Databases and Information Systems
CS303

Relational Algebra and Relational Calculus

24-08-2023

Formal Relational Query Languages

- Formal way to describe queries
- Analogous to Algorithmic analysis
  - Language independent
- Types:
  - Relational Algebra
  - Tuple Relational Calculus
  - Domain Relational Calculus
- Used for:
  - Mathematical framework to analyze queries
  - Query Representation
  - Query optimization

Relational Algebra

Relational Algebra

- Procedural:
  - Specifies what is the required result and also how to obtain it.
- Takes one or more relation as input and produces a new relation as a result
- One operator for every basic step that we have in the query

Basic Relational Operators: SELECT

- Denoted by σ
- Corresponds to WHERE of SQL queries
- $\sigma_{condition}$(relation)
- $\sigma_{dept\_name = \text{'Physics'}}$(instructor)

ID	name	dept_name	salary
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

ID	name	dept_name	salary
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

Basic Relational Operators: SELECT

- $\sigma_{\text{dept\_name = 'Physics'} \wedge \text{salary > 90000}}$ (instructor)
- > >= < <= are all allowed
- AND (Λ) OR (V) can be specified in the condition

ID	name	dept_name	salary
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

ID	name	dept_name	salary
22222	Einstein	Physics	95000

Basic Relational Operators: PROJECT

- Denoted by π
- $\pi_{list\_of\_attributes}(relation\_name)$
- $\pi_{ID,\ name,\ dept\_name}(instructor)$

ID	*name*	*dept_name*	*salary*
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

`instructor`

Basic Relational Operators: PROJECT

- $\Pi_{\text{ID, name, dept\_name}} (\sigma_{\text{dept\_name = 'Physics'}} (\text{instructor}))$

ID	*name*	*dept_name*	*salary*
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

`instructor`

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

Basic Relational Operators: PROJECT

- $\Pi_{\text{ID, name, dept\_name}} \left(\sigma_{\text{dept\_name = 'Physics'}} (\text{instructor}) \right)$

ID	*name*	*dept_name*	*salary*
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

`instructor`

ID	*name*	*dept_name*
22222	Einstein	Physics
33456	Gold	Physics

Basic Relational Operators: UNION

- Denoted by ∪
- $\pi_{course\_id}(\sigma_{semester = 'Spring' \wedge year = 2009}(section))$
- $\pi_{course\_id}(\sigma_{semester = Fall \wedge year = 2009}(section))$

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
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	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section

- $\pi_{course\_id}(\sigma_{semester = 'Spring' \wedge year = 2009}(section)) \cup \pi_{course\_id}(\sigma_{semester = Fall \wedge year = 2009}(section))$

Basic Relational Operators: UNION

- $\pi_{course\_id}\ (\sigma_{semester\ =\ 'Spring'\ \wedge\ year\ =\ 2009}\ (section))\ \cup\ \pi_{course\_id}\ (\sigma_{semester\ =\ Fall\ \wedge\ year\ =\ 2009}\ (section))$
- No duplicates (True for all relational operators)

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
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	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section

course_id
CS-101
CS-315
CS-319
CS-347
FIN-201
HIS-351
MU-199
PHY-101

Basic Relational Operators: UNION

- r U s is well defined only if the following holds:
  - Both r and s should have same number of attributes
  - Domain of i-th attribute of r should be the same as the domain of i-th attribute of s

- r and s can be relations from the database of other relational algebraic expressions

Basic Relational Operators: Set Difference

- Denoted by —
- $\pi_{course\_id}(\sigma_{semester = 'Spring' \wedge year = 2009}(section))$
- $\pi_{course\_id}(\sigma_{semester = Fall \wedge year = 2009}(section))$

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
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	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section

- $\pi_{course\_id}(\sigma_{semester = 'Spring' \wedge year = 2009}(section)) - \pi_{course\_id}(\sigma_{semester = Fall \wedge year = 2009}(section))$

Basic Relational Operators: Set Difference

- $\pi_{course\_id}\ (\sigma_{semester\ =\ \text{'Spring'}\ \land\ year\ =\ 2009}\ (section))\ -\ \pi_{course\_id}\ (\sigma_{semester\ =\ Fall\ \ \land\ year\ =\ 2009}\ (section))$
- No duplicates (True for all relational operators)

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
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	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section

course_id
CS-347
PHY-101

Basic Relational Operators: Cartesian Product

- Denoted by
- $\sigma_{dept\_name='Physics'}$ (instructor X teaches)

ID	name	dept_name	salary
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

teaches

inst.ID	name	dept_name	salary	teaches.ID	course_id	sec_id	semester	year
22222	Einstein	Physics	95000	10101	CS-437	1	Fall	2009
22222	Einstein	Physics	95000	10101	CS-315	1	Spring	2010
22222	Einstein	Physics	95000	12121	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
22222	Einstein	Physics	95000	32343	HIS-351	1	Spring	2010
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
33456	Gold	Physics	87000	10101	CS-437	1	Fall	2009
33456	Gold	Physics	87000	10101	CS-315	1	Spring	2010
33456	Gold	Physics	87000	12121	FIN-201	1	Spring	2010
33456	Gold	Physics	87000	15151	MU-199	1	Spring	2010
33456	Gold	Physics	87000	22222	PHY-101	1	Fall	2009
33456	Gold	Physics	87000	32343	HIS-351	1	Spring	2010
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

Basic Relational Operators: Set Difference

- Denoted by X
- $\Pi_{\text{name, course\_id}} (\sigma_{\text{instructor.id = teaches.id}} (\sigma_{\text{dept\_name='Physics'}} (\text{instructor X teaches})))$

inst.ID	*name*	*dept_name*	*salary*	*teaches.ID*	*course_id*	*sec_id*	*semester*	*year*
22222	Einstein	Physics	95000	10101	CS-437	1	Fall	2009
22222	Einstein	Physics	95000	10101	CS-315	1	Spring	2010
22222	Einstein	Physics	95000	12121	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
22222	Einstein	Physics	95000	32343	HIS-351	1	Spring	2010
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
33456	Gold	Physics	87000	10101	CS-437	1	Fall	2009
33456	Gold	Physics	87000	10101	CS-315	1	Spring	2010
33456	Gold	Physics	87000	12121	FIN-201	1	Spring	2010
33456	Gold	Physics	87000	15151	MU-199	1	Spring	2010
33456	Gold	Physics	87000	22222	PHY-101	1	Fall	2009
33456	Gold	Physics	87000	32343	HIS-351	1	Spring	2010
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

name	*course_id*
Einstein	PHY-101

Basic Relational Operators: Renaming

- Denoted by ρ
- ρ_X (A1, A2...An) ( )
- ρ_X ( )
- $\rho_{\text{temp\_inst (temp\_id, temp\_name)}} (\pi_{\text{ID, name}} (\text{instructor}))$

ID	*name*	*dept_name*	*salary*
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

Basic Relational Operators: Renaming

- Find the highest salary in the university
- $\pi_{instructor.salary\ (instructor)} - \pi_{instructor.salary}\ (\sigma_{instructor.salary\ <\ d.salary}\ (\ instructor \times\ \rho_{d}\ (\ instructor\ )\ )\ )$

ID	name	dept_name	salary
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

salary
65000
90000
40000
60000
87000
75000
62000
72000
80000
92000

salary
95000

Basic Relational Algebra: Formal Definition

- Basic expressions:
  - A relation in the database : Example: instructor, takes, ...
  - Constant relation
    - Example: { (22222, Einstein, Physics, 95000), (76543, Singh, Finance, 80000) }

- If E1 and E2 are already constructed then the following are relational algebraic expressions:
  - E1 U E2
  - E1 - E2
  - E1 X E2
  - σ_p (E1) where P is a condition on some attribute of E1
  - π_S (E1) where S is a subset of attributes of E1
  - $\rho_{X(A1..An)}$ (E1) where X(A1...An) is a new name for the result of E1

Additional Relational Operators

- Basic relational operators are sufficient to express any relational algebraic query
- But it becomes lengthy if we use only those
- Additional operators do not add any expressive power but simplify the queries

Additional Relational Operators : Intersection

- Denoted by $\cap$
- Π_{course_id} $(\sigma_{semester\ =\ 'Spring'\ \wedge\ year\ =\ 2009}$ (instructor)) $\cap$ Π_{course_id} $(\sigma_{semester\ =\ Fall\ \wedge\ year\ =\ 2009}$ (instructor))
- $r \cap s = r - (r - s)$

Additional Relational Operators : Natural Join

- Denoted
- instructor ⋈ teaches

b

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

ID	name	dept_name	salary
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	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

instructor

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

teaches

Additional Relational Operators : Natural Join

- Let r and s be two relational algebraic expressions with R and S as the set of attributes of r and s respectively. Let $R \cap S = \{A1, A2.. An\}$ Then:

- $r \bowtie s = \pi_{R \cup S} (\sigma_{r.A1 = s.A1 \wedge r.A2 = s.A2 \wedge .. r.An = s.An} (r \times s))$

- $\bowtie$ is associative: $(r \bowtie s) \bowtie t = r \bowtie (s \bowtie t) = r \bowtie s \bowtie t$

- Is $\bowtie$ order invariant? $r \bowtie s \bowtie t = r \bowtie t \bowtie s$?

Additional Relational Operators : Theta Join

- Variant of Natural join that combines select and cartesian product
- $r \bowtie_{\theta} s$ where θ is a condition on the attributes of r and s
- $r \bowtie_{\theta} s = \sigma_{\theta}(r \times s)$

Additional Relational Operators : Assignment operator

- Denoted by $\leftarrow$

- $R1 \leftarrow r \times s$
- $R2 \leftarrow \pi_A(\sigma_{p1}(R1))$
- $Result \leftarrow \sigma_{p2}(R2)$

- Result is equivalent to $\sigma_{p2}(\pi_A(\sigma_{p1}(r \times s)))$

Additional Relational Operators : Outer Join

- Left Outer Join : $r ⟕ s$
- Right Outer Join : $r ⟖ s$
- Full Outer Join : $r ⟗ s$

Additional Relational Operators : Outer Join

- Left Outer Join : $r ⟕ s = (r \bowtie s) \cup ((r - (r \bowtie s)) \times (null, null, ...null))$
- Right Outer Join : $r ⟖ s$ (exercise)
- Full Outer Join : $r ⟗ s$ (exercise)

Other Relational Operators

- There are many other operators like Division, Semi-join, Anti semi-join etc.
- All these can be expressed using the basic operators.
- Used as and when required