Formal Relational Query Languages](#)
- **Part 2: Database Design**
 - [Chapter 7: Database Design: The E-R Approach](#)
 - [Chapter 8: Relational Database Design](#)
 - [Chapter 9: Application Design](#)
- **Part 3: Data storage and querying**
 - [Chapter 10: Storage and File Structure](#)
 - [Chapter 11: Indexing and Hashing](#)
 - [Chapter 12: Query Processing](#)
 - [Chapter 13: Query Optimization](#)
- **Part 4: Transaction management**
 - [Chapter 14: Transactions](#)
 - [Chapter 15: Concurrency control](#)
 - [Chapter 16: Recovery System](#)
- **Part 5: System Architecture**
 - [Chapter 17: Database System Architectures](#)
 - [Chapter 18: Parallel Databases](#)
 - [Chapter 19: Distributed Databases](#)
- **Part 6: Data Warehousing, Mining, and IR**
 - [Chapter 20: Data Mining](#)
 - [Chapter 21: Information Retrieval](#)
- **Part 7: Specialty Databases**
 - [Chapter 22: Object-Based Databases](#)
 - [Chapter 23: XML](#)
- **Part 8: Advanced Topics**
 - [Chapter 24: Advanced Application Development](#)
 - [Chapter 25: Advanced Data Types](#)
 - [Chapter 26: Advanced Transaction Processing](#)
- **Part 9: Case studies**
 - [Chapter 27: PostgreSQL](#)
 - [Chapter 28: Oracle](#)
 - [Chapter 29: IBM DB2 Universal Database](#)
 - [Chapter 30: Microsoft SQL Server](#)
- **Online Appendices**
 - [Appendix A: Detailed University Schema](#)
 - [Appendix B: Advanced Relational Database Model](#)
 - [Appendix C: Other Relational Query Languages](#)
 - [Appendix D: Network Model](#)
 - [Appendix E: Hierarchical Model](#)

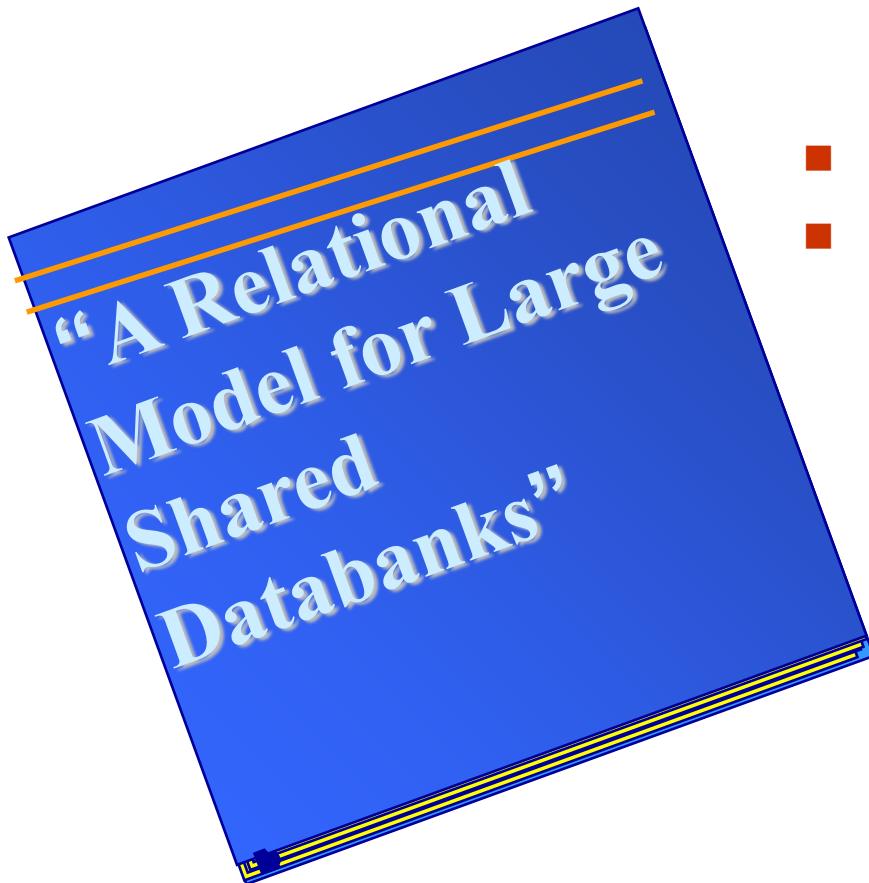


Chapter 2: Intro to Relational Model

- 2.1 Structure of Relational Databases
- 2.2 Database Schema
- 2.3 Keys
- 2.4 Schema Diagrams
- 2.5 Relational Query Languages
- 2.6 Relational Operations



In The Beginning...



- Everything in Table
- Set-oriented Query Language

E.F. Codd
-- 1970 CACM Paper
-- Turing Award

1970 →



Example of a Relation

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

attributes
(or columns)

tuples, records
(or rows)

- The set of allowed values for each attribute is called the **domain** of the attribute
- Attribute values are (normally) required to be **atomic**; that is, indivisible
- The special value **null** is a member of every domain
- The null value causes complications in the definition of many operations



Relation Schema and Instance

- A_1, A_2, \dots, A_n are *attributes*
- $R = (A_1, A_2, \dots, A_n)$ is a *relation schema*
Example: *instructor* = (*ID*, *name*, *dept_name*, *salary*)
- Formally, given sets D_1, D_2, \dots, D_n a **relation** r is a subset of $D_1 \times D_2 \times \dots \times D_n$
Thus, a relation is a set of n -tuples (a_1, a_2, \dots, a_n) where each $a_i \in D_i$
- The current values (**relation instance**) of a relation are specified by a table
- An element t of r is a *tuple*, represented by a *row* in a table



Relations are Unordered

- Order of tuples is irrelevant (tuples may be stored in an arbitrary order)
- Relations are basically set!
- Example: the *instructor* relation with unordered tuples

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000



Database

- A database consists of multiple relations
- Information about **an University enterprise** is broken up into parts

instructor

student

advisor

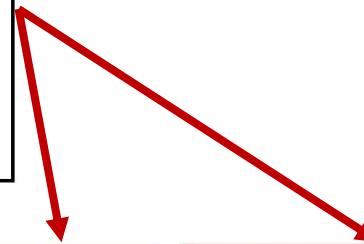
- Bad design:
university (instructor -ID, name, dept_name, salary, building, budget..)
results in
 - repetition of information (e.g., two students have the same instructor)
 - the need for null values (e.g., represent an student with no advisor)
- Normalization theory (Chapter 7) deals with how to design “good” relational schemas



University Relation

(instructor -ID, name, dept_name, salary, building, budget)

ID	name	salary	dept_name	building	budget
22222	Einstein	95000	Physics	Watson	70000
12121	Wu	90000	Finance	Painter	120000
32343	El Said	60000	History	Painter	50000
45565	Katz	75000	Comp. Sci.	Taylor	100000
98345	Kim	80000	Elec. Eng.	Taylor	85000
76766	Crick	72000	Biology	Watson	90000
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
83821	Brandt	92000	Comp. Sci	Taylor	100000
15151	Mozart	40000	Music	Packard	80000
33456	Gold	87000	Physics	Watson	70000
76543	Singh	80000	Finance	Painter	120000



ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

dept_name	building	budget
Comp. Sci.	Taylor	100000
Biology	Watson	90000
Elec. Eng.	Taylor	85000
Music	Packard	80000
Finance	Painter	120000
History	Painter	50000
Physics	Watson	70000

(a) The *instructor* table

(b) The *department* table

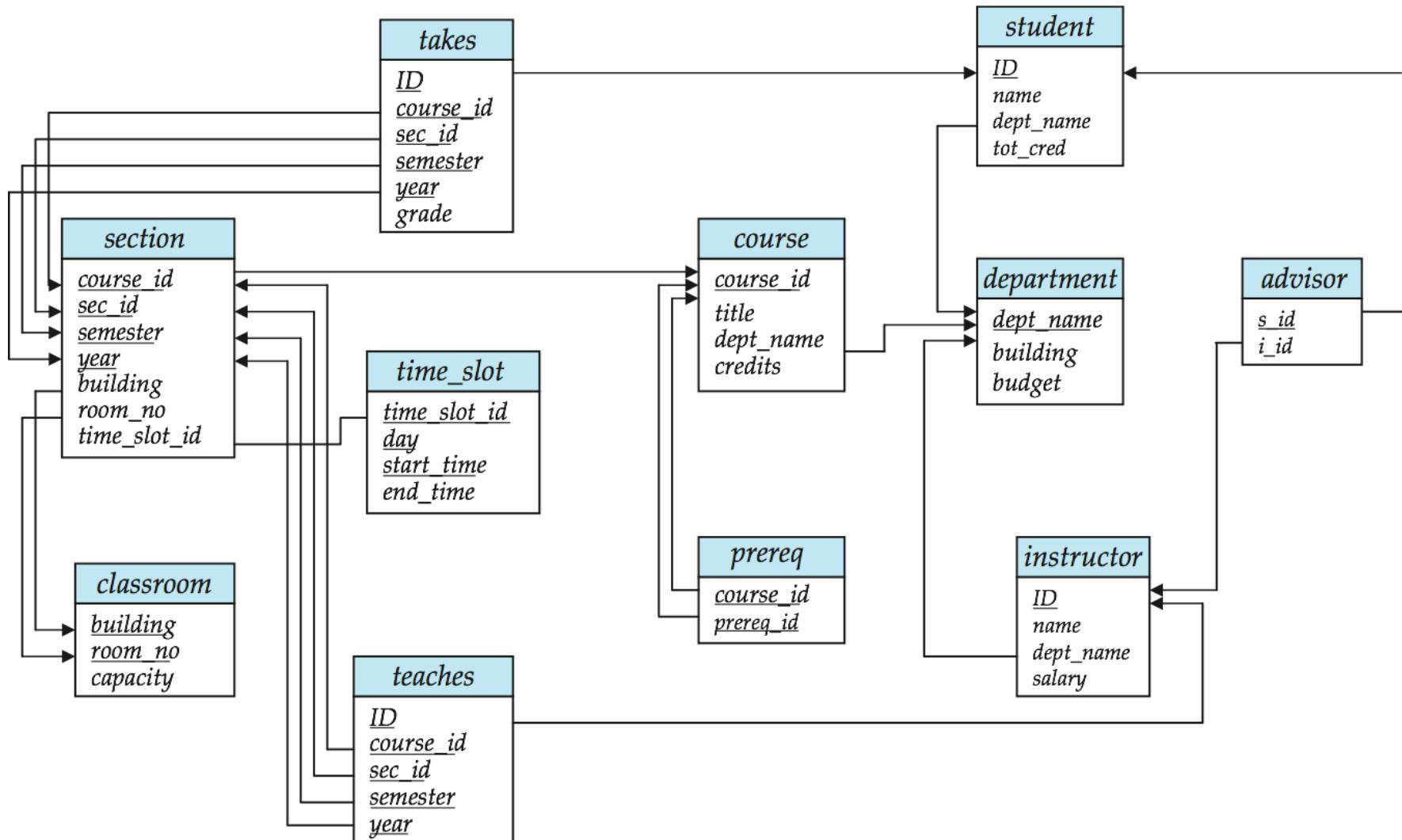


Keys

- Let $K \subseteq R$
- K is a **superkey** of R if values for K are sufficient to identify a unique tuple of each possible relation $r(R)$
 - Example: $\{ID\}$ and $\{ID, name\}$ are both superkeys of *instructor*.
- Superkey K is a **candidate key** if K is minimal
Example: $\{ID\}$ is a candidate key for *Instructor*
- One of the candidate keys is selected to be the **primary key**.
 - which one?
- **Foreign key** constraint: Value in one relation must appear in another
 - **Referencing** relation
 - **Referenced** relation



Schema Diagram for University Database





classroom(building, room_number, capacity)
department(dept_name, building, budget)
course(course_id, title, dept_name, credits)
instructor(ID, name, dept_name, salary)
section(course_id, sec_id, semester, year, building, room_number, time_slot_id)
teaches(ID, course_id, sec_id, semester, year)
student(ID, name, dept_name, tot_cred)
takes(ID, course_id, sec_id, semester, year, grade)
advisor(s_ID, i_ID)
time_slot(time_slot_id, day, start_time, end_time)
prereq(course_id, prereq_id)

Figure 2.9 Schema of the university database.



Figure 2.02: The Course relation

course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

Figure 2.05: The Department Relation

dept_name	building	budget
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

Figure 2.03: The Pre-requisite relation

course_id	prereq_id
BIO-301	BIO-101
BIO-399	BIO-101
CS-190	CS-101
CS-315	CS-101
CS-319	CS-101
CS-347	CS-101
EE-181	PHY-101

Figure 2.04: The Instructor relation

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000



Figure 2.06: The Section relation

course_id	sec_id	semester	year	building	room_number	time_slot_id
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

Figure 2.07: The Teaches relation

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009



Figure 2.10: selecting instructors with salary greater than \$85000

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
12121	Wu	Finance	90000
22222	Einstein	Physics	95000
33456	Gold	Physics	87000
83821	Brandt	Comp. Sci.	92000

Figure 2.13: selecting attributes ID and salary of instructors with salary greater than \$85000

<i>ID</i>	<i>salary</i>
12121	90000
22222	95000
33456	87000
83821	92000

Figure 2.11: selecting ID and salary attributes from the instructor relations

<i>ID</i>	<i>salary</i>
10101	65000
12121	90000
15151	40000
22222	95000
32343	60000
33456	87000
45565	75000
58583	62000
76543	80000
76766	72000
83821	92000
98345	80000

Figure 2.12: Natural Join of the instructor and department relations

<i>ID</i>	<i>name</i>	<i>salary</i>	<i>dept_name</i>	<i>building</i>	<i>budget</i>
10101	Srinivasan	65000	Comp. Sci.	Taylor	100000
12121	Wu	90000	Finance	Painter	120000
15151	Mozart	40000	Music	Packard	80000
22222	Einstein	95000	Physics	Watson	70000
32343	El Said	60000	History	Painter	50000
33456	Gold	87000	Physics	Watson	70000
45565	Katz	75000	Comp. Sci.	Taylor	100000
58583	Califieri	62000	History	Painter	50000
76543	Singh	80000	Finance	Painter	120000
76766	Crick	72000	Biology	Watson	90000
83821	Brandt	92000	Comp. Sci.	Taylor	100000
98345	Kim	80000	Elec. Eng.	Taylor	85000



Relational Query Languages

- Procedural vs. non-procedural (declarative)
- “Pure” formal query languages:
 - Relational algebra
 - Tuple relational calculus
 - Domain relational calculus
- Relational operators
 - Select
 - Project
 - Cartesian Product
 - Set Union
 - Set Minus



Selection (σ) of tuples

- Relation r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

- Select tuples with A=B and D > 5

- $\sigma_{A=B \text{ and } D > 5} (r)$

A	B	C	D
α	α	1	7
β	β	23	10



Projection (Π) of Columns (Attributes)

- Relation r :

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

- Select A and C
 - Projection
 - $\Pi_{A, C}(r)$

$$\begin{array}{c} \begin{array}{|c|c|} \hline A & C \\ \hline \alpha & 1 \\ \alpha & 1 \\ \beta & 1 \\ \beta & 2 \\ \hline \end{array} & = & \begin{array}{|c|c|} \hline A & C \\ \hline \alpha & 1 \\ \beta & 1 \\ \beta & 2 \\ \hline \end{array} \end{array}$$



Cartesian Product (x):

Cross-Product two relations

■ Relations r, s :

A	B
α	1
β	2

r

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

s

■ $r \times s$:

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b



Union (\cup) of two relations

■ Relations r, s :

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

■ $r \cup s$:

A	B
α	1
α	2
β	1
β	3



Set difference (-) of two relations

- Relations r, s :

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

- $r - s$:

A	B
α	1
β	1



Set Intersection (\cap) of two relations

- Relation r, s :

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

- $r \cap s$

A	B
α	2



Natural Join (\bowtie): Joining two relations

- Let r and s be relations on schemas R and S respectively.
- The “natural join” of relations R and S is a relation on schema $R \cup S$ obtained as follows:
 - Consider each pair of tuples t_r from r and t_s from s .
 - If t_r and t_s have the same value on each of the attributes in $R \cap S$, add a tuple t to the result, where
 - t has the same value as t_r on r
 - t has the same value as t_s on s

Relations r , s :

A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

r

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ε

s

- Natural Join
 - $r \bowtie s$

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	γ
δ	2	β	b	δ



Some of Relational Algebra Operators: Figure in-2.1

Symbol (Name)	Example of Use
σ (Selection)	$\sigma_{\text{salary} >= 85000}(\text{instructor})$ Return rows of the input relation that satisfy the predicate.
Π (Projection)	$\Pi_{ID, \text{salary}}(\text{instructor})$ Output specified attributes from all rows of the input relation. Remove duplicate tuples from the output.
\bowtie (Natural Join)	$\text{instructor} \bowtie \text{department}$ Output pairs of rows from the two input relations that have the same value on all attributes that have the same name.
\times (Cartesian Product)	$\text{instructor} \times \text{department}$ Output all pairs of rows from the two input relations (regardless of whether or not they have the same values on common attributes)
\cup (Union)	$\Pi_{name}(\text{instructor}) \cup \Pi_{name}(\text{student})$ Output the union of tuples from the two input relations.