Module 1.2 - Relational Algebra

1. Relational Data Model - Review

- 1.1 Relational Data (RD) Model Properties and Keys
  - History, Model, Keys, Indices, Constraints
- 1.2 Relational algebra
  - Basic relational operators
  - Joins
  - Division

1.1 RD Model: Properties and Keys

Learning Outcomes

- Explain and use basic relational algebra operators
- Explain and use equivalence of expressions
- Explain and use various forms of joins and other relational operators

Textbook Readings

- Chap 3 … readings (relational algebra)

Testing*

*Main (but not the only ones) sections of the textbook used for testing are identified in parentheses

- Relational algebra (3.4, 3.8)

Relational Operations

- Select (restrict)
  - Unary operator that yields a horizontal subset of a table
- Project
  - Unary operator that yields a vertical subset of a table (Duplicates?)
- Union
  - Combines all rows from two tables, excluding duplicate rows
  - *Union-compatible*: tables share the same number of columns, and their corresponding columns share compatible domains
- Difference
  - Yields all rows in one table that are not found in the other table
  - Tables must be union-compatible to yield valid results
- Product
  - Yields all possible pairs of rows from two tables
- Intersect
  - Yields only the rows that appear in both tables
  - Tables must be *union-compatible* to yield valid results

- **Operation Arguments**
  - Have to be Union compatible?
- **Operation result**
  - Columns/Attributes
    - Names and Number of?
  - Rows/tuples
    - Number of?

Relational Operations … continued

FIGURE 3.4 SELECT

Original table

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99

SELECT ALL yields

New table

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99

SELECT only PRICE less than $2.00 yields

P_CODE	P_DESCRIPT	PRICE
213345	9v battery	1.92
254467	100W bulb	1.47

SELECT only P_CODE = 311452 yields

P_CODE	P_DESCRIPT	PRICE
311452	Powerdrill	34.99

Relational Operations … continued

FIGURE 3.5 PROJECT

Original table

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99

PROJECT PRICE yields

New table

PRICE
5.26
25.15
10.99
1.92
1.47
34.99

PROJECT P_DESCRIPT and PRICE yields

P_DESCRIPT	PRICE
Flashlight	5.26
Lamp	25.15
Box Fan	10.99
9v battery	1.92
100W bulb	1.47
Powerdrill	34.99

PROJECT P_CODE and PRICE yields

P_CODE	PRICE
123456	5.26
123457	25.15
123458	10.99
213345	1.92
254467	1.47
311452	34.99

Relational Operations … continued

FIGURE 3.6 UNION

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99

UNION

P_CODE	P_DESCRIPT	PRICE
345678	Microwave	160.00
345679	Dishwasher	500.00
123458	Box Fan	10.99

yields

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99
345678	Microwave	160
345679	Dishwasher	500

FIGURE 3.7 INTERSECT

STU_FNAME	STU_LNAME
George	Jones
Jane	Smith
Peter	Robinson
Franklin	Johnson
Martin	Lopez

INTERSECT

EMP_FNAME	EMP_LNAME
Franklin	Lopez
William	Turner
Franklin	Johnson
Susan	Rogers

yields

STU_FNAME	STU_LNAME
Franklin	Johnson

Union-compatible: tables share the same number of columns, and their corresponding columns share compatible domains

Relational Operations … continued

FIGURE 3.8 DIFFERENCE

STU_FNAME	STU_LNAME
George	Jones
Jane	Smith
Peter	Robinson
Franklin	Johnson
Martin	Lopez

DIFFERENCE

EMP_FNAME	EMP_LNAME
Franklin	Lopez
William	Turner
Franklin	Johnson
Susan	Rogers

yields

STU_FNAME	STU_LNAME
George	Jones
Jane	Smith
Peter	Robinson
Martin	Lopez

Relational Operations … continued

FIGURE 3.9 PRODUCT

P_CODE	P_DESCRIPT	PRICE
123456	Flashlight	5.26
123457	Lamp	25.15
123458	Box Fan	10.99
213345	9v battery	1.92
254467	100W bulb	1.47
311452	Powerdrill	34.99

PRODUCT

STORE	AISLE	SHELF
23	W	5
24	K	9
25	Z	6

yields

P_CODE	P_DESCRIPT	PRICE	STORE	AISLE	SHELF
123456	Flashlight	5.26	23	W	5
123456	Flashlight	5.26	24	K	9
123456	Flashlight	5.26	25	Z	6
123457	Lamp	25.15	23	W	5
123457	Lamp	25.15	24	K	9
123457	Lamp	25.15	25	Z	6
123458	Box Fan	10.99	23	W	5
123458	Box Fan	10.99	24	K	9
123458	Box Fan	10.99	25	Z	6
213345	9v battery	1.92	23	W	5
213345	9v battery	1.92	24	K	9
213345	9v battery	1.92	25	Z	6
311452	Powerdrill	34.99	23	W	5
311452	Powerdrill	34.99	24	K	9
311452	Powerdrill	34.99	25	Z	6
254467	100W bulb	1.47	23	W	5
254467	100W bulb	1.47	24	K	9
254467	100W bulb	1.47	25	Z	6

Relational Operations … continued

FIGURE 3.7 INTERSECT

STU_FNAME	STU_LNAME
George	Jones
Jane	Smith
Peter	Robinson
Franklin	Johnson
Martin	Lopez

INTERSECT

EMP_FNAME	EMP_LNAME
Franklin	Lopez
William	Turner
Franklin	Johnson
Susan	Rogers

yields

STU_FNAME	STU_LNAME
Franklin	Johnson

Join Operation

Joins allow information to be intelligently combined from two or more tables

- Natural join: links tables by selecting only the rows with common values in their common attribute
- Equijoin: links tables on the basis of an equality condition that compares specified columns of each table
- Theta join: links tables using an inequality comparison operator
- Inner join: only returns matched records from the tables that are being joined
- Outer join: matched pairs are retained and unmatched values in the other table are left null
- Left outer join: yields all of the rows in the first table, including those that do not have a matching value in the second table
- Right outer join: yields all of the rows in the second table, including those that do not have matching values in the first table
- Divide
  - Uses one double-column table as the dividend and one single-column table as the divisor
  - Output is a single column that contains all values from the second column of the dividend that are associated with every row in the divisor

Join Operation – Example Tables

FIGURE 3.10 TWO TABLES THAT WILL BE USED IN JOIN ILLUSTRATIONS

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_ZIP	AGENT_CODE
1132445	Walker	32145	231
1217782	Adares	32145	125
1312243	Rakowski	34129	167
1321242	Rodriguez	37134	125
1542311	Smithson	37134	421
1657399	Vanloo	32145	231

Table name: AGENT

AGENT_CODE	AGENT_PHONE
125	6152439887
167	6153426778
231	6152431124
333	9041234445

4 columns/attributes
6 rows/tuples

2 columns/attributes
4 rows/tuples

Natural Join

FIGURE 3.11 NATURAL JOIN, STEP 1: PRODUCT

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1132445	Walker	32145	231	125	6152439887
1132445	Walker	32145	231	167	6153426778
1132445	Walker	32145	231	231	6152431124
1132445	Walker	32145	231	333	9041234445
1217782	Adares	32145	125	125	6152439887
1217782	Adares	32145	125	167	6153426778
1217782	Adares	32145	125	231	6152431124
1217782	Adares	32145	125	333	9041234445
1312243	Rakowski	34129	167	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1312243	Rakowski	34129	167	231	6152431124
1312243	Rakowski	34129	167	333	9041234445
1321242	Rodriguez	37134	125	125	6152439887
1321242	Rodriguez	37134	125	167	6153426778
1321242	Rodriguez	37134	125	231	6152431124
1321242	Rodriguez	37134	125	333	9041234445
1542311	Smithson	37134	421	125	6152439887
1542311	Smithson	37134	421	167	6153426778
1542311	Smithson	37134	421	231	6152431124
1542311	Smithson	37134	421	333	9041234445
1657399	Vanloo	32145	231	125	6152439887
1657399	Vanloo	32145	231	167	6153426778
1657399	Vanloo	32145	231	231	6152431124
1657399	Vanloo	32145	231	333	9041234445

FIGURE 3.12 NATURAL JOIN, STEP 2: SELECT

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1217782	Adares	32145	125	125	6152439887
1321242	Rodriguez	37134	125	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1132445	Walker	32145	231	231	6152431124
1657399	Vanloo	32145	231	231	6152431124

FIGURE 3.13 NATURAL JOIN, STEP 3: PROJECT

CUS_CODE	CUS_LNAME	CUS_ZIP	AGENT_CODE	AGENT_PHONE
1217782	Adares	32145	125	6152439887
1321242	Rodriguez	37134	125	6152439887
1312243	Rakowski	34129	167	6153426778
1132445	Walker	32145	231	6152431124
1657399	Vanloo	32145	231	6152431124

Left Outer Join

Note that there is an assumption here that AGENT_CODE IN the table CUSTOMER is not FK

FIGURE 3.10 TWO TABLES THAT WILL BE USED IN JOIN ILLUSTRATIONS

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_ZIP	AGENT_CODE
1132445	Walker	32145	231
1217782	Adares	32145	125
1312243	Rakowski	34129	167
1321242	Rodriguez	37134	125
1542311	Smithson	37134	421
1657399	Vanloo	32145	231

Table name: AGENT

AGENT_CODE	AGENT_PHONE
125	6152439887
167	6153426778
231	6152431124
333	9041234445

FIGURE 3.14 LEFT OUTER JOIN

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1217782	Adares	32145	125	125	6152439887
1321242	Rodriguez	37134	125	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1132445	Walker	32145	231	231	6152431124
1657399	Vanloo	32145	231	231	6152431124
1542311	Smithson	37134	421		

Right Outer Join

FIGURE 3.10 TWO TABLES THAT WILL BE USED IN JOIN ILLUSTRATIONS

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_ZIP	AGENT_CODE
1132445	Walker	32145	231
1217782	Adares	32145	125
1312243	Rakowski	34129	167
1321242	Rodriguez	37134	125
1542311	Smithson	37134	421
1657399	Vanloo	32145	231

Table name: AGENT

AGENT_CODE	AGENT_PHONE
125	6152439887
167	6153426778
231	6152431124
333	9041234445

FIGURE 3.15 RIGHT OUTER JOIN

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1217782	Adares	32145	125	125	6152439887
1321242	Rodriguez	37134	125	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1132445	Walker	32145	231	231	6152431124
1657399	Vanloo	32145	231	231	6152431124
				333	9041234445

Equi-join & Theta-join

Similar to Natural Join:

- Create Product
- Select using the relational arithmetic operation Theta (θ), where θ is one of the set { =, $\neq$, $\geq$, $\leq$, >, < }
- There is **no projection** in comparison to Natural Join

FIGURE 3.11 NATURAL JOIN, STEP 1: PRODUCT

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1132445	Walker	32145	231	125	6152439887
1132445	Walker	32145	231	167	6153426778
1132445	Walker	32145	231	231	6152431124
1132445	Walker	32145	231	333	9041234445
1217782	Adares	32145	125	125	6152439887
1217782	Adares	32145	125	167	6153426778
1217782	Adares	32145	125	231	6152431124
1217782	Adares	32145	125	333	9041234445
1312243	Rakowski	34129	167	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1312243	Rakowski	34129	167	231	6152431124
1312243	Rakowski	34129	167	333	9041234445
1321242	Rodriguez	37134	125	125	6152439887
1321242	Rodriguez	37134	125	167	6153426778
1321242	Rodriguez	37134	125	231	6152431124
1321242	Rodriguez	37134	125	333	9041234445
1542311	Smithson	37134	421	125	6152439887
1542311	Smithson	37134	421	167	6153426778
1542311	Smithson	37134	421	231	6152431124
1542311	Smithson	37134	421	333	9041234445
1657399	Vanloo	32145	231	125	6152439887
1657399	Vanloo	32145	231	167	6153426778
1657399	Vanloo	32145	231	231	6152431124
1657399	Vanloo	32145	231	333	9041234445

FIGURE 3.12 NATURAL JOIN, STEP 2: SELECT

CUS_CODE	CUS_LNAME	CUS_ZIP	CUSTOMER.AGENT_CODE	AGENT.AGENT_CODE	AGENT_PHONE
1217782	Adares	32145	125	125	6152439887
1321242	Rodriguez	37134	125	125	6152439887
1312243	Rakowski	34129	167	167	6153426778
1132445	Walker	32145	231	231	6152431124
1657399	Vanloo	32145	231	231	6152431124

Selection is done using the Theta (θ) operator

If θ is

"=" … Equi-join

">" … Greater-than Join

…

Divide

DIVIDE … used to answer questions about one set of data being associated with all values of data in another set of data

FIGURE 3.16 DIVIDE

P_CODE	CUS_CODE
123456	10400
123456	11501
123456	10030
123456	12550
234567	12350
234567	10040
234567	10500
234567	10030
234567	12550
345678	10400
345678	11630
345678	12550
456789	11630
567890	10500
567890	10030
567890	12550
678901	11500
678901	10400
678901	11630

DIVIDE

P_CODE
123456
234567
567890

yields

CUS_CODE
10030
12550

Data Dictionaries / Catalogues

- Data dictionary
  - Description of all tables in the database created by the user and designer
- System catalog
  - System data dictionary that describes all objects within the database

- Homonyms and synonyms must be avoided to lessen confusion
  - Homonym: same name is used to label different attributes
    - Example: pen … writing instrument / an enclosure to hold animals
  - Synonym: different names are used to describe the same attribute
    - Example: happy / joyful, help / aid

Ozsu Book – Sample DB

EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

Relational Algebra Operators (from Ozsu book)

- Fundamental
  - Selection
  - Projection
  - Union
  - Set difference
  - Cartesian product
- Additional
  - Intersection
  - θ-join
  - Natural join
  - Division
- Union compatibility
  - Same degree
  - Corresponding attributes defined over the same domain

Relational Operators from Ozsu Book - examples

Fundamental/Basic Rel. Algebra Operators

Selection	$\sigma_{TITLE="Elect.\ Eng."}(EMP)$
Projection	$\Pi_{PNO,BUDGET}(PROJ)$
Union	$R \cup S$
Set difference	R - S
Cartesian product	R x S

Additional Operators & Notation

- Additional Operators
  - Intersection $R \cap S = R - (R - S)$
  - θ-join $R \bowtie_F S$
  - Natural join $R \bowtie S$
  - Semijoin $R \ltimes_F S$
  - Division $R \div S$

- Set **operators**.
- Projection (Π)
- Selection (σ)
- Rename (ρ)
- Natural **join** (⋈)
- θ-**join** and equijoin.
- Semijoin (⋉)(⋊)

Relational Algebra - Division

- Given relations
  - R of degree k_1 ($R = \{A_1,\ldots,A_{k_1}\}$)
  - S of degree k_2 ($S = \{B_1,\ldots,B_{k_2}\}$)

Let $A = \{A_1,\ldots,A_{k_1}\}$ [i.e., $R(A)$] and $B = \{B_1,\ldots,B_{k_2}\}$ [i.e., $S(B)$] and $B \subseteq A$.

Then, $T = R \div S$ gives T of degree k_1-k_2 [i.e., $T(Y)$ where $Y = A$-B]
such that for a tuple t to appear in T,
the values in t must appear in R
in combination with *every tuple* in S.

Relational Algebra – Division … continued

Tables: D (x, y, z, a, b) S (a, b, c)

Result of: R = D : S … columns and rows ?

- *Columns in the result R: Columns of D minus columns in S*
- => R is an empty relation as D does not have a column that corresponds to the column c in the table S

Relational Algebra – Division … continued

Tables: D (X, Y, Z, A, B) S (A, B)

Result of: R = D : S … columns and rows ?

- Columns: R (X, Y, Z) … Columns of X minus columns in A
- Tuples/rows of R for the example R = D : S
  - For a tuple (x1, y1, z1) to be the in the result R
    - x1, y1, z1 must be in D combine with each tuple of S:
      - x1, y1, z1, a1, b1
      - x1, y1, z1, a2, b2
      - x1, y1, z1, a3, b3
  - For a tuple (x2, y2, z2) to be the in the result R
    - x2, y2, z2 must be in D combine with each tuple of S:
      - x2, y2, z2, a1, b1
      - x2, y2, z2, a2, b2
      - x2, y2, z2, a3, b3

TABLE D

X	Y	Z	A	B
x1	y1	z1	a1	b1
…	…	…	…	…
x1	y1	z1	a2	b2
…	…	…	…	…
x1	y1	z1	a3	b3
			…	…
x2	y2	z2	a1	b1
…	…	…	…	…
x2	y2	z2	a2	b2
x2	y2	z2	a3	b3

Result table R: (X, Y, Z)

Table S

A	B
a1	b1
a2	b2
a3	b3

Relational Algebra – Division … continued

- Example: Find each employee who works on both projects P3 and P4
- $ASG' = \Pi_{ENO, PNO}\, ASG$
- $PROJ' = \Pi_{PNO}\, (\sigma_{(PNO='P3' \vee PNO='P4')}\, PROJ\ )$
- *RESULT = ASG' ÷ PROJ'*

EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

(ASG' ÷ PROJ')

ENO
E3

Relational Algebra – Division … continued

- Example: Find each project that has all Mech Eng working on it
  (i.e., project that has all employees with title Mech Eng)
  - Result' = ASG ÷ EMP … ?
- $E' = \Pi_{ENO}(\sigma_{(TITLE='MECH.ENG.')} EMP)$ … { E3, E7 }
  - Result'' = ASG ÷ E' … ?
- $A' = \Pi_{ENO, PNO} ASG$
  - *RESULT = A' ÷ E'*

$$\frac{PNO}{P3}$$

EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

Relational Algebra - Conclusion

- More on Relational Algebra
  - E.g.,
    - https://en.wikipedia.org/wiki/Relational_algebra#Introduction

Questions / Answers