



## CS425 – Fall 2013 Boris Glavic Chapter 3: Formal Relational Query Languages

Modified from:  
Database System Concepts, 6<sup>th</sup> Ed.  
©Silberschatz, Korth and Sudarshan  
See [www.db-book.com](http://www.db-book.com) for conditions on re-use



## Chapter 3: Formal Relational Query Languages

- Relational Algebra
- Tuple Relational Calculus
- Domain Relational Calculus



**Textbook: Chapter 6**

CS425 – Fall 2013 – Boris Glavic

3.2

©Silberschatz, Korth and Sudarshan



## Relational Algebra

- Procedural language
- Six basic operators
  - select:  $\sigma$
  - project:  $\Pi$
  - union:  $\cup$
  - set difference:  $-$
  - Cartesian product:  $\times$
  - rename:  $\rho$
- The operators take one or two relations as inputs and produce a new relation as a result.
  - **composable**

CS425 – Fall 2013 – Boris Glavic

3.3

©Silberschatz, Korth and Sudarshan



## Select Operation – Example

- Relation  $r$

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

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

A	B	C	D
$\alpha$	$\alpha$	1	7
$\beta$	$\beta$	23	10

CS425 – Fall 2013 – Boris Glavic

3.4

©Silberschatz, Korth and Sudarshan



## Select Operation

- Notation:  $\sigma_p(r)$
- $p$  is called the **selection predicate**
- Defined as:

$$\sigma_p(r) = \{t \mid t \in r \wedge p(t)\}$$

Where  $p$  is a formula in propositional calculus consisting of **terms** connected by  $\wedge$  (**and**),  $\vee$  (**or**),  $\neg$  (**not**)  
Each **term** is one of:

<attribute>  $op$  <attribute> or <constant>

where  $op$  is one of: =,  $\neq$ , >,  $\geq$ , <,  $\leq$

- Example of selection:

$\sigma_{dept\_name='Physics'}(instructor)$



CS425 – Fall 2013 – Boris Glavic

3.5

©Silberschatz, Korth and Sudarshan



## Project Operation – Example

- Relation  $r$ :

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

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

A	C
$\alpha$	1
$\alpha$	1
$\beta$	1
$\beta$	2

=

A	C
$\alpha$	1
$\beta$	1
$\beta$	2

CS425 – Fall 2013 – Boris Glavic

3.6

©Silberschatz, Korth and Sudarshan

## Project Operation

- Notation:  $\Pi_{A_1, A_2, \dots, A_k}(r)$
- where  $A_1, A_2$  are attribute names and  $r$  is a relation name.
- The result is defined as the relation of  $k$  columns obtained by erasing the columns that are not listed
- Duplicate rows removed from result, since relations are sets
- Let  $A$  be a subset of the attributes of relation  $r$  then:
 
$$\pi_A(r) = \{t.A \mid t \in r\}$$
- Example: To eliminate the *dept\_name* attribute of *instructor*

$$\Pi_{ID, name, salary}(instructor)$$

Heavy Warning

CS425 – Fall 2013 – Boris Glavic 3.7 ©Silberschatz, Korth and Sudarshan

## Union Operation – Example

- Relations  $r, s$ :
 

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

A	B
$\alpha$	2
$\beta$	3
- $r \cup s$ :
 

A	B
$\alpha$	1
$\alpha$	2
$\beta$	1
$\beta$	3

CS425 – Fall 2013 – Boris Glavic 3.8 ©Silberschatz, Korth and Sudarshan

## Union Operation

- Notation:  $r \cup s$
- Defined as:
 
$$r \cup s = \{t \mid t \in r \vee t \in s\}$$
- For  $r \cup s$  to be valid.
  1.  $r, s$  must have the **same arity** (same number of attributes)
  2. The attribute domains must be **compatible** (example: 2<sup>nd</sup> column of  $r$  deals with the same type of values as does the 2<sup>nd</sup> column of  $s$ )
- Example: to find all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or in both
 
$$\Pi_{course\_id}(\sigma_{semester="Fall" \wedge year=2009}(section)) \cup \Pi_{course\_id}(\sigma_{semester="Spring" \wedge year=2010}(section))$$

CS425 – Fall 2013 – Boris Glavic 3.9 ©Silberschatz, Korth and Sudarshan

## Set difference of two relations

- Relations  $r, s$ :
 

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

A	B
$\alpha$	2
$\beta$	3
- $r - s$ :
 

A	B
$\alpha$	1
$\beta$	1

CS425 – Fall 2013 – Boris Glavic 3.10 ©Silberschatz, Korth and Sudarshan

## Set Difference Operation

- Notation  $r - s$
- Defined as:
 
$$r - s = \{t \mid t \in r \wedge t \notin s\}$$
- Set differences must be taken between **compatible** relations.
  - $r$  and  $s$  must have the **same arity**
  - attribute domains of  $r$  and  $s$  must be compatible
- Example: to find all courses taught in the Fall 2009 semester, but not in the Spring 2010 semester
 
$$\Pi_{course\_id}(\sigma_{semester="Fall" \wedge year=2009}(section)) - \Pi_{course\_id}(\sigma_{semester="Spring" \wedge year=2010}(section))$$

CS425 – Fall 2013 – Boris Glavic 3.11 ©Silberschatz, Korth and Sudarshan

## Cartesian-Product Operation – Example

- Relations  $r, s$ :
 

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

C	D	E
$\alpha$	10	a
$\beta$	10	a
$\beta$	20	b
$\gamma$	10	b
- $r \times s$ :
 

A	B	C	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

CS425 – Fall 2013 – Boris Glavic 3.12 ©Silberschatz, Korth and Sudarshan

### Cartesian-Product Operation

- Notation  $r \times s$
- Defined as:
 
$$r \times s = \{t, t' \mid t \in r \wedge t' \in s\}$$
- Assume that attributes of  $r(R)$  and  $s(S)$  are disjoint. (That is,  $R \cap S = \emptyset$ ).
- If attributes of  $r(R)$  and  $s(S)$  are not disjoint, then renaming must be used.

CS425 – Fall 2013 – Boris Glavic 3.13 ©Silberschatz, Korth and Sudarshan

### Composition of Operations

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

$r \times s$

A	B	C	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

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

A	B	C	D	E
$\alpha$	1	$\alpha$	10	a
$\beta$	2	$\beta$	10	a
$\beta$	2	$\beta$	20	b

CS425 – Fall 2013 – Boris Glavic 3.14 ©Silberschatz, Korth and Sudarshan

### Rename Operation

- Allows us to name, and therefore to refer to, the results of relational-algebra expressions.
- Allows us to refer to a relation by more than one name.
- Example:
 
$$\rho_X(E)$$
 returns the expression  $E$  under the name  $X$
- If a relational-algebra expression  $E$  has arity  $n$ , then
 
$$\rho_{X(A_1, A_2, \dots, A_n)}(r)$$
 returns the result of expression  $E$  under the name  $X$ , and with the attributes renamed to  $A_1, A_2, \dots, A_n$ .
 
$$\rho_X(r) = \{t(X) \mid t \in r\}$$

$$\rho_{X(A)}(r) = \{t(X).A \mid t \in r\}$$

CS425 – Fall 2013 – Boris Glavic 3.15 ©Silberschatz, Korth and Sudarshan

### Example Query

- Find the largest salary in the university
  - Step 1: find instructor salaries that are less than some other instructor salary (i.e. not maximum)
    - using a copy of *instructor* under a new name  $d$ 

$$\pi_{instructor.salary}(\sigma_{instructor.salary < d.salary}(instructor \times \rho_d(instructor)))$$
  - Step 2: Find the largest salary
    - $$\pi_{salary}(instructor) - \pi_{instructor.salary}(\sigma_{instructor.salary < d.salary}(instructor \times \rho_d(instructor)))$$

CS425 – Fall 2013 – Boris Glavic 3.16 ©Silberschatz, Korth and Sudarshan

### Example Queries

- Find the names of all instructors in the Physics department, along with the *course\_id* of all courses they have taught
  - Query 1
 
$$\pi_{instructor.ID, course\_id}(\sigma_{dept\_name='Physics'}(\sigma_{instructor.ID=teaches.ID}(instructor \times teaches)))$$
  - Query 2
 
$$\pi_{instructor.ID, course\_id}(\sigma_{instructor.ID=teaches.ID}(\sigma_{dept\_name='Physics'}(instructor \times teaches)))$$


CS425 – Fall 2013 – Boris Glavic 3.17 ©Silberschatz, Korth and Sudarshan

### Formal Definition (Syntax)

- A basic expression in the relational algebra consists of either one of the following:
  - A relation in the database
  - A constant relation: e.g.,  $\{(1), (2)\}$
- Let  $E_1$  and  $E_2$  be relational-algebra expressions; the following are all relational-algebra expressions:
  - $E_1 \cup E_2$
  - $E_1 - E_2$
  - $E_1 \times E_2$
  - $\sigma_P(E_1)$ ,  $P$  is a predicate on attributes in  $E_1$
  - $\pi_S(E_1)$ ,  $S$  is a list consisting of some of the attributes in  $E_1$
  - $\rho_X(E_1)$ ,  $X$  is the new name for the result of  $E_1$


Warning

CS425 – Fall 2013 – Boris Glavic 3.18 ©Silberschatz, Korth and Sudarshan




## Formal Definition (Semantics)

- Let  $E$  be a relational algebra expression. We use  $[E](I)$  to denote the evaluation of  $E$  over a database instance  $I$ 
  - For simplicity we will often drop  $I$  and  $[\ ]$
- The result of evaluating a simple relational algebra expression  $E$  over a database is defined as
  - Simple relation:  $[R](I) = R(I)$
  - Constant relation:  $[C](I) = C$



CS425 – Fall 2013 – Boris Glavic 3.19 ©Silberschatz, Korth and Sudarshan



## Formal Definition (Semantics)

- Let  $E_1$  and  $E_2$  be relational-algebra expressions.
 
$$[E_1 \cup E_2] = \{t \mid t \in [E_1] \vee t \in [E_2]\}$$


$$[E_1 - E_2] = \{t \mid t \in [E_1] \wedge t \notin [E_2]\}$$

$$[E_1 \times E_2] = \{t, t' \mid t \in [E_1] \wedge t' \in [E_2]\}$$


$$[\sigma_p(E_1)] = \{t \mid t \in [E_1] \wedge p(t)\}$$

$$[\pi_A(E_1)] = \{t.A \mid t \in [E_1]\}$$

$$[\rho_X(E_1)] = \{t(X) \mid t \in [E_1]\}$$




CS425 – Fall 2013 – Boris Glavic 3.20 ©Silberschatz, Korth and Sudarshan



## Null Values

- It is possible for tuples to have a null value, denoted by *null*, for some of their attributes
- *null* signifies an unknown value or that a value does not exist.
- The result of any arithmetic expression involving *null* is *null*.
- Aggregate functions simply ignore null values (as in SQL)
- For duplicate elimination and grouping, null is treated like any other value, and two nulls are assumed to be the same (as in SQL)


CS425 – Fall 2013 – Boris Glavic 3.21 ©Silberschatz, Korth and Sudarshan



## Null Values

- Comparisons with null values return the special truth value: *unknown*
  - If *false* was used instead of *unknown*, then  $not(A < 5)$  would not be equivalent to  $A \geq 5$
- Three-valued logic using the truth value *unknown*:
  - OR:  $(unknown \text{ or } true) = true,$   
 $(unknown \text{ or } false) = unknown$   
 $(unknown \text{ or } unknown) = unknown$
  - AND:  $(true \text{ and } unknown) = unknown,$   
 $(false \text{ and } unknown) = false,$   
 $(unknown \text{ and } unknown) = unknown$
  - NOT:  $(not \text{ unknown}) = unknown$
  - In SQL "*P* is *unknown*" evaluates to true if predicate *P* evaluates to *unknown*
- Result of select predicate is treated as *false* if it evaluates to *unknown*

CS425 – Fall 2013 – Boris Glavic 3.22 ©Silberschatz, Korth and Sudarshan




## Additional Operations

We define additional operations that do not add any expressive power to the relational algebra, but that simplify common queries.

- Set intersection
- Natural join
- Assignment
- Outer join

CS425 – Fall 2013 – Boris Glavic 3.23 ©Silberschatz, Korth and Sudarshan



## Set-Intersection Operation

- Notation:  $r \cap s$
- Defined as:
 
$$r \cap s = \{t \mid t \in r \wedge t \in s\}$$
- Assume:
  - $r, s$  have the *same arity*
  - attributes of  $r$  and  $s$  are compatible
- Note:  $r \cap s = r - (r - s)$ 
  - That is adding intersection to the language does not make it more expressive

CS425 – Fall 2013 – Boris Glavic 3.24 ©Silberschatz, Korth and Sudarshan

### Set-Intersection Operation – Example

- Relation  $r, s$ :
 

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

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

$r$                        $s$

- $r \cap s$ 

A	B
$\alpha$	2

CS425 – Fall 2013 – Boris Glavic                      3.25                      ©Silberschatz, Korth and Sudarshan

### Natural-Join Operation

- Notation:  $r \bowtie s$
- Let  $r$  and  $s$  be relations on schemas  $R$  and  $S$  respectively. Then,  $r \bowtie s$  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  $s$
- Example:
  - $R = (A, B, C, D)$
  - $S = (E, B, D)$
  - Result schema =  $(A, B, C, D, E)$
  - $r \bowtie s$  is defined as:
 
$$\Pi_{r.A, r.B, r.C, r.D, s.E} (\sigma_{r.B = s.B \wedge r.D = s.D} (r \times s))$$

CS425 – Fall 2013 – Boris Glavic                      3.26                      ©Silberschatz, Korth and Sudarshan

### Natural-Join Operation (cont.)

- Let  $r$  and  $s$  be relations on schemas  $R$  and  $S$  respectively. Then,  $r \bowtie s$  is defined as:
 
$$X = R \cap S$$

$$S' = S - R$$

$$r \bowtie s = \pi_{R, S'} (\sigma_{r.X = s.X} (r \times s))$$

CS425 – Fall 2013 – Boris Glavic                      3.27                      ©Silberschatz, Korth and Sudarshan

### Natural Join Example

- Relations  $r, s$ :
 

A	B	C	D
$\alpha$	1	$\alpha$	a
$\beta$	2	$\gamma$	a
$\gamma$	4	$\beta$	b
$\alpha$	1	$\gamma$	a
$\delta$	2	$\beta$	b

B	D	E
1	a	$\alpha$
3	a	$\beta$
1	a	$\gamma$
2	b	$\delta$
3	b	$\epsilon$
- $r \bowtie s$ 

A	B	C	D	E
$\alpha$	1	$\alpha$	a	$\alpha$
$\alpha$	1	$\alpha$	a	$\gamma$
$\alpha$	1	$\gamma$	a	$\alpha$
$\alpha$	1	$\gamma$	a	$\gamma$
$\delta$	2	$\beta$	b	$\delta$

CS425 – Fall 2013 – Boris Glavic                      3.28                      ©Silberschatz, Korth and Sudarshan

### Natural Join and Theta Join

- Find the names of all instructors in the Comp. Sci. department together with the course titles of all the courses that the instructors teach
  - $\Pi_{name, title} (\sigma_{dept\_name = \text{"Comp. Sci."} (instructor \bowtie teaches))$
- Natural join is associative
  - $(instructor \bowtie teaches) \bowtie course$  is equivalent to  $instructor \bowtie (teaches \bowtie course)$
- Natural join is commutative (we ignore attribute order)
  - $instruct \bowtie teaches$  is equivalent to  $teaches \bowtie instructor$
- The **theta join** operation  $r \bowtie_{\theta} s$  is defined as
 
$$r \bowtie_{\theta} s = \sigma_{\theta} (r \times s)$$

CS425 – Fall 2013 – Boris Glavic                      3.29                      ©Silberschatz, Korth and Sudarshan

### Assignment Operation

- The assignment operation ( $\leftarrow$ ) provides a convenient way to express complex queries.
  - Write query as a sequential program consisting of
    - a series of assignments
    - followed by an expression whose value is displayed as a result of the query.
  - Assignment must always be made to a temporary relation variable.

$$E_1 \leftarrow \sigma_{salary > 40000} (instructor)$$

$$E_2 \leftarrow \sigma_{salary < 10000} (instructor)$$

$$E_3 \leftarrow E_1 \cup E_2$$

CS425 – Fall 2013 – Boris Glavic                      3.30                      ©Silberschatz, Korth and Sudarshan

## Outer Join

- An extension of the join operation that avoids loss of information.
- Computes the join and then adds tuples from one relation that does not match tuples in the other relation to the result of the join.
- Uses *null* values:
  - *null* signifies that the value is unknown or does not exist
  - All comparisons involving *null* are (roughly speaking) **false** by definition.
    - ▶ We shall study precise meaning of comparisons with nulls later

## Outer Join – Example

- Relation *instructor1*

ID	name	dept_name
10101	Srinivasan	Comp. Sci.
12121	Wu	Finance
15151	Mozart	Music

- Relation *teaches1*

ID	course_id
10101	CS-101
12121	FIN-201
76766	BIO-101

## Outer Join – Example

- Join  
*instructor* ⋈ *teaches*

ID	name	dept_name	course_id
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201

- Left Outer Join  
*instructor* ⋈<sub>L</sub> *teaches*

ID	name	dept_name	course_id
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
15151	Mozart	Music	<i>null</i>

## Outer Join – Example

- Right Outer Join  
*instructor* ⋈<sub>R</sub> *teaches*

ID	name	dept_name	course_id
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
76766	<i>null</i>	<i>null</i>	BIO-101

- Full Outer Join  
*instructor* ⋈<sub>F</sub> *teaches*

ID	name	dept_name	course_id
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
15151	Mozart	Music	<i>null</i>
76766	<i>null</i>	<i>null</i>	BIO-101

## Outer Join using Joins

- Outer join can be expressed using basic operations

$$r \bowtie_{L} s = (r \bowtie s) \cup ((r - \Pi_R(r \bowtie s)) \times \{(null, \dots, null)\})$$

$$r \bowtie_{R} s = (r \bowtie s) \cup (\{(null, \dots, null)\} \times (s - \Pi_S(r \bowtie s)))$$

$$r \bowtie_{F} s = (r \bowtie s) \cup ((r - \Pi_R(r \bowtie s)) \times \{(null, \dots, null)\}) \cup (\{(null, \dots, null)\} \times (s - \Pi_S(r \bowtie s)))$$

## Division Operator

- Given relations  $r(R)$  and  $s(S)$ , such that  $S \subseteq R$ ,  $r \div s$  is the largest relation  $t(R-S)$  such that  $t \times s \subseteq r$ 
  - Alternatively, all tuples from  $r$  (R-S) such that all their extensions on  $R \cap S$  with tuples from  $s$  exist in  $R$
- E.g. let  $r(ID, course\_id) = \Pi_{ID, course\_id}(takes)$  and  $s(course\_id) = \Pi_{course\_id}(\sigma_{dept\_name='Biology'}(course))$  then  $r \div s$  gives us students who have taken all courses in the Biology department
- Can write  $r \div s$  as

$$E_1 \leftarrow \Pi_{R-S}(r)$$

$$E_2 \leftarrow \Pi_{R-S}((E_1 \times s) - \Pi_{R-S,S}(r \bowtie s))$$

$$r \div s = E_1 - E_2$$

### Division Operator Example

- Return the name of all persons that read all newspapers

reads	
name	newspaper
Peter	Times
Bob	Wall Street
Alice	Times
Alice	Wall Street

newspaper
Times
Wall Street

$$E_1 \leftarrow \Pi_{name}(reads)$$

$$E_2 \leftarrow ((E_1 \times newspaper) - \Pi_{name,newspaper}(reads \bowtie newspaper))$$

$$reads \div newspaper = E_1 - E_2$$

$$\{reads \div newspaper\} = \{Alice\}$$

CS425 – Fall 2013 – Boris Glavic 3.37 ©Silberschatz, Korth and Sudarshan

### Extended Relational-Algebra-Operations

- Generalized Projection
- Aggregate Functions

CS425 – Fall 2013 – Boris Glavic 3.38 ©Silberschatz, Korth and Sudarshan

### Generalized Projection

- Extends the projection operation by allowing arithmetic functions to be used in the projection list.

$$\pi_{F_1, \dots, F_n}(E)$$

- $E$  is any relational-algebra expression
- Each of  $F_1, F_2, \dots, F_n$  are arithmetic expressions and function calls involving constants and attributes in the schema of  $E$ .
- Given relation  $instructor(ID, name, dept\_name, salary)$  where salary is annual salary, get the same information but with monthly salary

$$\Pi_{ID, name, dept\_name, salary/12}(instructor)$$

- Adding functions increases expressive power!
  - In standard relational algebra there is no way to change attribute values

CS425 – Fall 2013 – Boris Glavic 3.39 ©Silberschatz, Korth and Sudarshan

### Aggregate Functions and Operations

- Aggregation function** takes a set of values and returns a single value as a result.
  - avg**: average value
  - min**: minimum value
  - max**: maximum value
  - sum**: sum of values
  - count**: number of values
- Aggregate operation** in relational algebra

$$G_1, G_2, \dots, G_n \mathcal{G}_{F_1(A_1), F_2(A_2), \dots, F_n(A_n)}(E)$$

$E$  is any relational-algebra expression

- $G_1, G_2, \dots, G_n$  is a list of attributes on which to group (can be empty)
- Each  $F_j$  is an aggregate function
- Each  $A_j$  is an attribute name

- Note: Some books/articles use  $\gamma$  instead of  $\mathcal{G}$  (Calligraphic G)

CS425 – Fall 2013 – Boris Glavic 3.40 ©Silberschatz, Korth and Sudarshan

### Aggregate Operation – Example

- Relation  $r$ .

A	B	C
$\alpha$	$\alpha$	7
$\alpha$	$\beta$	7
$\beta$	$\beta$	3
$\beta$	$\beta$	10

- $\mathcal{G}_{sum(c)}(r)$

sum(c)
27

CS425 – Fall 2013 – Boris Glavic 3.41 ©Silberschatz, Korth and Sudarshan

### Aggregate Operation – Example


- Find the average salary in each department

$$dept\_name \mathcal{G}_{avg(salary)}(instructor)$$

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

dept_name	avg_salary
Biology	72000
Comp. Sci.	77333
Elec. Eng.	80000
Finance	85000
History	61000
Music	40000
Physics	91000


CS425 – Fall 2013 – Boris Glavic 3.42 ©Silberschatz, Korth and Sudarshan



## Aggregate Functions (Cont.)

- What are the names for attributes in aggregation results?
  - Need some convention!
    - ▶ E.g., use the expression as a name **avg(salary)**
  - For convenience, we permit renaming as part of aggregate operation
 
$$\text{dept\_name } \mathcal{G} \text{ avg(salary) as avg\_sal (instructor)}$$


CS425 – Fall 2013 – Boris Glavic 3.43 ©Silberschatz, Korth and Sudarshan



## Modification of the Database

- The content of the database may be modified using the following operations:
  - Deletion
  - Insertion
  - Updating
- All these operations can be expressed using the assignment operator
- Example: Delete instructors with salary over \$1,000,000
 
$$R \leftarrow R - (\sigma_{\text{salary} > 1000000}(R))$$

CS425 – Fall 2013 – Boris Glavic 3.44 ©Silberschatz, Korth and Sudarshan




## Restrictions for Modification

- Consider a modification where  $R=(A,B)$  and  $S=(C)$ 


$$R \leftarrow \sigma_{C > 5}(S)$$
- This would change the schema of R!
  - Should not be allowed
- Requirements for modifications
  - The name **R** on the left-hand side of the assignment operator refers to an existing relation in the database schema
  - The expression on the right-hand side of the assignment operator should be union-compatible with **R**

CS425 – Fall 2013 – Boris Glavic 3.45 ©Silberschatz, Korth and Sudarshan



## Tuple Relational Calculus


CS425 – Fall 2013 – Boris Glavic 3.46 ©Silberschatz, Korth and Sudarshan



## Tuple Relational Calculus

- A nonprocedural query language, where each query is of the form  $\{t \mid P(t)\}$
- It is the set of all tuples  $t$  such that predicate  $P$  is true for  $t$
- $t$  is a *tuple variable*,  $t[A]$  denotes the value of tuple  $t$  on attribute  $A$
- $t \in r$  denotes that tuple  $t$  is in relation  $r$
- $P$  is a *formula* similar to that of the predicate calculus

CS425 – Fall 2013 – Boris Glavic 3.47 ©Silberschatz, Korth and Sudarshan




## Predicate Calculus Formula

1. Set of attributes and constants
2. Set of comparison operators: (e.g.,  $<$ ,  $\leq$ ,  $=$ ,  $\neq$ ,  $>$ ,  $\geq$ )
3. Set of logical connectives: and ( $\wedge$ ), or ( $\vee$ ), not ( $\neg$ )
4. Implication ( $\Rightarrow$ ):  $x \Rightarrow y$ , if  $x$  if true, then  $y$  is true
 
$$x \Rightarrow y = \neg x \vee y$$
5. Set of quantifiers:
  - ▶  $\exists t \in r(Q(t))$  = "there exists" a tuple in  $t$  in relation  $r$  such that predicate  $Q(t)$  is true
  - ▶  $\forall t \in r(Q(t)) = Q$  is true "for all" tuples  $t$  in relation  $r$

CS425 – Fall 2013 – Boris Glavic 3.48 ©Silberschatz, Korth and Sudarshan






## Example Queries

- Find the *ID*, *name*, *dept\_name*, *salary* for instructors whose salary is greater than \$80,000
 
$$\{t \mid t \in \text{instructor} \wedge t[\text{salary}] > 80000\}$$
- As in the previous query, but output only the *ID* attribute value
 
$$\{t \mid \exists s \in \text{instructor} (t[ID] = s[ID] \wedge s[\text{salary}] > 80000)\}$$

Notice that a relation on schema (*ID*) is implicitly defined by the query, because

- t* is not bound to any relation by the predicate
- we implicitly state that *t* has an ID attribute ( $t[ID] = s[ID]$ )


CS425 – Fall 2013 – Boris Glavic 3.49 ©Silberschatz, Korth and Sudarshan



## Example Queries

- Find the names of all instructors whose department is in the Watson building
 
$$\{t \mid \exists s \in \text{instructor} (t[\text{name}] = s[\text{name}] \wedge \exists u \in \text{department} (u[\text{dept\_name}] = s[\text{dept\_name}] \wedge u[\text{building}] = \text{"Watson"}))\}$$
- Find the set of all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or both
 
$$\{t \mid \exists s \in \text{section} (t[\text{course\_id}] = s[\text{course\_id}] \wedge s[\text{semester}] = \text{"Fall"} \wedge s[\text{year}] = 2009) \vee \exists u \in \text{section} (t[\text{course\_id}] = u[\text{course\_id}] \wedge u[\text{semester}] = \text{"Spring"} \wedge u[\text{year}] = 2010)\}$$


CS425 – Fall 2013 – Boris Glavic 3.50 ©Silberschatz, Korth and Sudarshan



## Example Queries

- Find the set of all courses taught in the Fall 2009 semester, and in the Spring 2010 semester
 
$$\{t \mid \exists s \in \text{section} (t[\text{course\_id}] = s[\text{course\_id}] \wedge s[\text{semester}] = \text{"Fall"} \wedge s[\text{year}] = 2009) \wedge \exists u \in \text{section} (t[\text{course\_id}] = u[\text{course\_id}] \wedge u[\text{semester}] = \text{"Spring"} \wedge u[\text{year}] = 2010)\}$$
- Find the set of all courses taught in the Fall 2009 semester, but not in the Spring 2010 semester
 
$$\{t \mid \exists s \in \text{section} (t[\text{course\_id}] = s[\text{course\_id}] \wedge s[\text{semester}] = \text{"Fall"} \wedge s[\text{year}] = 2009) \wedge \neg \exists u \in \text{section} (t[\text{course\_id}] = u[\text{course\_id}] \wedge u[\text{semester}] = \text{"Spring"} \wedge u[\text{year}] = 2010)\}$$


CS425 – Fall 2013 – Boris Glavic 3.51 ©Silberschatz, Korth and Sudarshan



## Safety of Expressions

- It is possible to write tuple calculus expressions that generate infinite relations.
- For example,  $\{t \mid t \in r\}$  results in an infinite relation if the domain of any attribute of relation *r* is infinite
- To guard against the problem, we restrict the set of allowable expressions to safe expressions.
- An expression  $\{t \mid P(t)\}$  in the tuple relational calculus is *safe* if every component of *t* appears in one of the relations, tuples, or constants that appear in *P*.
  - NOTE: this is more than just a syntax condition.
    - E.g.  $\{t \mid t[A] = 5 \vee \text{true}\}$  is not safe --- it defines an infinite set with attribute values that do not appear in any relation or tuples or constants in *P*.


CS425 – Fall 2013 – Boris Glavic 3.52 ©Silberschatz, Korth and Sudarshan



## Universal Quantification

- Find all students who have taken all courses offered in the Biology department
  - $\{t \mid \exists r \in \text{student} (t[ID] = r[ID]) \wedge (\forall u \in \text{course} (u[\text{dept\_name}] = \text{"Biology"} \Rightarrow \exists s \in \text{takes} (t[ID] = s[ID] \wedge s[\text{course\_id}] = u[\text{course\_id}])))\}$
  - Note that without the existential quantification on student, the above query would be unsafe if the Biology department has not offered any courses.

CS425 – Fall 2013 – Boris Glavic 3.53 ©Silberschatz, Korth and Sudarshan



## Domain Relational Calculus

CS425 – Fall 2013 – Boris Glavic 3.54 ©Silberschatz, Korth and Sudarshan

## Domain Relational Calculus

- A nonprocedural query language equivalent in power to the tuple relational calculus
- Each query is an expression of the form:
 
$$\{ \langle x_1, x_2, \dots, x_n \rangle \mid P(x_1, x_2, \dots, x_n) \}$$
  - $x_1, x_2, \dots, x_n$  represent domain variables
    - ▶ Variables that range of attribute values
  - $P$  represents a formula similar to that of the predicate calculus
  - Tuples can be formed using  $\langle \rangle$ 
    - ▶ E.g.,  $\langle \text{'Einstein'}, \text{'Physics'} \rangle$

CS425 – Fall 2013 – Boris Glavic 3.55 ©Silberschatz, Korth and Sudarshan

## Example Queries

- Find the  $ID, name, dept\_name, salary$  for instructors whose salary is greater than \$80,000
  - $\{ \langle i, n, d, s \rangle \mid \langle i, n, d, s \rangle \in instructor \wedge s > 80000 \}$
- As in the previous query, but output only the  $ID$  attribute value
  - $\{ \langle i \rangle \mid \langle i, n, d, s \rangle \in instructor \wedge s > 80000 \}$
- Find the names of all instructors whose department is in the Watson building
 
$$\{ \langle n \rangle \mid \exists i, d, s (\langle i, n, d, s \rangle \in instructor \wedge \exists b, a (\langle d, b, a \rangle \in department \wedge b = \text{'Watson'})) \}$$

CS425 – Fall 2013 – Boris Glavic 3.56 ©Silberschatz, Korth and Sudarshan

## Example Queries

- Find the set of all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or both
 
$$\{ \langle c \rangle \mid \exists a, s, y, b, r, t (\langle c, a, s, y, b, t \rangle \in section \wedge s = \text{'Fall'} \wedge y = 2009) \vee \exists a, s, y, b, r, t (\langle c, a, s, y, b, t \rangle \in section \wedge s = \text{'Spring'} \wedge y = 2010) \}$$

This case can also be written as

$$\{ \langle c \rangle \mid \exists a, s, y, b, r, t (\langle c, a, s, y, b, t \rangle \in section \wedge ((s = \text{'Fall'} \wedge y = 2009) \vee (s = \text{'Spring'} \wedge y = 2010))) \}$$
- Find the set of all courses taught in the Fall 2009 semester, and in the Spring 2010 semester
 
$$\{ \langle c \rangle \mid \exists a, s, y, b, r, t (\langle c, a, s, y, b, t \rangle \in section \wedge s = \text{'Fall'} \wedge y = 2009) \wedge \exists a, s, y, b, r, t (\langle c, a, s, y, b, t \rangle \in section \wedge s = \text{'Spring'} \wedge y = 2010) \}$$

CS425 – Fall 2013 – Boris Glavic 3.57 ©Silberschatz, Korth and Sudarshan

## Safety of Expressions

The expression:

$$\{ \langle x_1, x_2, \dots, x_n \rangle \mid P(x_1, x_2, \dots, x_n) \}$$

is safe if all of the following hold:

1. All values that appear in tuples of the expression are values from  $dom(P)$  (that is, the values appear either as constants in  $P$  or in a tuple of a relation mentioned in  $P$ ).
2. For every "there exists" subformula of the form  $\exists x (P_1(x))$ , the subformula is true if and only if there is a value of  $x$  in  $dom(P_1)$  such that  $P_1(x)$  is true.
3. For every "for all" subformula of the form  $\forall x (P_1(x))$ , the subformula is true if and only if  $P_1(x)$  is true for all values  $x$  from  $dom(P_1)$ .

CS425 – Fall 2013 – Boris Glavic 3.58 ©Silberschatz, Korth and Sudarshan

## Universal Quantification

- Find all students who have taken all courses offered in the Biology department
  - $\{ \langle i \rangle \mid \exists n, d, tc (\langle i, n, d, tc \rangle \in student \wedge (\forall ci, ti, dn, cr (\langle ci, ti, dn, cr \rangle \in course \wedge dn = \text{'Biology'} \Rightarrow \exists si, se, y, g (\langle ci, ci, si, se, y, g \rangle \in takes))) \}$
  - Note that without the existential quantification on student, the above query would be unsafe if the Biology department has not offered any courses.


\* Above query fixes bug in page 246, last query

CS425 – Fall 2013 – Boris Glavic 3.59 ©Silberschatz, Korth and Sudarshan

## Relationship between Relational Algebra and Tuple (Domain) Calculus

- **Codd's theorem**
  - Relational algebra and tuple calculus are equivalent
- That means that every query expressible in relational algebra can also be expressed in tuple calculus and vice versa
- Since domain calculus is as expressive as tuple calculus the same holds for the domain calculus
- Note: Here relational algebra refers to the standard version (no aggregation and projection with functions)


CS425 – Fall 2013 – Boris Glavic 3.60 ©Silberschatz, Korth and Sudarshan



## Recap


- **Relational algebra**
  - Standard relational algebra:
    - Selection, projection, renaming, cross product, union, set difference
  - Null values
  - Semantic sugar operators:
    - Intersection, joins, division,
  - Extensions:
    - Aggregation, extended projection
- **Tuple Calculus**
  - safety
- **Domain Calculus**

CS425 – Fall 2013 – Boris Glavic 3.61 ©Silberschatz, Korth and Sudarshan



## End of Chapter 3


Modified from:  
Database System Concepts, 6<sup>th</sup> Ed.  
©Silberschatz, Korth and Sudarshan  
See [www.db-book.com](http://www.db-book.com) for conditions on re-use



## Outline

- Introduction
- Relational Data Model
- Formal Relational Languages (relational algebra)
- **SQL - Introduction**
- Database Design
- Transaction Processing, Recovery, and Concurrency Control
- Storage and File Structures
- Indexing and Hashing
- Query Processing and Optimization


CS425 – Fall 2013 – Boris Glavic 3.63 ©Silberschatz, Korth and Sudarshan



## Figure 6.01

<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	Califrieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000


CS425 – Fall 2013 – Boris Glavic 3.64 ©Silberschatz, Korth and Sudarshan



## Figure 6.02

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

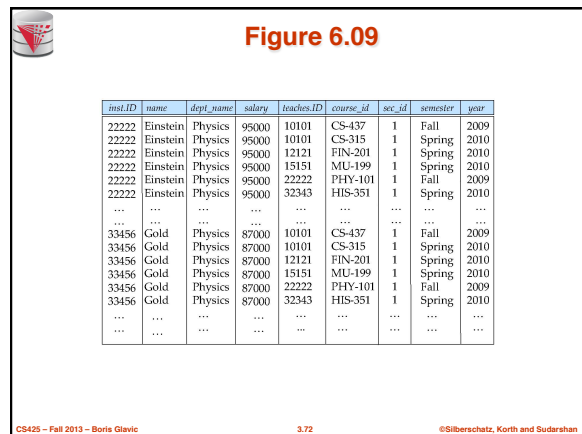
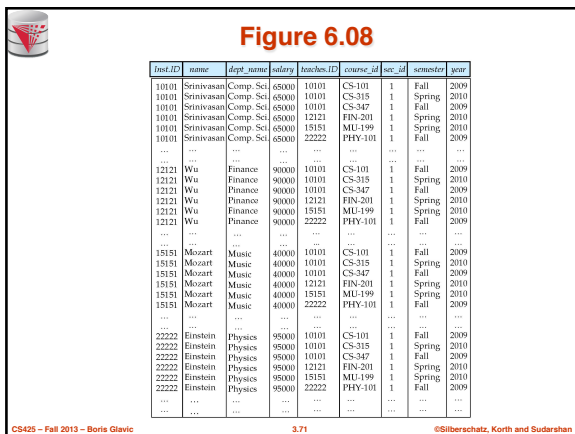
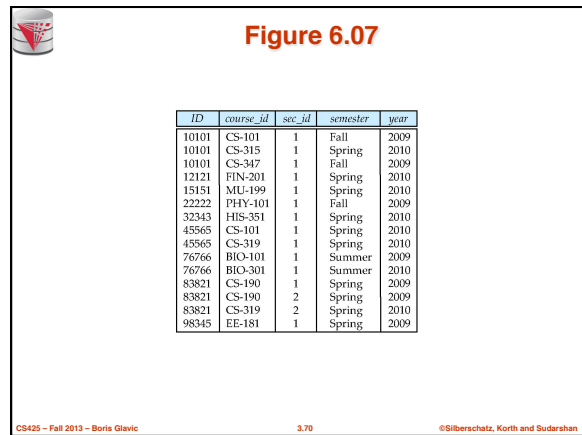
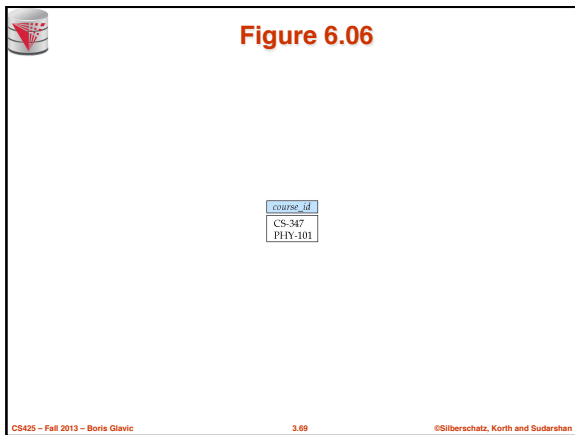
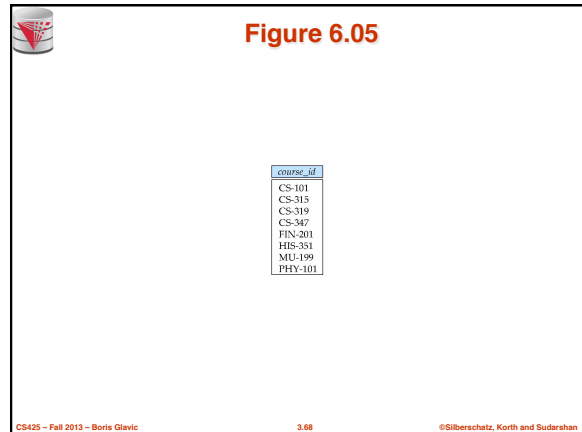
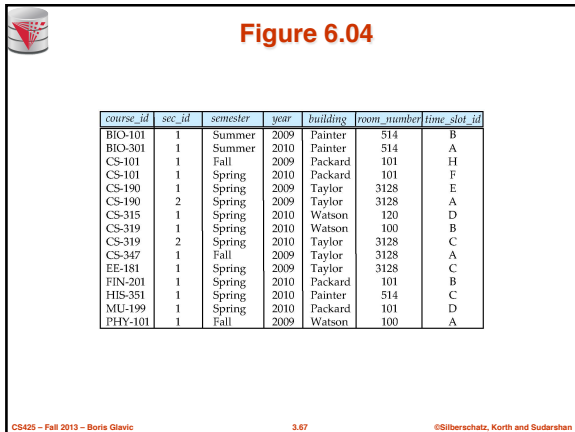
CS425 – Fall 2013 – Boris Glavic 3.65 ©Silberschatz, Korth and Sudarshan

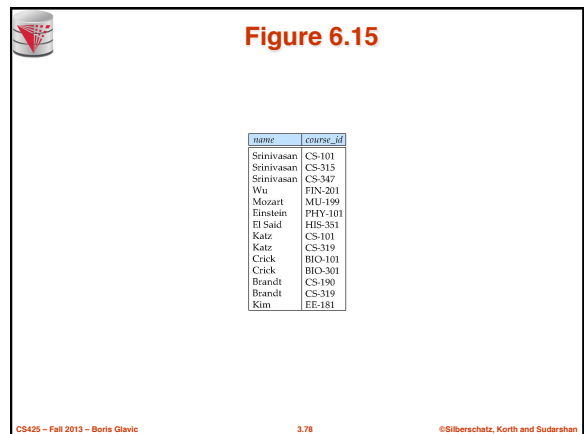
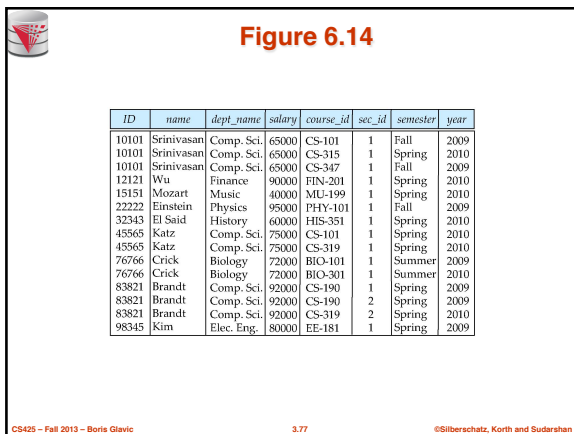
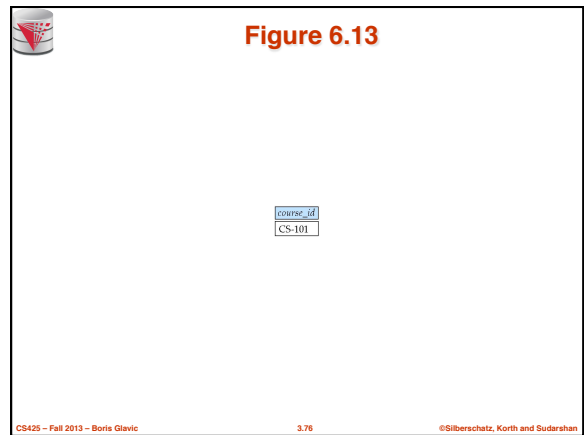
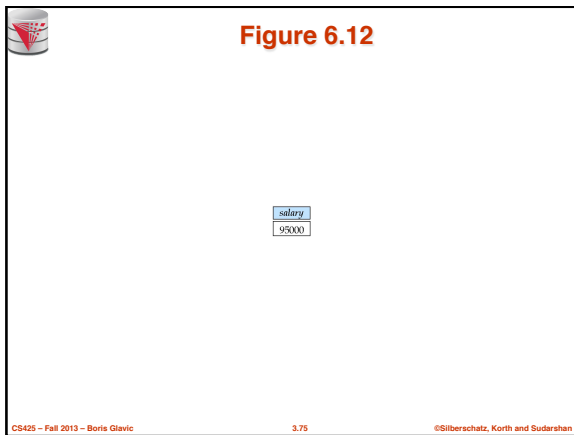
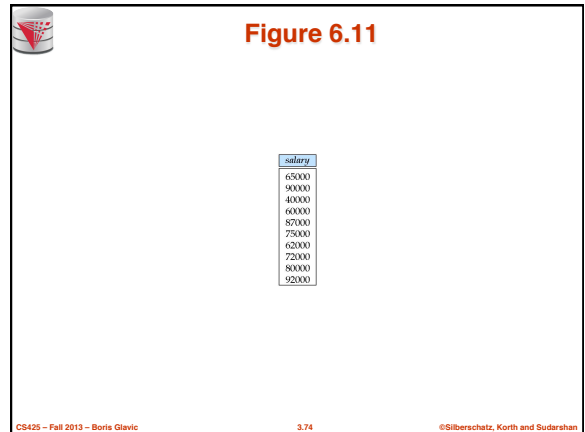
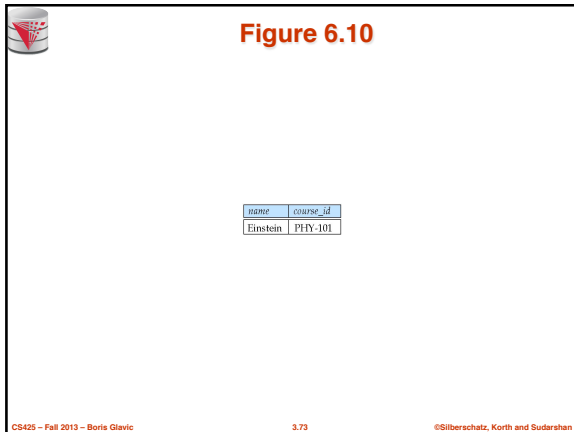


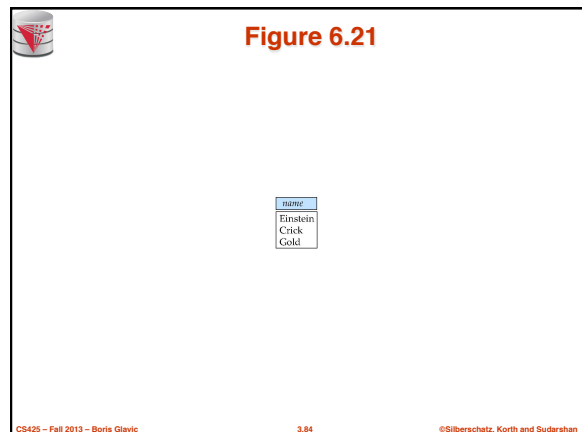
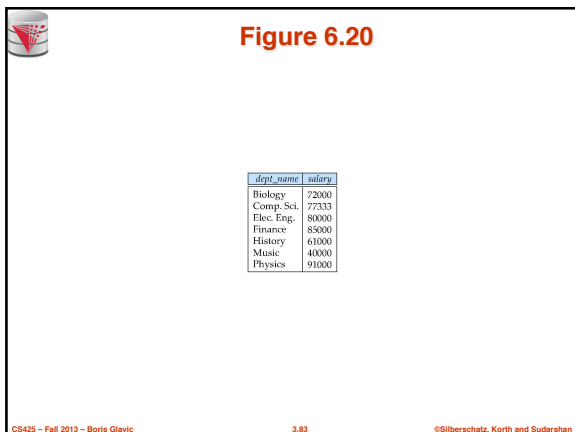
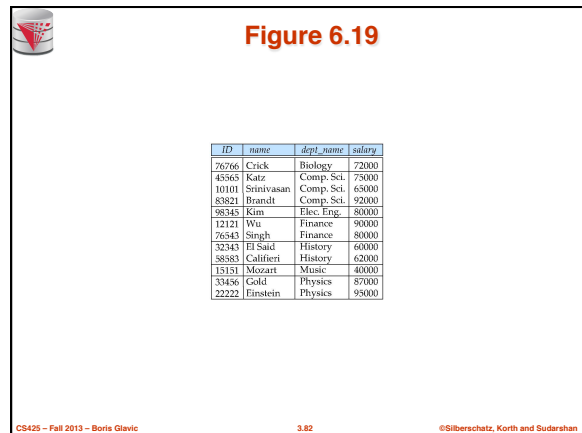
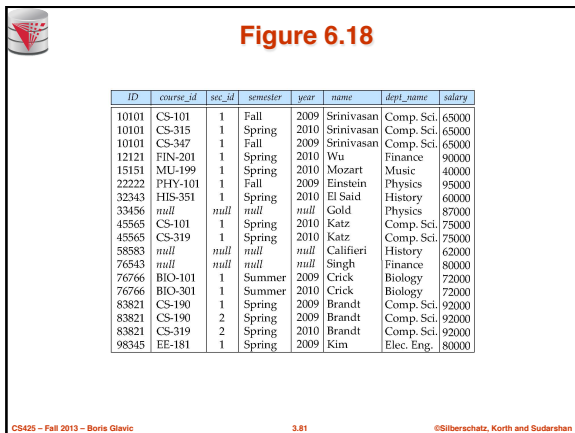
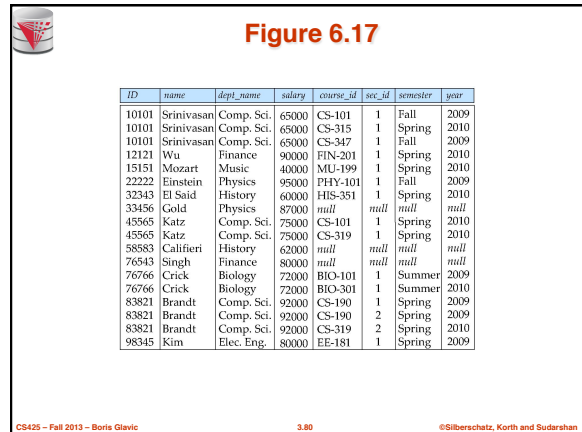
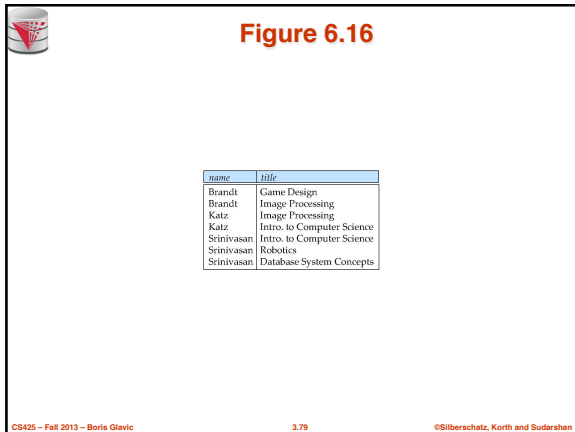
## Figure 6.03


<i>ID</i>	<i>name</i>	<i>salary</i>
10101	Srinivasan	65000
12121	Wu	90000
15151	Mozart	40000
22222	Einstein	95000
32343	El Said	60000
33456	Gold	87000
45565	Katz	75000
58583	Califrieri	62000
76543	Singh	80000
76766	Crick	72000
83821	Brandt	92000
98345	Kim	80000

CS425 – Fall 2013 – Boris Glavic 3.66 ©Silberschatz, Korth and Sudarshan










## Deletion

- A delete request is expressed similarly to a query, except instead of displaying tuples to the user, the selected tuples are removed from the database.
- Can delete only whole tuples; cannot delete values on only particular attributes
- A deletion is expressed in relational algebra by:
 
$$r \leftarrow r - E$$
 where  $r$  is a relation and  $E$  is a relational algebra query.

CS425 – Fall 2013 – Boris Glavic 3.85 ©Silberschatz, Korth and Sudarshan



## Deletion Examples

- Delete all account records in the Perryridge branch.
 
$$account \leftarrow account - \sigma_{branch\_name = "Perryridge"}(account)$$
- Delete all loan records with amount in the range of 0 to 50
 
$$loan \leftarrow loan - \sigma_{amount \geq 0 \text{ and } amount \leq 50}(loan)$$
- Delete all accounts at branches located in Needham.
 
$$r_1 \leftarrow \sigma_{branch\_city = "Needham"}(account \bowtie branch)$$


$$r_2 \leftarrow \pi_{account\_number, branch\_name, balance}(r_1)$$

$$r_3 \leftarrow \pi_{customer\_name, account\_number}(r_2 \bowtie depositor)$$

$$account \leftarrow account - r_2$$

$$depositor \leftarrow depositor - r_3$$


CS425 – Fall 2013 – Boris Glavic 3.86 ©Silberschatz, Korth and Sudarshan



## Insertion

- To insert data into a relation, we either:
  - specify a tuple to be inserted
  - write a query whose result is a set of tuples to be inserted
- in relational algebra, an insertion is expressed by:
 
$$r \leftarrow r \cup E$$
 where  $r$  is a relation and  $E$  is a relational algebra expression.
- The insertion of a single tuple is expressed by letting  $E$  be a constant relation containing one tuple.

CS425 – Fall 2013 – Boris Glavic 3.87 ©Silberschatz, Korth and Sudarshan



## Insertion Examples


- Insert information in the database specifying that Smith has \$1200 in account A-973 at the Perryridge branch.
 
$$account \leftarrow account \cup \{("A-973", "Perryridge", 1200)\}$$

$$depositor \leftarrow depositor \cup \{("Smith", "A-973")\}$$
- Provide as a gift for all loan customers in the Perryridge branch, a \$200 savings account. Let the loan number serve as the account number for the new savings account.
 
$$r_1 \leftarrow (\sigma_{branch\_name = "Perryridge"}(borrower \bowtie loan))$$

$$account \leftarrow account \cup \pi_{loan\_number, branch\_name, 200}(r_1)$$

$$depositor \leftarrow depositor \cup \pi_{customer\_name, loan\_number}(r_1)$$


CS425 – Fall 2013 – Boris Glavic 3.88 ©Silberschatz, Korth and Sudarshan



## Updating

- A mechanism to change a value in a tuple without changing *all* values in the tuple
- Use the generalized projection operator to do this task
 
$$r \leftarrow \pi_{F_1, F_2, \dots, F_i}(r)$$
- Each  $F_j$  is either
  - the  $j^{\text{th}}$  attribute of  $r$ , if the  $j^{\text{th}}$  attribute is not updated, or,
  - if the attribute is to be updated  $F_j$  is an expression, involving only constants and the attributes of  $r$ , which gives the new value for the attribute

CS425 – Fall 2013 – Boris Glavic 3.89 ©Silberschatz, Korth and Sudarshan



## Update Examples

- Make interest payments by increasing all balances by 5 percent.
 
$$account \leftarrow \pi_{account\_number, branch\_name, balance * 1.05}(account)$$
- Pay all accounts with balances over \$10,000 6 percent interest and pay all others 5 percent
 
$$account \leftarrow \pi_{account\_number, branch\_name, balance * 1.06}(\sigma_{BAL > 10000}(account)) \cup \pi_{account\_number, branch\_name, balance * 1.05}(\sigma_{BAL \leq 10000}(account))$$

CS425 – Fall 2013 – Boris Glavic 3.90 ©Silberschatz, Korth and Sudarshan



### Example Queries

- Find the names of all customers who have a loan and an account at bank.

$$\Pi_{customer\_name} (borrower) \cap \Pi_{customer\_name} (depositor)$$

- Find the name of all customers who have a loan at the bank and the loan amount

$$\Pi_{customer\_name, loan\_number, amount} (borrower \bowtie loan)$$



### Example Queries

- Find all customers who have an account from at least the "Downtown" and the Uptown" branches.

- Query 1

$$\Pi_{customer\_name} (\sigma_{branch\_name = "Downtown"} (depositor \bowtie account)) \cap$$

$$\Pi_{customer\_name} (\sigma_{branch\_name = "Uptown"} (depositor \bowtie account))$$

- Query 2

$$\Pi_{customer\_name, branch\_name} (depositor \bowtie account) \div \rho_{temp(branch\_name)} (\{ ("Downtown"), ("Uptown") \})$$

Note that Query 2 uses a constant relation.



### Bank Example Queries

- Find all customers who have an account at all branches located in Brooklyn city.

$$\Pi_{customer\_name, branch\_name} (depositor \bowtie account) \div \Pi_{branch\_name} (\sigma_{branch\_city = "Brooklyn"} (branch))$$