Database Management Systems (DBMS)

Lec 5: Relational model of data (Cont.)

Ramesh K. Jallu

IIIT Raichur

Date: 5/2/21

Recap

- Some warmup exercise
- The join operation
  - $\boldsymbol{R} \bowtie_{\text{<join condition>}} \boldsymbol{S} = \sigma_{\text{<join condition>}}(\boldsymbol{R} \times \boldsymbol{S})$

Today's plan

- Answers to the take-home queries
- The other variations of the join operation
  - Equijoin and Natural join
- The division operation (÷)

Exercise: Bank database

- Branch(**Name,** Assets, City)
- Customer(**ID**, Customer_Name, Street, City)
- Loan (**Loan_Number**, Branch_name, Amount)
- Borrower(**BC_ID**, **BL_Number**)
- Account (**Account_Number**, AB_name, Balance)
- Depositor (**DC_ID** , **A_number**)

Take-home queries

1. Find the customer name for the last three queries with the same requirement

   - Find all customers IDs who have loan 10,000/- or account balance > 5,000/-
   - Find the IDs of all customers who have a loan at the Raichur branch
   - Find the IDs of all customers who have a loan at the Raichur branch but do not have an account at any branch of the bank

2. Find the largest account balance
   - Hint: Use Rename, Cartesian, and Set Minus operations

Answers

1. Use the join operation with the previous solutions with join condition …..
   - $\pi_{BC\_ID}\ (\sigma_{Amount\ =\ 10000}(Borrower)) \cup \pi_{DC\_ID}\ (\sigma_{Account\_Number\ =\ A\_Number}(Account \times Depoister))$
   - $\pi_{BC\_ID}\ (\sigma_{Branch\_Name\ =\ 'Raichur'\ \textbf{AND}\ BL\_Number\ =\ Loan\_Number}\ (Borrower \times Loan))$
   - $TEMP \leftarrow \pi_{BC\_ID}\ (\sigma_{Branch\_Name\ =\ 'Raichur'\ \textbf{AND}\ BL\_Number\ =\ Loan\_Number}\ (Borrower \times Loan))$
     $TEMP - \pi_{DC\_ID}(Depoister)$

2. Strategy:
   1. Find those balances that are ***not*** the largest
   2. Use set difference to find those account balances that were ***not*** found in the Step 1

Account (**Account_Number**, AB_name, Balance)

Account

Account_Number	AB_name	Balance
63165065258	Raichur	88,833
54674498364	Hyderabad	55,544
59369845487	Delhi	3,42,572
36465594547	Pune	77,847
76994489457	Chennai	37,545

TEMP

Acc_Bal
88,833
55,544
3,42,572
77,847
37,545

$\rho_{TEMP}\ (Acc\_Bal)(\pi_{\text{Balance}}(\text{Account}))$

Balance	Acc_Bal
88,833	3,42,572
55,544	88,833
55,544	3,42,572
77,847	88,833
77,847	3,42,572
37,545	88,833
37,545	55,544
37,545	3,42,572
37,545	77,847

$\sigma_{Balance < Acc\_Bal}(\pi_{\text{Balance}}(\text{Account})\ x\ TEMP)$

$\pi_{\text{Balance}}(\sigma_{Balance < Acc\_Bal}\ (\pi_{\text{Balance}}(\text{Account})\ x\ TEMP))$

Balance
88,833
55,544
77,847
37,545

$\pi_{\text{Balance}}(\text{Account})\ -\ (\quad)$

Balance
3,42,572

Variations of join: Equi join and Natural join

- A general join condition: <cond> **AND**<cond>**AND**….**AND**<cond>
  - **Theta join**: A_i θ B_j such that $dom(A_i) = dom(B_j)$, and θ is one of the comparison operators {=, <, ≤, >, ≥, ≠}
  - **Equi join**: Join operation with join condition with equality comparisons, I.e., θ is "="

Example from last class

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

Query: Retrieve the name of the manager of each department

$$\text{DEPT\_MGR} \leftarrow \text{DEPARTMENT} \bowtie_{\text{Mgr\_ssn=Ssn}} \text{EMPLOYEE}$$
$$\text{RESULT} \leftarrow \pi_{\text{Dname, Lname,Fname}}(\text{DEPT\_MGR})$$

DEPT_MGR

Dname	Dnumber	Mgr_ssn	...	Fname	Minit	Lname	Ssn	...
Research	5	333445555	...	Franklin	T	Wong	333445555	...
Administration	4	987654321	...	Jennifer	S	Wallace	987654321	...
Headquarters	1	888665555	...	James	E	Borg	888665555	...

RESULT

Dname	Lname	Fname
Research	Wong	Franklin
Administration	Wallace	Jennifer
Headquarters	Borg	James

Natural join (*)

- An equi join operation followed by the removal of the superfluous attributes
- Notation: $R*S$
- The natural join of two relations $R(A_1,A_2,\ldots,A_n)$ and $S(B_1,B_2,\ldots,B_m)$, where $Ai = Bj$, is a new relation $Q(A_1,A_2,\ldots,A_i,\ldots,A_n,B_1,B_2,\ldots,B_{j-1},B_{j+1},B_m)$
  - Q contains those tuples satisfying ***equi join*** condition
  - Q removes B_i

Observations

- ***Join attribute***: Natural Join requires that the two join attributes have the same name in both relations
- If this is not the case, a renaming operation is applied first
- There can be a list of join attributes from each relation, and each corresponding pair must have the same name

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

PROJ_DEPT ← PROJECT * $\rho_{(\text{Dname, Dnum, Mgr\_ssn, Mgr\_start\_date})}$(DEPARTMENT

PROJ_DEPT

Pname	Pnumber	Plocation	Dnum	Dname	Mgr_ssn	Mgr_start_date
ProductX	1	Bellaire	5	Research	333445555	1988-05-22
ProductY	2	Sugarland	5	Research	333445555	1988-05-22
ProductZ	3	Houston	5	Research	333445555	1988-05-22
Computerization	10	Stafford	4	Administration	987654321	1995-01-01
Reorganization	20	Houston	1	Headquarters	888665555	1981-06-19
Newbenefits	30	Stafford	4	Administration	987654321	1995-01-01

DEPT_LOCS ← DEPARTMENT * DEPT_LOCATIONS

DEPT_LOCS

Dname	Dnumber	Mgr_ssn	Mgr_start_date	Location
Headquarters	1	888665555	1981-06-19	Houston
Administration	4	987654321	1995-01-01	Stafford
Research	5	333445555	1988-05-22	Bellaire
Research	5	333445555	1988-05-22	Sugarland
Research	5	333445555	1988-05-22	Houston

The division operation

- Notation: ***R ÷ S***

- Used to deal queries which contain the keyword ***All/every***
  - Retrieve students IDs who enrolled in *every* course
  - Retrieve details of a customer who has account in *all* the banks in a city
  - Retrieve the names of employees who work on *all* the projects that ‘John Smith’ works on

How it works?

1. Division is applied to two relations ***R(X)*** and ***S(Y)*** such that $Y \subseteq X$
2. Let $Z = X - Y$ I.e., $X = Y \cup Z$
3. ***T(Z) = R(X) ÷ S(Y)***
   - For every tuple ***t*** in ***T(Z)***, the values in ***t*** must appear in ***R*** in combination with every tuple in ***S***

Explanation

1. ***R(X)*** such that ***X*** = **{A_1,A_2,…,A_n,B_1,B_2,…,B_m}** and ***S(Y)*** such that ***Y*** =**{B_1,B_2,…,B_m}**

2. Let ***Z*** = ***X*** – ***Y*** = **{A_1,A_2,…,A_n}**

3. ***R(X)*** ÷ ***S(Y)*** Contains attributes **A_1,A_2,…,A_n** with tuple values <a_1,a_2,…,a_n> such that for every tuple <b_1,b_2,…,b_m> in ***S***, <a_1,a_2,…,a_n,b_1,b_2,…,b_m> is in ***R***

Example

STUDENT

Student_ID	Course_ID
01CS19	MA101
16CS19	CS201
01CS19	MA203
01CS19	CS101
16CS19	MA101
18CS20	MA101
01CS19	CS201

COURSE

Course_ID
MA101
CS201
MA203
CS101

RESULT

Student_ID
01CS19

RESULT = STUDENT ÷ COURSE

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

Exercise

Query: Retrieve the names of employees who work on *all* the projects that ‘John Smith’ works on

Strategy: 1. Retreive the projects that 'John Smith' works on: R_1

2. Retreive the employees who work on which projects: R_2

3. $R_2 \div R_1$

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

$$SMITH \leftarrow \sigma_{Fname = \text{'John' AND } Lname = \text{'Smith'}} (EMPLOYEE)$$

$$SMITH\_PNOS \leftarrow \pi_{Pno}(WORKS\_ON \bowtie_{Essn = Ssn} SMITH)$$

$$SSN\_PNOS \leftarrow \pi_{Essn,\ Pno} (WORKS\_ON)$$

$$SSNS (Ssn) \leftarrow SSN\_PNOS \div SMITH\_PNOS$$

$$RESULT \leftarrow \pi_{Fname,\ Lname} (SSNS * EMPLOYEE)$$

SSN_PNOS

Essn	Pno
123456789	1
123456789	2
666884444	3
453453453	1
453453453	2
333445555	2
333445555	3
333445555	10
333445555	20
999887777	30
999887777	10
987987987	10
987987987	30
987654321	30
987654321	20
888665555	20

SMITH_PNOS

Pno
1
2

SSNS

Ssn
123456789
453453453

OPERATION	PURPOSE	NOTATION
SELECT	Selects all tuples that satisfy the selection condition from a relation R.	$\sigma_{\text{<selection condition>}}(R)$
PROJECT	Produces a new relation with only some of the attributes of R, and removes duplicate tuples.	$\pi_{\text{<attribute list>}}(R)$
THETA JOIN	Produces all combinations of tuples from R_1 and R_2 that satisfy the join condition.	$R_1 \bowtie_{\text{<join condition>}} R_2$
EQUIJOIN	Produces all the combinations of tuples from R_1 and R_2 that satisfy a join condition with only equality comparisons.	$R_1 \bowtie_{\text{<join condition>}} R_2$, OR $R_1 \bowtie_{\text{(<join attributes 1>), (<join attributes 2>)}} R_2$
NATURAL JOIN	Same as EQUIJOIN except that the join attributes of R_2 are not included in the resulting relation; if the join attributes have the same names, they do not have to be specified at all.	$R_1 *_{\text{<join condition>}} R_2$, OR $R_1 *_{\text{(<join attributes 1>), (<join attributes 2>)}} R_2$ OR $R_1 * R_2$
UNION	Produces a relation that includes all the tuples in R_1 or R_2 or both R_1 and R_2; R_1 and R_2 must be union compatible.	$R_1 \cup R_2$
INTERSECTION	Produces a relation that includes all the tuples in both R_1 and R_2; R_1 and R_2 must be union compatible.	$R_1 \cap R_2$
DIFFERENCE	Produces a relation that includes all the tuples in R_1 that are not in R_2; R_1 and R_2 must be union compatible.	$R_1 - R_2$
CARTESIAN PRODUCT	Produces a relation that has the attributes of R_1 and R_2 and includes as tuples all possible combinations of tuples from R_1 and R_2.	$R_1 \times R_2$
DIVISION	Produces a relation $R(X)$ that includes all tuples $t[X]$ in $R_1(Z)$ that appear in R_1 in combination with every tuple from $R_2(Y)$, where $Z = X \cup Y$.	$R_1(Z) \div R_2(Y)$

Complete set

- $\{\sigma, \pi, \cup, \rho, -, \times\}$ is complete set
- Any other relational operation can be expressed as a combination of these
- How can we express the *intersection* operation?
- The division operation can be expressed as a sequence of π, ×, and – operations as follows:
  1. $T_1 \leftarrow \pi_Z(R)$
  2. $T_2 \leftarrow \pi_Z((S \times T_1) - R)$
  3. $T \leftarrow T_1 - T_2$

Example

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

$1.T_1 \leftarrow \pi_Z(R)$
$2.T_2 \leftarrow \pi_Z((S \times T_1) - R)$
$3.T \leftarrow T_1 - T_2$

X = {A,B}, Y={A}, Z = {B}

T_1

B
b1
b2
b3
b4

$S \times T_1$

A	B
a_1	b_1
a_1	b_2
a_1	b_3
a_1	b_4
a_2	b_1
a_2	b_2
a_2	b_3
a_2	b_4
a_3	b_1
a_3	b_2
a_3	b_3
a_3	b_4

$(S \times T_1) - R$

A	B
a_1	b_3
a_2	b_2

T_2

B
b_3
b_2

T

B
b_1
b_4

Thank you!