CS2102 Database Systems

Relational Calculus

NUS National University of Singapore | School of Computing

Leading The World With Asia's Best

Background

- The concept of relational calculus was first proposed by Edgar F. Codd. See http://en.wikipedia.org/wiki/Edgar_F._Codd

- Relational calculus is based on predicate calculus, and operates on relations

- It is used to measure the power of relational languages. A language that can be used to produce any relation which can be derived using relational calculus is said to be *relationally complete*.

Relational Calculus

- Relational calculus is declarative
  - Describes a relation in terms of one or more database relations
  - Query languages, like SQL, are similar in construction to relational calculus

- Relational algebra is procedural
  - Describes how to build a new relation from one or more relations in the database

Relational Calculus

- Basic concepts: variables, constants, comparison ops, logical connectors, quantifiers
- Two variants
  - Tuple Relational Calculus (TRC)

    - Variables range over tuples

    e.g. $\{C \mid C \in Course \wedge C.credit \geq 4\}$
  - Domain Relational Calculus (DRC)

    - Variables range over attribute domain values

    e.g. $\{<I, N, C> \mid <I, N, C> \in Course \wedge C > 4\}$

Relational Calculus

- Expressions in a calculus are called formulas
- An answer tuple is an assignment of constants to variables that make the formula evaluate to true

e.g. $\{C \mid C \in \text{Course} \wedge C.\text{credit} \geq 4\}$

Tuple Relational Calculus

- Query has the form {T | p(T)}
  - T is a tuple variable
  - p(T) is a formula that describes T
  - Result of query is the set of all tuples *t* for which the formula p(T) evaluates to true with T = *t*

e.g. $\{C \mid C \in \text{Course} \wedge C.\text{credit} \geq 4\}$

Simple TRC Queries

- Find all lecturers

  $\{T \mid T \in Lecturer\}$

Lecturer

id	**name**
9876	Stephane Bressan
9865	Mong Li Lee
9843	Shen Hengtao
9821	Ling Tok Wang

- Find courses where the credit is at least 4

  $\{C \mid C \in Course \wedge C.credit \geq 4\}$

Course

id	**name**	**credit**
1234	Intro to computing	2
2107	Intro to DB	4
2234	Intro to OS	2
3604	Advanced OS	4

Syntax of TRC Queries

- An atomic formula is one of the following:
  - R $\in$ Rel
  - R.*a* op S.*b*
  - R.*a* op constant

where

  - Rel is a relation name
  - R, S are tuple variables with attributes *a* and *b* respectively
  - op is an operator in the set {>, <, =, ≥, ≤, ≠}

e.g. $\{C \mid C \in \text{Course} \wedge C.\text{credit} \geq 4\}$

Syntax of TRC Queries

- A formula is recursively defined as one of the following:
  - Any atomic formula
  - $\neg$ p, p $\wedge$ q, p $\vee$ q, p $\Rightarrow$ q where p and q are formulas
  - $\exists$ R (p(R)) where R is a tuple variable and p(R) is a formula in which R appears
  - $\forall$ R (p(R)) where R is a tuple variable and p(R) is a formula in which R appears

  e.g. $\{C \mid C \in \text{Course} \wedge C.\text{credit} \geq 4\}$

Combination of Formulas

p	q	$\neg$ p	p $\wedge$ q	p $\vee$ q	p $\Rightarrow$ q
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

Example Database

- Sailors (sid, name, rating, age)
- Boats (bid, bname, color)
- Reserves (sid, bid, day)

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Example TRC Queries

- Find the names and ages of sailors with a rating above 7

$$\{P \mid \exists S\,(S \in Sailors \wedge S.rating > 7 \wedge P.name = S.sname \wedge P.age = S.age)\}$$

- P is a tuple variable with two fields, name and age
- Result of query is a relation with two fields, name and age
- Atomic formula P.name=S.sname and P.age=S.age give values to the fields of an answer tuple P

Example TRC Queries

- Find the names and ages of sailors with a rating above 7

$$\{P \mid \exists S \, (S \in \text{Sailors} \wedge S.\text{rating} > 7 \wedge P.\text{name} = S.\text{sname} \wedge P.\text{age} = S.\text{age})\}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

name	age
Lubber	55
Andy	25
Rusty	35
Zorbia	16
Horatio	35

Example TRC Queries

- Find the names and ages of sailors with a rating above 7

  $\{P \mid \exists S\ (S \in Sailors \wedge S.rating > 7 \wedge$

  $P.name = S.sname \wedge P.age = S.age)\}$

- Compare to SQL:

```
SELECT S.sname, S.age
FROM Sailors S
WHERE S.rating > 7
```

Example TRC Queries

- Find the sailor name, boat id and reservation date for each reservation

$$\{ P \mid \exists R \ \exists S \ (R \in \text{Reserves} \wedge S \in \text{Sailors} \wedge R.\text{sid} = S.\text{sid} \wedge P.\text{bid} = R.\text{bid} \wedge P.\text{date} = R.\text{date} \wedge P.\text{sname} = S.\text{sname}) \}$$

- For each Reserves tuple, look for a Sailors tuple with the same sid
- For each pair of such tuples, construct answer tuple P with fields sname, bid and day by copying values from the corresponding fields from the two tuples

Example TRC Queries

- Find the sailor name, boat id and reservation date for each reservation

$$\{ P \mid \exists R \ \exists S \ (R \in \text{Reserves} \wedge S \in \text{Sailors} \wedge R.\text{sid} = S.\text{sid} \ \wedge \ P.\text{bid} = R.\text{bid} \ \wedge \ P.\text{date} = R.\text{date} \ \wedge \ P.\text{sname} = S.\text{sname}) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

bid	date	sname
101	10/10/98	Dustin
102	10/10/98	Dustin
103	10/8/98	Dustin
104	10/7/98	Dustin
102	11/10/98	Lubber
103	11/6/98	Lubber
104	11/12/98	Lubber
101	9/5/98	Horatio
102	9/8/98	Horatio
103	9/8/98	Horatio

Example TRC Queries

- Find the names of sailors who have reserved boat 103

$$\{ P \mid \exists S \ \exists R \ (S \in Sailors \wedge R \in Reserves \wedge R.sid = S.sid \wedge R.bid = 103 \wedge P.sname = S.sname) \}$$

- For each Sailors tuple, look for a tuple in Reserves that shows that this sailor has reserved boat 103
- Answer tuple P contains just one field, sname

Example TRC Queries

- Find the names of sailors who have reserved boat 103

$$\{ P \mid \exists S \ \exists R (S \in \text{Sailors} \wedge R \in \text{Reserves} \wedge R.\text{sid} = S.\text{sid} \wedge R.\text{bid} = 103 \wedge P.\text{sname} = S.\text{sname}) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

sname
Dustin
Lubber
Horatio

Example TRC Queries

- Find the names of sailors who have reserved boat 103

  $\{ P \mid \exists S \ \exists R \ (S \in Sailors \wedge R \in Reserves \wedge$
  $R.sid = S.sid \wedge R.bid = 103 \wedge P.sname = S.sname) \}$

- Compare to SQL:

```
SELECT S.sname
FROM Sailors S, Reserves R
WHERE S.sid= R.sid AND R.bid = 103
```

Example TRC Queries

- Find the names of sailors who have reserved a red boat

  $\{ P \mid \exists S \ \exists R \exists B \ (S \in Sailors \wedge R \in Reserves \wedge B \in Boats$
  $\wedge \ R.sid = S.sid \ \wedge \ R.bid = B.bid \ \wedge \ B.color = \text{'red'}$
  $\wedge \ P.sname = S.sname) \}$

- Retrieve all sailor tuples S for which there exist tuples R in Reserves and B in Boats such that S.sid=R.sid, R.bid=B.bid and B.color='red'

Example TRC Queries

- Find the names of sailors who have reserved a red boat

$$\{ P \mid \exists S \exists R \exists B (S \in \text{Sailors} \wedge R \in \text{Reserves} \wedge B \in \text{Boats} \wedge R.\text{sid} = S.\text{sid} \wedge R.\text{bid} = B.\text{bid} \wedge B.\text{color} = \text{'red'} \wedge P.\text{sname} = S.\text{sname}) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

sname
Dustin
Lubber
Horatio

Example TRC Queries

- Find the names of sailors who have reserved at least two boats

$$\{ P \mid \exists S \ \exists R1 \ \exists R2 \ (S \in Sailors \wedge R1 \in Reserves \wedge R2 \in Reserves \wedge S.sid = R1.sid \wedge R1.sid = R2.sid \wedge R1.bid \neq R2.bid \wedge P.sname = S.sname) \}$$

Example TRC Queries

- Find the names of sailors who have reserved at least two boats

$$\{ P \mid \exists S \ \exists R1 \exists R2 (S \in Sailors \wedge R1 \in Reserves \wedge R2 \in Reserves \wedge S.sid = R1.sid \wedge R1.sid = R2.sid \wedge R1.bid \neq R2.bid \wedge P.sname = S.sname) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

sname
Dustin
Lubber
Horatio

Example TRC Queries

- Find the names of sailors who have reserved **all** boats

  $\{ P \mid \exists S \ \forall B \ \exists R \ (S \in Sailors \wedge$

$(B \in Boats \Rightarrow (R \in Reserves \wedge S.sid = R.sid \wedge R.bid = B.bid))$

$\wedge \ P.sname = S.sname) \}$

  - Retrieve sailors S such that for all boats B there is a Reserves tuple showing that sailor S has reserved boat B

Example TRC Queries

- Find the names of sailors who have reserved **all** boats

$$\{ P \mid \exists S \ \forall B \ \exists R \ (S \in \text{Sailors} \wedge (B \in \text{Boats} \Rightarrow (R \in \text{Reserves} \wedge S.\text{sid} = R.\text{sid} \wedge R.\text{bid} = B.\text{bid})) \wedge P.\text{sname} = S.\text{sname}) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

sname
Dustin

Example TRC Queries

- Find the names of sailors who have reserved all red boats

$$\{ P \mid \exists S \ \forall B \ \exists R \ (S \in Sailors \wedge$$
$$((B \in Boats \wedge B.color = \text{'red'}) \Rightarrow$$
$$(R \in Reserves \wedge S.sid = R.sid \wedge R.bid = B.bid))$$
$$\wedge P.sname = S.sname) \}$$

- For each candidate sailor tuple, if a boat is red, then the sailor must have reserved it.

- Find the names of sailors who have reserved all red boats

$$\{ P \mid \exists S \ \forall B \ \exists R \ (S \in Sailors \wedge ((B \in Boats \wedge B.color = \text{'red'}) \Rightarrow (R \in Reserves \wedge S.sid = R.sid \wedge R.bid = B.bid)) \wedge P.sname = S.sname) \}$$

Sailors

sid	sname	rating	age
22	Dustin	7	45
29	Brutus	1	33
31	Lubber	8	55
32	Andy	8	25
58	Rusty	10	35
64	Horatio	7	35
71	Zorbia	10	16
74	Horatio	9	35
85	Art	3	25
95	Bob	3	63

Reserves

sid	bid	date
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

sid	sname	rating	age
22	Dustin	7	45
31	Lubber	8	55

Domain Relational Calculus

- Query has the form $\{<x_1, \ldots, x_n> \mid p(x_1, \ldots, x_n)\}$
  - Each x_i is a domain variable or a constant
  - $p(x_1, \ldots, x_n)$ is a DRC formula
  - Result of query is the set of all tuples $<x_1, \ldots, x_n>$ for which the formula evaluates to true

Syntax of DRC Queries

- An atomic formula in DRC is one of the following:
  - $<x_1, \ldots, x_n> \in$ Rel
  - X op Y
  - X op constant

where

- Rel is a relation with n attributes
- Each x_i is a variable or a constant
- X, Y are domain variables
- op is an operator in the set {>, <, =, ≥, ≤, ≠}

Syntax of DRC Queries

- A formula in DRC is recursively defined as one of the following:
  - Any atomic formula
  - $\neg$ p, p $\wedge$ q, p $\vee$ q, p $\Rightarrow$ q where p and q are formulas
  - $\exists$ X (p(X)) where X is a domain variable and p(X) is a formula in which X appears
  - $\forall$ X (p(X)) where X is a domain variable and p(X) is a formula in which X appears

Example DRC Queries

- Find all sailors with a rating above 7

  $\{\langle I, N, T, A\rangle \mid \langle I, N, T, A\rangle \in Sailors \wedge T > 7\}$

- Each attribute is given a variable name
- Condition $\langle I, N, T, A\rangle \in$ Sailors ensures that domain variables are restricted to the fields of the same tuple
- Compare to TRC

  TRC: $\{P \mid \exists S (S \in Sailors \wedge S.rating > 7 \wedge P.name = S.sname \wedge P.age = S.age) \}$

  - Specify T > 7 instead of S.rating > 7
  - Specify tuple <I, N, T, A> in result instead of S

Example DRC Queries

- Find the names of sailors who have reserved boat 103

$\{ \langle N \rangle \mid \exists I \ \exists T \ \exists A \ \exists I_R \ \exists B_R \ \exists D$

$(\langle I, N, T, A \rangle \in Sailors \wedge \langle I_R, B_R, D \rangle \in Reserves \ \wedge$

$I = I_R \wedge B_R = 103) \}$

OR

$\{ \langle N \rangle \mid \exists I \ \exists T \ \exists A \ \exists D$

$(\langle I, N, T, A \rangle \in Sailors \wedge \langle I, 103, D \rangle \in Reserves) \}$

- Only the sname field is retained in answer

Example DRC Queries

- Find the names of sailors who have reserved a red boat

$$\{ \langle N \rangle \mid \exists I\ \exists T\ \exists A\ \exists B\ \exists D\ \exists BN$$
$$(\langle I, N, T, A \rangle \in Sailors \wedge \langle I, B, D \rangle \in Reserves \wedge$$
$$\langle B, BN, \text{‘red’} \rangle \in Boats) \}$$

Example DRC Queries

- Find the names of sailors who have reserved at least two boats

$$\{ \langle N \rangle \mid \exists I \exists T \exists A \exists B1 \exists B2 \exists D1 \exists D2$$
$$(\langle I, N, T, A \rangle \in Sailors \wedge \langle I, B1, D1 \rangle \in Reserves \wedge$$
$$\langle I, B2, D2 \rangle \in Reserves \wedge B1 \neq B2) \}$$

- Repeated use of variable I ensures that the same sailor has reserved both the boats

Example DRC Queries

- Find the names of sailors who have reserved all boats

$$\{ \langle N \rangle \mid \exists I\ \exists T\ \exists A\ \forall B\ \forall BN\ \forall C\ \exists D$$
$$(\langle I, N, T, A \rangle \in Sailors\ \wedge$$
$$(\langle B, BN, C \rangle \in Boats \Rightarrow\ \langle I, B, D \rangle \in Reserves)\ )\ \}$$

OR

$$\{ \langle N \rangle \mid \exists I\ \exists T\ \exists A\ \forall B\ \forall BN\ \forall C\ \exists D$$
$$(Sailors\ (I, N, T, A)\ \wedge$$
$$(Boats\ (B, BN, C) \Rightarrow Reserves\ (I, B, D)\ )\ )\ \}$$

Example DRC Queries

- Find the names of sailors who have reserved all red boats

  $\{ \langle N \rangle \mid \exists I \ \exists T \ \exists A \ \forall B \ \forall BN \ \forall C \ \exists D$

  $(\langle I, N, T, A \rangle \in Sailors \ \wedge$

  $((\langle B, BN, C \rangle \in Boats \wedge C = \text{'red'}) \Rightarrow \langle I, B, D \rangle \in Reserves)) \}$

- Find all sailors such that, for every red boat, there is a tuple in Reserves that shows the sailor has reserved it.

Safety of Queries in Relational Calculus

- A query is **safe** if the evaluation result is a finite set of tuples; otherwise, it is **unsafe**.
- Query {S | ¬ (S ∈ Sailors)} is unsafe
  - Ask for all tuples S such that S is not in Sailors
  - Answer set is infinite

Summary

- Relational calculus is a non-procedural formal query language
- Relational calculus has the same expressive power as relational algebra
  - Every query that can be expressed in relational algebra can be expressed as a safe query in TRC/DRC; and vice versa
- A query language (e.g. SQL) is relationally complete if it can express every query that is expressible in relational algebra/calculus