

Formal Relational Query Languages

Chapter - 6

Instructor: Nítesh Kumar Jha níteshkumarjha@soa.ac.ín

ITER,S'O'A(DEEMED TO BE UNIVERSITY)

May 2024

Outline

- Relational Algebra
- Relational Calculus
- I. Tuple Relational Calculus
- II. Domain Relational Calculus

Relational Algebra

- Procedural language
- Six basic operators
 - select: σ
 - project: ∏
 - union: ∪
 - set difference: –
 - Cartesian product: x
 - rename: ρ
- The operators take one or two relations as inputs and produce a new relation as a result.

Select Operation

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

$$\sigma_p(\mathbf{r}) = \{t \mid t \in r \text{ and } p(t)\}$$

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

<attribute> op <attribute> or <constant> where op is one of: =, \neq , >, \geq . <. \leq

Example of selection:

Project Operation

Notation:

$$\prod_{A_1,A_2,\ldots,A_k}(r)$$

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
- Example: To eliminate the *dept_name* attribute of *instructor*

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

Union Operation

- Notation: $r \cup s$
- Defined as:

$$r \cup s = \{t \mid t \in r \text{ or } t \in s\}$$

- 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^{nd} column of r deals with the same type of values as does the 2^{nd} column of s)

course_id	sec_id	semester	year	building	room_number	time_slot_id
BIO-101	1	Summer	2009	Painter	514	В
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	В
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	С
FIN-201	1	Spring	2010	Packard	101	В
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 6.4 The section relation.

Example: to find all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or in both

$$\Pi_{course_id}(\sigma_{semester="Fall"} \land year=2009(section)) \cup$$

$$\Pi_{course_id}(\sigma_{semester="Spring"} \land year=2010 (section))$$

Set Difference Operation

- Notation r s
- Defined as:

$$r-s = \{t \mid t \in r \text{ and } t \notin s\}$$

- 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

$$\Pi_{course_id}(\sigma_{semester="Fall"} \land year=2009(section)) - \Pi_{course_id}(\sigma_{semester="Spring"} \land year=2010(section))$$



Figure 6.6 Courses offered in the Fall 2009 semester but not in Spring 2010 semester.

Set-Intersection Operation

- Notation: $r \cap s$
- Defined as:
- $r \cap s = \{ t \mid t \in r \text{ and } t \in s \}$
- Assume:
 - r, s have the same arity
 - attributes of r and s are compatible
- Note: $r \cap s = r (r s)$

Cartesian-Product Operation

- Notation r x s
- Defined as:

$$r \times s = \{t \mid q \mid t \in r \text{ and } q \in s\}$$

- 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

inst.ID	name	dept_name	salary	teaches.ID	course_id	coc id	semester	year
10101	Srinivasan		95000	10101	CS-101	1	Fall	2009
10101	Srinivasan	Physics	95000	10101	CS-101 CS-315	1		2009
10101		Physics	95000	10101	CS-315 CS-347	1	Spring Fall	2010
	Srinivasan	Physics				1		
10101	Srinivasan	Physics	95000	10101	FIN-201	1	Spring	2010
10101	Srinivasan	Physics	95000	15151	MU-199	1	Spring Fall	2010
10101	Srinivasan	Physics	95000	22222	PHY-101	1	rall	2009
10101	 TA7			10101			 E 11	2000
12121	Wu	Physics	95000	10101	CS-101	1	Fall	2009
12121	Wu	Physics	95000	10101	CS-315	1	Spring	2010
12121	Wu	Physics	95000	10101	CS-347	1	Fall	2009
12121	Wu	Physics	95000	10101	FIN-201	1	Spring	2010
12121	Wu	Physics	95000	15151	MU-199	1	Spring	2010
12121	Wu	Physics	95000	22222	PHY-101	1	Fall	2009
15151	Mozart	Physics	95000	10101	CS-101	1	Fall	2009
15151	Mozart	Physics	95000	10101	CS-315	1	Spring	2010
15151	Mozart	Physics	95000	10101	CS-347	1	Fall	2009
15151	Mozart	Physics	95000	10101	FIN-201	1	Spring	2010
15151	Mozart	Physics	95000	15151	MU-199	1	Spring	2010
15151	Mozart	Physics	95000	22222	PHY-101	1	Fall	2009
22222	Einstein	Physics	95000	10101	CS-101	1	Fall	2009
22222	Einstein	Physics	95000	10101	CS-315	1	Spring	2010
22222	Einstein	Physics	95000	10101	CS-347	1	Fall	2009
22222	Einstein	Physics	95000	10101	FIN-201	1	Spring	2010
22222	Einstein	Physics	95000	15151	MU-199	1	Spring	2010
22222	Einstein	Physics	95000	22222	PHY-101	1	Fall	2009
		•••				•••		

Figure 6.8 Result of instructor × teaches.

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

 $\rho_{x(A_1,A_2,...,A_n)}(E)$ returns the result of expression E under the name X, and with the attributes renamed to A_1 , A_2 ,, A_n .

The query to find the largest salary in the university can be written as:

$$\Pi_{salary}$$
 (instructor) – $\Pi_{instructor.salary}$ ($\sigma_{instructor.salary}$ (instructor $\times \rho_d$ (instructor)))

Formal Definition

- A basic expression in the relational algebra consists of either one of the following:
 - A relation in the database
 - A constant relation
- 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

6.1.3.1 The Set-Intersection Operation

The first additional relational-algebra operation that we shall define is set intersection (\cap). Suppose that we wish to find the set of all courses taught in both the Fall 2009 and the Spring 2010 semesters. Using set intersection, we can write

$$\Pi_{course_id}$$
 ($\sigma_{semester} = \text{``Fall''} \land year = 2009 (section)$) \cap Π_{course_id} ($\sigma_{semester} = \text{``Spring''} \land year = 2010 (section)$)

The result relation for this query appears in Figure 6.13.

Note that we can rewrite any relational-algebra expression that uses set intersection by replacing the intersection operation with a pair of set-difference operations as:

$$r \cap s = r - (r - s)$$

Thus, set intersection is not a fundamental operation and does not add any power to the relational algebra. It is simply more convenient to write $r \cap s$ than to write r - (r - s).

6.1.3.2 The Natural-Join Operation

We are now ready for a formal definition of the natural join. Consider two relations r(R) and s(S). The **natural join** of r and s, denoted by $r \bowtie s$, is a relation on schema $R \cup S$ formally defined as follows:

$$r \bowtie s = \prod_{R \cup S} (\sigma_{r.A_1 = s.A_1 \land r.A_2 = s.A_2 \land ... \land r.A_n = s.A_n} (r \times s))$$

where $R \cap S = \{A_1, A_2, ..., A_n\}.$

Please note that if r(R) and s(S) are relations without any attributes in common, that is, $R \cap S = \emptyset$, then $r \bowtie s = r \times s$.

Let us consider one more example of the use of natural join, to write the query "Find the names of all instructors in the Comp. Sci. department together with the course titles of all the courses that the instructors teach."

$$\Pi_{name, title}$$
 ($\sigma_{dept_name = "Comp. Sci."}$ (instructor $\bowtie teaches \bowtie course$))

ID	name	dept_name	salary	course_id	sec_id	semester	year
10101	Srinivasan	Comp. Sci.	65000	CS-101	1	Fall	2009
10101	Srinivasan	Comp. Sci.	65000	CS-315	1	Spring	2010
10101	Srinivasan	Comp. Sci.	65000	CS-347	1	Fall	2009
12121	Wu	Finance	90000	FIN-201	1	Spring	2010
15151	Mozart	Music	40000	MU-199	1	Spring	2010
22222	Einstein	Physics	95000	PHY-101	1	Fall	2009
32343	El Said	History	60000	HIS-351	1	Spring	2010
45565	Katz	Comp. Sci.	75000	CS-101	1	Spring	2010
45565	Katz	Comp. Sci.	75000	CS-319	1	Spring	2010
76766	Crick	Biology	72000	BIO-101	1	Summer	2009
76766	Crick	Biology	72000	BIO-301	1	Summer	2010
83821	Brandt	Comp. Sci.	92000	CS-190	1	Spring	2009
83821	Brandt	Comp. Sci.	92000	CS-190	2	Spring	2009
83821	Brandt	Comp. Sci.	92000	CS-319	2	Spring	2010
98345	Kim	Elec. Eng.	80000	EE-181	1	Spring	2009

Figure 6.14 The natural join of the *instructor* relation with the *teaches* relation.

6.1.3.3 The Assignment Operation

It is convenient at times to write a relational-algebra expression by assigning parts of it to temporary relation variables. The **assignment** operation, denoted by \leftarrow , works like assignment in a programming language. To illustrate this operation, consider the definition of the natural-join operation. We could write $r \bowtie s$ as:

$$temp1 \leftarrow R \times S$$

 $temp2 \leftarrow \sigma_{r.A_1 = s.A_1 \land r.A_2 = s.A_2 \land ... \land r.A_n = s.A_n}$ (temp1)
 $result = \Pi_{R \cup S}$ (temp2)

The evaluation of an assignment does not result in any relation being displayed to the user. Rather, the result of the expression to the right of the \leftarrow is assigned to the relation variable on the left of the \leftarrow . This relation variable may be used in subsequent expressions.

6.1.3.4 Outer join Operations

The outer-join operation is an extension of the join operation to deal with missing information.

We can use the *outer-join* operation to avoid this loss of information. There are actually three forms of the operation: *left outer join*, denoted \bowtie ; *right outer join*, denoted \bowtie ; and *full outer join*, denoted \bowtie . All three forms of outer join compute the join, and add extra tuples to the result of the join. For example, the results of the expression *instructor* \bowtie *teaches* and *teaches* \bowtie *instructor* appear in Figures 6.17 and 6.18, respectively.

The **left outer join** (□×) takes all tuples in the left relation that did not match with any tuple in the right relation, pads the tuples with null values for all other attributes from the right relation, and adds them to the result of the natural join. In Figure 6.17, tuple (58583, Califieri, History, 62000, *null*, *null*, *null*, *null*), is such a tuple. All information from the left relation is present in the result of the left outer join.

ID	name	dept_name	salary	course_id	sec_id	semester	year
10101	Srinivasan	Comp. Sci.	65000	CS-101	1	Fall	2009
10101	Srinivasan	Comp. Sci.	65000	CS-315	1	Spring	2010
10101	Srinivasan	Comp. Sci.	65000	CS-347	1	Fall	2009
12121	Wu	Finance	90000	FIN-201	1	Spring	2010
15151	Mozart	Music	40000	MU-199	1	Spring	2010
22222	Einstein	Physics	95000	PHY-101	1	Fall	2009
32343	El Said	History	60000	HIS-351	1	Spring	2010
33456	Gold	Physics	87000	null	null	null	null
45565	Katz	Comp. Sci.	75000	CS-101	1	Spring	2010
45565	Katz	Comp. Sci.	75000	CS-319	1	Spring	2010
58583	Califieri	History	62000	null	null	null	null
76543	Singh	Finance	80000	null	null	null	null
76766	Crick	Biology	72000	BIO-101	1	Summer	2009
76766	Crick	Biology	72000	BIO-301	1	Summer	2010
83821	Brandt	Comp. Sci.	92000	CS-190	1	Spring	2009
83821	Brandt	Comp. Sci.	92000	CS-190	2	Spring	2009
83821	Brandt	Comp. Sci.	92000	CS-319	2	Spring	2010
98345	Kim	Elec. Eng.	80000	EE-181	1	Spring	2009

6.1.3.4 Outer join Operations

The **right outer join** (\bowtie) is symmetric with the left outer join: It pads tuples from the right relation that did not match any from the left relation with nulls and adds them to the result of the natural join. In Figure 6.18, tuple (58583, *null*, *null*, *null*, *null*, Califieri, History, 62000), is such a tuple. Thus, all information from the right relation is present in the result of the right outer join.

The full outer join(DC) does both the left and right outer join operations, padding tuples from the left relation that did not match any from the right relation, as well as tuples from the right relation that did not match any from the left relation, and adding them to the result of the join.

ID	course_id	sec_id	semester	year	name	dept_name	salary
10101	CS-101	1	Fall	2009	Srinivasan	Comp. Sci.	65000
10101	CS-315	1	Spring	2010	Srinivasan	Comp. Sci.	65000
10101	CS-347	1	Fall	2009	Srinivasan	Comp. Sci.	65000
12121	FIN-201	1	Spring	2010	Wu	Finance	90000
15151	MU-199	1	Spring	2010	Mozart	Music	40000
22222	PHY-101	1	Fall	2009	Einstein	Physics	95000
32343	HIS-351	1	Spring	2010	El Said	History	60000
33456	null	null	null	null	Gold	Physics	87000
45565	CS-101	1	Spring	2010	Katz	Comp. Sci.	75000
45565	CS-319	1	Spring	2010	Katz	Comp. Sci.	75000
58583	null	null	null	null	Califieri	History	62000
76543	null	null	null	null	Singh	Finance	80000
76766	BIO-101	1	Summer	2009	Crick	Biology	72000
76766	BIO-301	1	Summer	2010	Crick	Biology	72000
83821	CS-190	1	Spring	2009	Brandt	Comp. Sci.	92000
83821	CS-190	2	Spring	2009	Brandt	Comp. Sci.	92000
83821	CS-319	2	Spring	2010	Brandt	Comp. Sci.	92000
98345	EE-181	1	Spring	2009	Kim	Elec. Eng.	80000

Figure 6.18 Result of teaches ⋈ instructor.

(**Division operation**): The division operator of relational algebra, " \div ", is defined as follows. Let r(R) and s(S) be relations, and let $S \subseteq R$; that is, every attribute of schema S is also in schema R. Then $r \div s$ is a relation on schema R - S (that is, on the schema containing all attributes of schema R that are not in schema S). A tuple t is in t if and only if both of two conditions hold:

- t is in $\Pi_{R-S}(r)$
- For every tuple t_s in s, there is a tuple t_r in r satisfying both of the following:

a.
$$t_r[S] = t_s[S]$$

$$b. t_r[R - S] = t$$

Relation schemas

- CUSTOMER(C_ID, NAME, PHONE_NO, ADDRESS)
- DEPOSITOR(C_ID, A_NO)
- ACCOUNT(A_NO,BALANCE,B_NAME)
- BORROWER(C_ID, L_NO)
- LOAN(L_NO, L_AMT, B_NAME)
- BRANCH(B_NAME, B_CITY, B_ASSETS)

Relational algebra queries

1. Find out the account details with balance greater than 10000.

2. Find out the loan details where the loan amount is greater than 50000 and loan taken is from khandagiri branch.

3. Find out the phone number and address of all the customers.

4. Find out the phone number and address of customer name sachin5. Find out the account number and branch name of all those account where the balance is less than 1000.

6. Find out the customer id of those customer who have either an account in the bank or else have borrower loan from the bank.

7. Find out the customer Id of customer who have account in the bank but not taken loan.

8. Find out the name of customer along with their account number.

9. Find out the name, phone number and balance of customer having balance more than 50000.

10. Find out the name of customer who have taken loan from any branch with branch city Bhubaneswar 11. Find out the branch name and branch city of customer having customer name sachin has a account 12. Find out the customer Id along with their loan number and account number

Tuple Relational Calculus

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

Predicate Calculus Formula

- 1. Set of attributes and constants
- 2. Set of comparison operators: (e.g., \langle , \leq , =, \neq , \rangle , \geq)
- 3. Set of connectives: and (\land) , or (\lor) , not (\neg)
- 4. Implication (\Rightarrow) : $x \Rightarrow y$, if x if true, then y is true

$$X \Rightarrow y \equiv \neg X \lor y$$

- 5. Set of quantifiers:
 - ▶ $\exists t \in r(Q(t)) \equiv$ "there exists" a tuple in t in relation r such that predicate Q(t) is true
 - $\forall t \in r(Q(t)) \equiv Q$ is true "for all" tuples t in relation r

■ Find the *ID*, name, dept_name, salary for instructors whose salary is greater than \$80,000

$$\{t \mid t \in instructor \land t [salary] > 80000\}$$

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

As in the previous query, but output only the *ID* attribute value

$$\{t \mid \exists \ s \in \text{instructor} \ (t[ID] = s[ID] \land s[salary] > 80000)\}$$

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

 Find the names of all instructors whose department is in the Watson building

```
\{t \mid \exists s \in instructor (t [name] = s [name] \\ \land \exists u \in department (u [dept_name] = s[dept_name] " \\ \land u [building] = "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 section \ (t [course\_id] = s [course\_id] \land s [semester] = "Fall" \land s [year] = 2009 \ \lor \exists u \in section \ (t [course\_id] = u [course\_id] \land u [semester] = "Spring" \land u [year] = 2010 )\}
```

Q1: Find out the account detail where the balance is greater than 50000.

Q2: Find out the loan details for branch name Khandagiri, and loan amount greater than 50000. Q3: Find out the account number for which the balance is greater than 50000.

Q4: Find out the name of customer along with their phone number.

Q5: Find out the name of customers who have an account but not taken loan.

■ Q6: Find out the branch city of the branch name Khandagiri

Q7: Find out the loan amount and branch name of the loan with loan number=L001.

Q8: Find out the account number of the customer along with their customer name.

Q9: Find out the account number of customer name sachin.

Universal Quantification

Q10: Find out all customer name who have an account at all branches located in Bhubaneswar.

Safety of Expressions (Cont.)

 Consider again that query to find all students who have taken all courses offered in the Biology department

```
    {t | ∃ r ∈ student (t [ID] = r [ID]) ∧
    (∀ u ∈ course (u [dept_name]="Biology" ⇒
    ∃ s ∈ takes (t [ID] = s [ID] ∧
    s [course_id] = u [course_id]))}
```

Without the existential quantification on student, the above query would be unsafe if the Biology department has not offered any courses.

Domain Relational Calculus

Domain Relational Calculus

- A nonprocedural query language equivalent in power to the tuple relational calculus
- Each query is an expression of the form:

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

- $x_1, x_2, ..., x_n$ represent domain variables
- P represents a formula similar to that of the predicate calculus

- Find the *ID*, *name*, *dept_name*, *salary* for instructors whose salary is greater than \$80,000
 - $\{ < i, n, d, s > | < i, n, d, s > \in instructor \land s > 80000 \}$
- As in the previous query, but output only the ID attribute value
 - $\{ < i > | < i, n, d, s > \in instructor \land s > 80000 \}$
- 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 \land \exists b, a \ (\langle d, b, a \rangle \in department \land b = "Watson") \} \}
```

Find the set of all courses taught in the Fall 2009 semester, or in the Spring 2010 semester, or both

```
{<c> | ∃ a, s, y, b, r, t (<c, a, s, y, b, r, t > ∈ section ∧ s = \text{``Fall''} \land y = 2009)
v∃ a, s, y, b, r, t (<c, a, s, y, b, r, t > ∈ section] ∧ s = \text{``Spring''} \land y = 2010)}
```

This case can also be written as

$$\{ \langle c \rangle \mid \exists \ a, \ s, \ y, \ b, \ r, \ t \ (\langle c, \ a, \ s, \ y, \ b, \ r, \ t \rangle \in section \land ((s = "Fall" \land y = 2009)) \lor (s = "Spring" \land y = 2010)) \}$$

■ Find the set of all courses taught in the Fall 2009 semester, and in the Spring 2010 semester

Universal Quantification

Find all students who have taken all courses offered in the Biology department

```
    {< i > | ∃ n, d, tc ( < i, n, d, tc > ∈ student ∧
    (∀ ci, ti, dn, cr ( < ci, ti, dn, cr > ∈ course ∧ dn = "Biology"
    ⇒ ∃ si, se, y, g ( <i, ci, si, se, y, g > ∈ takes ))}
```

Find out the account detail where the balance is greater than 50000.

■ Find out the customer details of customer having phone number 7407983296

Find out the name of all customers along with their phone numbers.

Find out the loan number and amount of all loan taken from khandagiri branch.

Find out the phone number and address of customer named sachin.

Find out all customer Id where the account are maintained at khandagiri branch.

Find out the account number of customer name sachin.

■ Find out the loan amount of a customer with customer ID C001.

■ Find out the branch city where customer with customer Id C001 has his account.

End of Chapter 6