

BITS, PILANI – K. K. BIRLA GOA CAMPUS

Database Systems (CS F212)

by

Dr. Mrs. Shubhangi Gawali
Dept. of CS and IS



Relational Algebra

- Six basic operators
 - select
 - project
 - union
 - set difference
 - Cartesian product
 - rename
- The operators take one or more relations as inputs and give a new relation as a result.

Select Operation – Example

• Relation *r*

Α	В	С	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

• $\sigma_{A=B \land D > 5}(r)$

Α	В	С	D
α	α	1	7
β	β	23	10

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:

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

• Example of selection:

$$\sigma_{\textit{branch-name}="Perryridge"}(\textit{account})$$

Project Operation – Example

• Relation *r*:

Project Operation

Notation:

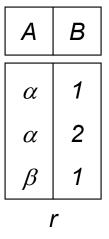
$$\prod_{A1, A2, ..., Ak} (r)$$

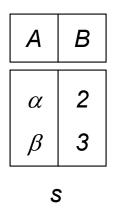
where A_1 , A_2 are attribute names and r is a relation name.

- 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
- E.g. To eliminate the *branch-name* attribute of *account* $\Pi_{account-number,\ balance}$ (account)

Union Operation – Example

• Relations *r, s:*





 $r \cup s$:

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* (e.g., 2nd column of *r* deals with the same type of values as does the 2nd column of *s*)
- E.g. to find all customers with either an account or a loan

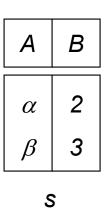
 $\prod_{customer-name}$ (depositor) $\cup \prod_{customer-name}$ (borrower)

Set Difference Operation – Example

Relations

r, s:

Α	В	
α	1	
α	2	
β	1	
r		



r − *s*:

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

Cartesian-Product Operation-Example

Relations *r*, *s*:

Α	В	
α	1	
β	2	
r		

С	D	Ε
$\begin{array}{c c} \alpha & \\ \beta & \\ \beta & \\ \gamma & \end{array}$	10 10 20 10	a a b b
· · · · · · · · · · · · · · · · · · ·		

r x s:

Α	В	С	D	E
α	1	α	10	а
α	1	β	10	а
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	а
β	2	β	10	а
β	2	β	20	b
β	2	γ	10	b

Cartesian-Product Operation

- Notation r x s
- Defined as:

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

• 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.

Composition of Operations

- Can build expressions using multiple operations
- Example:

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

rxs

• $\sigma_{A=C}(r x s)$

Α	В	C	D	E
α	1	α	10	а
α	1	β	10	а
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	а
β	2	β	10	а
β	2	β	20	b
β	2	γ	10	b

Α	В	С	D	E
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	1	$\begin{array}{c c} \alpha \\ \beta \end{array}$	10 20	а
β	2 2	β	20	a b

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 XIf a relational-algebra expression E has arity n, then

$$\rho_{x (A1, A2, ..., An)}(E)$$

returns the result of expression *E* under the name *X*, and with the

attributes renamed to A1, A2,, An.

Additional Operations

- Set intersection
- Natural join
- Division
- Assignment

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

Set-Intersection Operation - Example

Relation

r, s:

Α	В
αα	1 2
β	1

A B
α 2
β 3

S

r

 $r \cap s$

Α	В
α	2

Natural Join Operation – Example

• Relations r, s:

Α	В	С	D
α	1	α	а
β	2	γ	а
γ	4	β	b
α	1	γ	а
δ	2	β	b
r			

В	D	E
1	а	α
3	а	β
1	а	$\frac{\gamma}{\delta}$
2	b	δ
3	b	\in
S		

 $r^{\bowtie}s$

Α	В	С	D	E
α	1	α	а	α
α	1	α	а	γ
α	1	γ	а	α
α	1	γ	а	γ
δ	2	β	b	δ

Natural Join – Example

• Relation *loan*

loan-numberbranch-nameamountL-170Downtown3000L-230Redwood4000L-260Perryridge1700

• Relation borrower

customer-name	loan-number	
Jones	L-170	
Smith	L-230	
Hayes	L-155	

loan X borrower

loan.loan- number	branch- name	amount	borrower. loan- number	customer -name
L-170	Downtown	3000	L-170	Jones
L-170	Downtown	3000	L-230	Smith
L-170	Downtown	3000	L-155	Hayes
L-230	Redwood	4000	L-170	Jones
L-230	Redwood	4000	L-230	Smith
L-230	Redwood	4000	L-155	Hayes
L-260	Perryridge	1700	L-170	Jones
L-260	Perryridge	1700	L-230	Smith
L-260	Perryridge	1700	L-155	Hayes

loan |X| borrower

loan.loan- number	branch- name	amount	borrower. loan- number	customer -name
L-170	Downtown	3000	L-170	Jones
L-170	Downtown	3000	L-230	Smith X
L-170	Downtown	3000	L-155	Hayes X
L-230	Redwood	4000	L-170	Jones X
L-230	Redwood	4000	L-230	Smith
L-230	Redwood	4000	L-155	Hayes X
L-260	Perryridge	1700	L-170	Jones X
L-260	Perryridge	1700	L-230	Smith X
L-260	Perryridge	1700	L-155	Hayes X

loan |X| borrower

loan.loan- number	branch- name	amount	borrower. loan- number	customer -name
L-170	Downtown	3000	L-170	Jones
L-230	Redwood	4000	L-230	Smith

loan- number	branch- name	amount	customer -name
L-170	Downtown	3000	Jones
L-230	Redwood	4000	Smith

Division Operation – Example

Relations r, s:

α	1
α	2
α	3
β	1
γ	1
δ	1
δ	3
δ	4
\in	6
\in	1
β	2
	_

1 2

S

 $r \div s$: A

 $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

r

Another Division Example

Relations *r*, *s*:

Α	В	С	D	E
α	а	α	а	1
α	а	γ	а	1
α	а		b	1
β	а	γ	а	1
β	а	γ	b	3
γ		γ	а	1
$\begin{bmatrix} \alpha \\ \alpha \\ \alpha \\ \beta \\ \beta \\ \gamma \\ \gamma \\ \gamma \end{bmatrix}$	a a	γ γ γ γ	b	1
γ	а	β	b	1
r				

D E
a 1
b 1

r ÷ s:

Α	В	С
α	а	γ
γ	а	γ

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.
- Example: Write $r \div s$ as

$$temp1 \leftarrow \prod_{R-S} (r)$$

 $temp2 \leftarrow \prod_{R-S} ((temp1 \times s) - \prod_{R-S,S} (r))$
 $result = temp1 - temp2$

- The result to the right of the \leftarrow is assigned to the relation variable on the left of the \leftarrow .
- May use variable in subsequent expressions.

Extended Relational-Algebra-Operations

- Generalized Projection
- Outer Join
- Aggregate Functions

Generalized Projection

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

$$\prod_{\mathsf{F1},\,\mathsf{F2},\,\ldots,\,\mathsf{Fn}}(E)$$

- E is any relational-algebra expression
- Each of F_1 , F_2 , ..., F_n are are arithmetic expressions involving constants and attributes in the schema of E.
- Given relation credit-info(customer-name, limit, credit-balance), find how much more each person can spend:

 $\prod_{customer-name, \ limit-credit-balance}$ (credit-info)

Aggregate Functions and Operations

 Aggregation function takes a collection 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

$$G_{1, G_{2, ..., G_{n}}} \mathcal{G}_{F_{1}(A_{1}), F_{2}(A_{2}), ..., F_{n}(A_{n})} (E)$$

- E is any relational-algebra expression
- $-G_1, G_2 ..., G_n$ is a list of attributes on which to group (can be empty)
- Each F_i is an aggregate function
- Each A_i is an attribute name

Aggregate Operation – Example

• Relation *r*

Α	В	С
α	α	7
α	β	7
β	β	3
β	β	10

$$g_{\text{sum(c)}}(r)$$

*sum-C*27

Aggregate Operation – Example

• Relation *account* grouped by *branch-name*:

branch-name	account-number	balance
Perryridge	A-102	400
Perryridge	A-201	900
Brighton	A-217	750
Brighton	A-215	750
Redwood	A-222	700

branch-name $g_{sum(balance)}$ (account)

branch-name	balance
Perryridge	1300
Brighton	1500
Redwood	700

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

Outer Join – Example

• Relation *loan*

loan-number	branch-name	amount
L-170	Downtown	3000
L-230	Redwood	4000
L-260	Perryridge	1700

Relation borrower

customer-name	loan-number
Jones	L-170
Smith	L-230
Hayes	L-155

Outer Join – Example

Inner Join

loan ⋈ *Borrower*

loan-number	branch-name	amount	customer-name
L-170	Downtown	3000	Jones
L-230	Redwood	4000	Smith

■ Left Outer Join

Ioan Borrower

loan-number	branch-name	amount	customer-name
L-170	Downtown	3000	Jones
L-230	Redwood	4000	Smith
L-260	Perryridge	1700	null

Outer Join – Example

Right Outer Join

loan ⋈ borrower

loan-number	branch-name	amount	customer-name
L-170	Downtown	3000	Jones
L-230	Redwood	4000	Smith
L-155	null	null	Hayes

■ Full Outer Join

loan □×□*borrower*

loan-number	branch-name	amount	customer-name
L-170	Downtown	3000	Jones
L-230	Redwood	4000	Smith
L-260	Perryridge	1700	null
L-155	null	null	Hayes

Modification of the Database

- The content of the database may be modified using the following operations:
 - Deletion
 - Insertion
 - Updating
- All these operations are expressed using the assignment operator.

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.

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 ← loan −
$$\sigma$$
 amount ≥ 0 and amount ≤ 50 (loan)

Delete all accounts at branches located in Needham.

```
r_1 \leftarrow \sigma_{branch-city} = \text{``Needham''} (account \bowtie branch)
r_2 \leftarrow \Pi_{branch-name, account-number, 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
```

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.

Insertion Examples

• Insert information in the database specifying that Smith has \$1200 in account A-973 at the Perryridge branch.

```
account \leftarrow account \cup \{("Perryridge", A-973, 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 \prod_{branch-name, account-number, 200}(r_1)
depositor \leftarrow depositor \cup \prod_{customer-name, loan-number}(r_1)
```

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

$$r \leftarrow \prod_{F1, F2, ..., Fl_r} (r)$$

- Each F_i is either
 - the *i*th attribute of *r*, if the *i*th attribute is not updated, or,
 - if the attribute is to be updated F_i is an expression, involving only constants and the attributes of r, which gives the new value for the attribute

Update Examples

Make interest payments by increasing all balances by 5 percent.

$$account \leftarrow \prod_{AN, BN, BAL * 1.05} (account)$$

where AN, BN and BAL stand for account-number, branch-name and balance, respectively.

Pay all accounts with balances over \$10,000 6 percent interest and pay all others 5 percent

```
account \leftarrow \quad \prod_{AN,\ BN,\ BAL\ ^*\ 1.06} (\sigma_{BAL\ >\ 10000} (account)) \\ \cup \quad \prod_{AN,\ BN,\ BAL\ ^*\ 1.05} (\sigma_{BAL\ \le\ 10000} (account))
```

Relational Schema

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
 - Q1. List the name of all Employees

∏ name (Employee)

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)

Q2.List the name & telno of all junior engineers.

 $\Pi_{\text{name, telno}}(\sigma_{\text{post}} = 'junior engineer', (Employee))$

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q3. List the name & post of those employees who live in mumbai & have salary > 10,000.

 $\Pi_{\text{name,post}}(\sigma_{\text{city}} = \text{'Mumbai'} \land \text{sal>10,000} \text{(Employee)})$

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)

Q4. Find the names of employees where projno=123; ANS 1

 Π_{name} ($\sigma_{\text{pno}=123}$ Λ Assigned.eno=Employee.eno (Assigned X Employee))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)

Q4. Find the names of employees where projno=123; ANS 2

 Π_{name} ($\sigma_{\text{Assigned.eno}}$ (Employee X ($\sigma_{\text{pno}=123}$ (Assigned))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q5. Find the location where employee no 5 has worked.

ANS₁

 Π location (σ eno=5 Λ Assigned.pno=project.pno (Project X Assigned))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q5. Find the location where employee no 5 has worked.

ANS 2

 Π location (σ Assigned.pno=project.pno (Project X (σ eno=5 (Assigned))))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q6. Find the city of the managers of all the projects located at Dadar. Assume Managerid field contains eno.

 Π_{city} ($\sigma_{location}$ = 'Dadar' Λ Eno= Managerid (Project X Employee))

- Employee(eno,Name,Telno,post,DOJ, sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q7. Find the eno who are assigned the pno 123 after the manager of the project joined the organisation.

```
\Pi_{eno} (\sigma_{date>DOJ} \Lambda Assigned.pno=123 (Assigned X (\sigma_{eno=managerid} (Employee X (\sigma_{pno} (Project))))))
```

• A \leftarrow (σ pno=123(Project))

• B \leftarrow ($\sigma_{\text{eno=managerid}}$ (Employee X A))

• C ← (σ date>DOJ Λ Assigned.pno=123 (Assigned X B))

• R←—∏ eno (C)

- Employee(eno,Name,Telno,post,DOJ, sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q7. Find the eno who are assigned the pno 123 after the manager of the project joined the organisation.

 Π_{eno} ($\sigma_{date>DOJ}$ Λ Assigned.pno=123 Λ Project.pno = 123 Λ eno=Managerid (Assigned X Employee X Project))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q8. Find the Employees who are living in the same city as that of employee 'XYZ'.

temp1
$$\leftarrow \prod$$
 city (σ Name= 'XYZ' Employee)

 \prod eno (σ Name<> 'XYZ'^ temp1.city = Employee.city (temp1 X Employee))

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q8. Find the Employees who are living in the same city as that of employee 'XYZ'.

 Π_{Name} (σ_{Name} (σ_{Name} (σ_{Name}) (Employee X ρ_{E})

- Employee(eno,Name,Telno,post,DOJ,sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q9. Find the complete details of all the employees who are assigned to project no 123.

Employee $|X| \prod_{Eno} (\sigma_{pno=123} (Assigned))$

- Employee(eno,Name,Telno,post,DOJ, sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q10. Find the employee who are working at the same location where they reside.

 $\prod_{\text{name}} (\sigma_{\text{city} = \text{location}} (\text{Project} | X | \text{Assigned} | X | \text{Employee})$

- Employee(eno,Name,Telno,post,DOJ, sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q11. Find the employees who are working with employee xyz.

Step1: Find employee whose name is xyz

$$\sigma$$
 name = 'xyz' (Employee)

Step2: Projects on which xyz is working

A
$$\leftarrow$$
 \prod pno (Assigned |X|(σ name = 'xyz'(Employee)

Step3: Employees working on that project

$$B \leftarrow \prod_{eno} (Assigned | X | A)$$

Step 4: Name of Employees working on the project

- Employee(eno,Name,Telno,post,DOJ, sal,city)
- Project (pno, Managerid, location)
- Assigned(eno,pno,date,task)
- Q12. Find the projects whose location is same as their manager's city.

 Π_{pno} ($\sigma_{manager=eno}$ Λ city= location (Project X Employee))

Q13. Find Maximum salary

Employee

Name	Salary
Α	10k
В	20k
С	30k

F

E.Name	E.Salary
Α	10k
В	20k
С	30k

Employee X E

Emp. Name	Emp. Sal	E. Name	E. Sal
Α	10k	Α	10k
Α	10k	В	20k
Α	10k	С	30k
В	20k	A	10k
В	20k	В	20k
В	20k	С	30k
С	30k	A	10k
С	30k	В	20k
С	30k	С	30k

Emp.	Emp.
Name	Sal
Α	10K
В	20K
С	30K

E. Name	E. Sal
Α	10K
Α	10K
В	20K

Emp.Name	Emp. Sal
С	30k

Solution

 $\Pi_{emp.sal}$ (Emp)- $\Pi_{E.sal}$ ($\sigma_{emp.sal}$ -E.sal (Employee X ρ_{E} Employee))

OR

 $\Pi_{emp.sal}$ (Emp)- $\Pi_{Emp.sal}$ ($\sigma_{emp.sal}$ <E.sal (Employee X ρ_{E} Employee))

For Finding Minimum salary

 $\Pi_{emp.sal}$ (Emp)- $\Pi_{E.sal}$ ($\sigma_{emp.sal < E.sal}$ (Employee X ρ_{F} Employee))

Exercise 2

Database schema

- Professor(<u>ssn</u>, profname, status, salary)
- Course(crscode, crsname, credits)
- Taught(crscode, semester, ssn)

Assumptions: (1) Each course has only one instructor in each semester; (2) all professors have different salaries; (3) all professors have different names; (4) all courses have different names; (5) status can take values from "Full", "Associate", and "Assistant".

- Find those professors who have taught 'C1' but have never 'C2'.
- Find those professors who have taught both 'C2' and 'C3'.
- Find those professors who have never taught 'C2'.
- Find those professors who taught 'C2' and 'C3' in the same semester.

- Find those professors who taught 'C1' or 'C2' but not both.
- Find those courses that have never been taught.
- Find those courses that have been taught at least in two semesters.
- Find the names of professors who ever taught 'C3'.

- Find the names of Assistant Professors who ever taught 'C2'.
- Find the names of professors who ever taught at least two courses in one semester.
- List all the course names that professor 'Dr. Bhoomi Desai' taught in Even sem of 2009.
- List those courses that have been taught ONLY by Assistant professors.