# Database Systems

# Lecture 3:
# Intro to Relational Model

**Dr. Momtazi**
**momtazi@aut.ac.ir**

based on the slides of the course book

# Example of a Relation

attributes
(or columns)

| *ID* | *name* | *dept_name* | *salary* |
|---|---|---|---|
| 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 |

tuples
(or rows)

# Attribute Types

- 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. Indicated that the value is “unknown”
- The null value causes complications in the definition of many operations

# Relation Schema and Instance

- $A_1, A_2, \ldots, A_n$ are *attributes*

- $R = (A_1, A_2, \ldots, A_n)$ is a *relation schema*

  Example:

  *instructor* = (*ID*, *name*, *dept_name*, *salary*)

- Formally, given sets $D_1, D_2, \ldots. D_n$ a **relation *r*** is a subset of

  $D_1 \times D_2 \times \ldots \times D_n$

  Thus, a relation is a set of *n*-tuples $(a_1, a_2, \ldots, a_n)$ 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)
- Example: *instructor* relation with unordered tuples

| *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 |

# 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**.

- **Foreign key** constraint: Value in one relation must appear in another
  - **Referencing** relation
  - **Referenced** relation
  - Example – *dept_name* in *instructor* is a foreign key from *instructor* referencing *department*

# Schema Diagram for University Database

# Example: instructor relation

| *ID* | *name* | *dept_name* | *salary* |
|---|---|---|---|
| 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 |

**Figure 2.1** The *instructor* relation.

# Example: 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 2.2** The *course* relation.

# Example: prereq 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 2.3** The *prereq* relation.

# Example: 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 2.5** The *department* relation.

# Example: time_slot relation

| *time_slot_id* | *day* | *start_time* | *end_time* |
|---|---|---|---|
| A | M | 8:00 | 8:50 |
| A | W | 8:00 | 8:50 |
| A | F | 8:00 | 8:50 |
| B | M | 9:00 | 9:50 |
| B | W | 9:00 | 9:50 |
| B | F | 9:00 | 9:50 |
| C | M | 11:00 | 11:50 |
| C | W | 11:00 | 11:50 |
| C | F | 11:00 | 11:50 |
| D | M | 13:00 | 13:50 |
| D | W | 13:00 | 13:50 |
| D | F | 13:00 | 13:50 |
| E | T | 10:30 | 11:45 |
| E | R | 10:30 | 11:45 |
| F | T | 14:30 | 15:45 |
| F | R | 14:30 | 15:45 |
| G | M | 16:00 | 16:50 |
| G | W | 16:00 | 16:50 |
| G | F | 16:00 | 16:50 |
| H | W | 10:00 | 12:30 |

Apago PDF Enhancer

**Figure A.12** The *time_slot* relation.

# Example: 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 2.6** The *section* relation.

# Example: 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 2.7** The *teaches* relation.

# Relational Query Languages

- Procedural vs .non-procedural, or declarative
- “Pure” languages:
  - Relational algebra
  - Tuple relational calculus
  - Domain relational calculus
- The above 3 pure languages are equivalent in computing power
- We will concentrate in this chapter on relational algebra
  - Consists of 6 basic operations

# Select Operation
# selection of rows (tuples)

- Relation r

| $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|
| $\alpha$ | $\alpha$ | 1 | 7 |
| $\alpha$ | $\beta$ | 5 | 7 |
| $\beta$ | $\beta$ | 12 | 3 |
| $\beta$ | $\beta$ | 23 | 10 |

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

# Select Operation
# selection of rows (tuples)

- Relation r

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

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

| A | B | C | D |
|---|---|---|---|
| α | α | 1 | 7 |
| β | β | 23 | 10 |

# Example: select

| *ID* | *name* | *dept_name* | *salary* |
|---|---|---|---|
| 12121 | Wu | Finance | 90000 |
| 22222 | Einstein | Physics | 95000 |
| 33456 | Gold | Physics | 87000 |
| 83821 | Brandt | Comp. Sci. | 92000 |

**Figure 2.10** Result of query selecting *instructor* tuples with salary greater than $85000.

# Project Operation
# selection of columns (Attributes)

- Relation *r*:

| $A$ | $B$ | $C$ |
|---|---|---|
| $\alpha$ | 10 | 1 |
| $\alpha$ | 20 | 1 |
| $\beta$ | 30 | 1 |
| $\beta$ | 40 | 2 |

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

# Project Operation
# selection of columns (Attributes)

- Relation *r*:

| A | B | C |
|---|---|---|
| α | 10 | 1 |
| α | 20 | 1 |
| β | 30 | 1 |
| β | 40 | 2 |

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

| A | C |
|---|---|
| α | 1 |
| α | 1 |
| β | 1 |
| β | 2 |

=

| A | C |
|---|---|
| α | 1 |
| β | 1 |
| β | 2 |

# Example: project

| *ID* | *salary* |
|---|---|
| 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 2.11** Result of query selecting attributes *ID* and *salary* from the *instructor* relation.

# Example: select and project

| *ID* | *salary* |
|---|---|
| 12121 | 90000 |
| 22222 | 95000 |
| 33456 | 87000 |
| 83821 | 92000 |

**Figure 2.13** Result of selecting attributes *ID* and *salary* of instructors with salary greater than $85,000.

# Union of two relations

- Relations *r, s:*

| A | B |
|---|---|
| α | 1 |
| α | 2 |
| β | 1 |

*r*

| A | B |
|---|---|
| α | 2 |
| β | 3 |

*s*

- r ∪ s:

# Union 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 |
| α | 2 |
| β | 1 |
| β | 3 |

# Set difference of two relations

- Relations *r*, *s*:

| $A$ | $B$ |
|---|---|
| $\alpha$ | 1 |
| $\alpha$ | 2 |
| $\beta$ | 1 |

*r*

| $A$ | $B$ |
|---|---|
| $\alpha$ | 2 |
| $\beta$ | 3 |

*s*

- *r – s:*

# 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 of two relations

- Relation *r, s*:

| $A$ | $B$ |
|---|---|
| $\alpha$ | 1 |
| $\alpha$ | 2 |
| $\beta$ | 1 |

$r$

| $A$ | $B$ |
|---|---|
| $\alpha$ | 2 |
| $\beta$ | 3 |

$s$

- $r \cap s$

# Set intersection of two relations

- Relation *r, s*:

| $A$ | $B$ |
|---|---|
| $\alpha$ | 1 |
| $\alpha$ | 2 |
| $\beta$ | 1 |

*r*

| $A$ | $B$ |
|---|---|
| $\alpha$ | 2 |
| $\beta$ | 3 |

*s*

- $r \cap s$

| $A$ | $B$ |
|---|---|
| $\alpha$ | 2 |

# Joining two relations -- Cartesian-product

- 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* x *s*:

# Joining two relations -- Cartesian-product

- 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* x *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 |

# Cartesian-product – naming issue

- Relations *r, s*:

| $A$ | $B$ |
|---|---|
| $\alpha$ | 1 |
| $\beta$ | 2 |

*r*

| $B$ | $D$ | $E$ |
|---|---|---|
| $\alpha$ | 10 | a |
| $\beta$ | 10 | a |
| $\beta$ | 20 | b |
| $\gamma$ | 10 | b |

*s*

- *r* x *s*:

| $A$ | $r.B$ | $s.B$ | $D$ | $E$ |
|---|---|---|---|---|
| $\alpha$ | 1 | $\alpha$ | 10 | a |
| $\alpha$ | 1 | $\beta$ | 10 | a |
| $\alpha$ | 1 | $\beta$ | 20 | b |
| $\alpha$ | 1 | $\gamma$ | 10 | b |
| $\beta$ | 2 | $\alpha$ | 10 | a |
| $\beta$ | 2 | $\beta$ | 10 | a |
| $\beta$ | 2 | $\beta$ | 20 | b |
| $\beta$ | 2 | $\gamma$ | 10 | b |

# Renaming a Table

- Allows us to refer to a relation, (say E) by more than one name.

$$\rho_x(E)$$

returns the expression *E* under the name *X*

- Relations *r*

| *A* | *B* |
|---|---|
| α | 1 |
| β | 2 |

*r*

- $r \times \rho_s(r)$

| *r.A* | *r.B* | *s.A* | *s.B* |
|---|---|---|---|
| α | 1 | α | 1 |
| α | 1 | β | 2 |
| β | 2 | α | 1 |
| β | 2 | β | 2 |

# Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(r \times s)$
- $r \times s$

| 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 |

- $\sigma_{A=C}(r \times s)$

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

# Joining two relations – Natural Join

- Let $r$ and $s$ be relations on schemas $R$ and $S$ respectively.
  Then, 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 $s$

# Natural Join Example

- 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
  - $r \bowtie s$

$$\Pi_{A,\ r.B,\ C,\ r.D,\ E}\left(\sigma_{r.B = s.B \wedge r.D = s.D}\ (r \times s)\right)$$

# Natural Join Example

- 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 ⋈ 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 |

$$\Pi_{A,\ r.B,\ C,\ r.D,\ E}(\sigma_{r.B = s.B \wedge r.D = s.D}(r \times s)))$$

# Example: natural join

| *ID* | *name* | *salary* | *dept_name* | *building* | *budget* |
|---|---|---|---|---|---|
| 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 |

**Figure 2.12** Result of natural join of the *instructor* and *department* relations.

# Notes about Relational Languages

- Each Query input is a table (or set of tables)
- Each query output is a table.
- All data in the output table appears in one of the input tables
- Can we compute:
  - SUM
  - AVG
  - MAX
  - MIN

# Summary of Relational Algebra Operators

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

Questions?