On the Correspondence Between the Tuple Relational Calculus (TRC) and SQL (Draft)

Dirk Van Gucht¹

¹Indiana University

Learning outcomes

- Pure SQL is a formal logic:
 - It can express statements (constraints-boolean queries)
 - It can express set (relation) abstractions (queries)
- Pure SQL corresponds in a one-to-one manner with Tuple Relation Calculus (a variant of Predicate Logic):¹
 - The semantics of TRC provides the semantics for Pure SQL: this is at the basis of correct query evaluation and constraint verification
 - The proof theory of TRC provides techniques to rewrite SQL queries into equivalent queries: this is
 at the basis of query optimization which is essential for efficient query evaluation and constaint
 verification

¹More accurately, Pure SQL is equivalent with *SafeTuple Relation Calculus*, a logic that places some restrictions on the free tuples in TRC formulas

Overview

- Tuple Relational Calculus (TRC) syntax
- Tuple Relation Calculus (TRC) constraints
- Tuple Relational Calculus (TRC) queries
- Expressing TRC constraints as Boolean SQL queries
- Expressing TRC queries as SQL queries

Database Model for TRC

The database model for TRC is the *Relational Database Model*. Formally,

- a set of *attributes* A;
- each attribute A has as an associated domain dom(A).
- a set of *relation schemas* \mathcal{R} ;
- each relation schema R has an associated list of attributes att(R).
- a *relational database schema* \mathbf{R} is a sequence of relation schemas R_1, R_2, \dots, R_n :
 - $\mathbf{R} = (R_1, \dots, R_n).$

Example: database schema

```
Student(<u>sid</u>:integer, sname:text)
Book(<u>bookno</u>:integer, title:text, price:integer)
Buys(sid:integer, bookno:integer)
```

TRC terms

Terms serve to represent objects in TRC formulas.

- A tuple variable t has an associated sequence of attributes att(t) and will represent a tuple object with components across this sequence of attributes;
- If a tuple variable t has $\operatorname{att}(t) = (A_1, \dots, A_k)$, then t has a sequence of associated *tuple variable components* denoted by $t.A_1, \dots, t.A_k$, and these represent objects in $\operatorname{dom}(A_1)$ through $\operatorname{dom}(A_k)$ respectively; and
- for each domain value of an attribute there is a constant that names this domain value.

TRC Formulas $\mathcal{F}(R)$ (Syntax)

$R(t)$ $t.A\theta s.B$ $t.A\theta a$	$att(R) = att(t) \ dom(A) = dom(B) \ a \in dom(A)$	$\theta:=,\neq,<,\leq,>,\textit{or}\geq\\\theta:=,\neq,<,\leq,>,\textit{or}\geq$
$ \begin{vmatrix} F_1 \land F_2 \\ F_1 \lor F_2 \\ F_1 \to F_2 \\ \neg F \end{vmatrix} $	conjunction (and) disjunction (or) implication (if then) negation (not)	
∃ <i>t F</i> ∀ <i>t F</i>	existential quantification universal quantification	
(<i>F</i>)		

Precedence order (\succ): $\neg \succ \land \succ \lor \succ \rightarrow$

TRC Formulas $\mathcal{F}(R)$ in abbreviated form (Syntax)

$$R(t)$$

 $t.A\theta s.B$ $att(R) = att(t)$
 $dom(A) = dom(B)$ $\theta := , \neq , <,$
 $\leq , > , or \geq$ $t.A\theta a$ $a \in dom(A)$ $\theta := , \neq , <,$
 $\leq , > , or \geq$ $F_1 \wedge F_2$
 $F_1 \vee F_2$
 $F_1 \rightarrow F_2$
 $\neg F$ conjunction (and)
disjunction (or)
implication (if then)
negation (not) $\exists t F$
 $\forall t F$
 $\exists t_1 \in R_1, \dots, t_k \in R_k(F)$ existential quantification
universal quantification
series of existential quantifiers

series of universal quantifiers

 $\exists t_1 \in R_1, \ldots, t_k \in R_k(F)$ $\forall t_1 \in R_1, \ldots, t_k \in R_k(F \to G)$

TRC Formulas $\mathcal{F}(R)$ in abbreviated form (Syntax)

$$\exists t_1 \in R_1, \ldots, t_k \in R_k(F)$$

is an abbreviated form for the formula

$$\exists t_1 \ldots \exists t_k (R_1(t_1) \wedge \cdots \wedge R_k(t_k) \wedge F)$$

$$\forall t_1 \in R_1, \ldots, t_k \in R_k(F \to G)$$

is an abbreviated form for the formula

$$\forall t_1 \ldots \forall t_k ((R_1(t_1) \wedge \cdots \wedge R_k(t_k) \wedge F) \rightarrow G)$$

Formulas (Examples)

 $\exists \, \textit{s}(\texttt{Student}(\textit{s}) \land \exists \, \textit{t}(\texttt{Buys}(\textit{t}) \land \textit{s.sname} = \textbf{Eric} \land \textit{t.sid} = \textit{s.sid}))$

 $\exists s \exists t (Student(s) \land Buys(t) \land s.sname = Eric \land t.sid = s.sid)$

 $\exists s (\mathtt{Student}(s) \land \forall t ((\mathtt{Buys}(t) \land t.sid = s.sid) \rightarrow t.bookno \neq 1000)$

Formulas in abbreviated form

```
Student(s)
Student(s) \land s.sname = Eric
\exists s \in \text{Student}(s.sname = \text{Eric})
Student(s) \land \exists t \in \text{Buys}(s.sname = \text{Eric} \land t.sid = s.sid)
\exists s \in \text{Student}, t \in \text{Buys}(s.sname = \text{Eric} \land t.sid = s.sid)
\exists s \in \text{Student}(\exists t \in \text{Buys}(s.sname = \text{Eric} \land t.sid = s.sid))
```

 $\exists s \in \text{Student}(\forall t \in \text{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000))$

Free and bound variables of TRC formulas

A tuple variable is *bound* if gets bound by an associated existential or universal *quantifier*. It is *free* otherwise.²

TRC Formula H	free(H)	bound(H)
R(t)	{ t}	{}
t.Aθs.B	{t, s}	[{}
$t.A\theta$ a	{ <i>t</i> }	[]
$F_1 \wedge F_2$	$free(F_1) \cup free(F_2)$	$bound(F_1) \cup bound(F_2)$
$F_1 \vee F_2$	$free(F_1) \cup free(F_2)$	$bound(F_1) \cup bound(F_2)$
$F_1 \rightarrow F_2$	$free(F_1) \cup free(F_2)$	$bound(F_1) \cup bound(F_2)$
$\neg F$	free(F)	bound(F)
∃ <i>t F</i>	$free(F) - \{t\}$	$bound(F) \cup \{t\}$
∀ <i>t F</i>	$free(F) - \{t\}$	$bound(F) \cup \{t\}$
$\exists t_1 \in R_1, \ldots, t_k \in R_k(F)$	$free(F) - \{t_1, \ldots, t_k\}$	$bound(F) \cup \{t_1, \ldots, t_k\}$
$\forall t_1 \in R_1, \ldots, t_k \in R_k(F \to G)$	$free(F \rightarrow G) - \{t_1, \ldots, t_k\}$	$bound(F \rightarrow G) \cup \{t_1, \ldots, t_k\}$
(F)	free(F)	bound(F)

 $F(R_1,...,R_n;t_1,...,t_k)$:

- R_1 through R_0 are relation schemas appearing in formula F; and
- t_1 through t_k are the free variables of formula F.

²For an SQL query, free variables are exclusively those that are introduced in the outermost FROM clause of the query.

Formula and its free and bound variables (Examples)

Formula	free	bound
Student(s)	{s}	{}
$\texttt{Student}(s) \land s.sname = \textbf{Eric}$	{s}	{}
$\exists s(Student(s) \land s.sname = Eric)$	{}	{s}
$\texttt{Student}(s) \land \exists \ t \in \texttt{Buys}(\textbf{Eric} \land t.\textit{sid} = s.\textit{sid})$	{s}	{ <i>t</i> }
$\exists \ s \ \exists \ t (\texttt{Student}(s) \land \texttt{Buys}(t) \land s. \textit{sname} = Eric \land t. \textit{sid} = s. \textit{sid})$	{}	{s, t}
$\exists \ \mathit{s}(\mathtt{Student}(\mathit{s}) \land \exists \ \mathit{t}(\mathtt{Buys}(\mathit{t}) \land \mathit{s.sname} = Eric \land \mathit{t.sid} = \mathit{s.sid}))$	{}	{s, t}
$\exists \; s(\texttt{Student}(s) \land \forall \; t(\texttt{Buys}(t) \land t.sid = s.sid \rightarrow t.bookno \neq \texttt{1000}))$	{}	{s, t}

Fundamental equivalence involving implication and universal quantification

Truth table of $F \rightarrow G$:

F	G	F o G	$\neg F \lor G$	$\neg (F \land \neg G)$
t	t	t	t	t
t	f	f	f	f
f	t	t	t	t
f	f	t	t	t

$$F \to G \qquad \Leftrightarrow \neg F \lor G \\ \Leftrightarrow \neg (F \land \neg G)$$

$$\neg (F \to G) \qquad \Leftrightarrow \neg (\neg (F \land \neg G)) \\ \Leftrightarrow F \land \neg G$$

$$\forall t_1 \in R_1, \dots t_k \in R_k(F \to G) \qquad \Leftrightarrow \neg \exists t_1 \in R_1, \dots t_k \in R_k \neg (F \to G) \\ \Leftrightarrow \neg \exists t_1 \in R_1, \dots t_k \in R_k (F \land \neg G)$$

TRC constraints

A formula F is a *constraint* if $free(F) = \{\}$.

A constraint is a *statement* that is either true or false.

Primary keys and foreign keys are examples of constraints.

TRC constraints (Examples)

There is a student whose name is Eric $\exists s(Student(s) \land s.sname = Eric)$

There is a student whose name is Eric and who bought a book $\exists s(\text{Student}(s) \land s.sname = \text{Eric} \land \exists t(\text{Buys}(t) \land t.sid = s.sid))$

 $= 2.00 \times 4.00 \times 10^{-10} \times 10^{$

There is a student who only bought books with bookno \neq 1000 $\exists s(\texttt{Student}(s) \land \forall t((\texttt{Buys}(t) \land t.sid = s.sid) \rightarrow t.bookno \neq 1000))$

sid is a primary key for the Student relation $\forall s_1 \forall s_2 ((\text{Student}(s_1) \land \text{Student}(s_2) \land s_1.sid = s_2.sid) \rightarrow s_1.sname = s_2.sname)$

TRC constraints in abbreviated form (Examples)

There is a student whose name is Eric $\exists s \in \text{Student}(s.sname = \text{Eric})$

There is a student whose name is Eric and who bought a book $\exists s \in \text{Student}(s.sname = \text{Eric} \land \exists t \in \text{Buys}(t.sid = s.sid))$

There is a student who only bought books with bookno \neq 1000 $\exists s \in \text{Student}(\forall t \in \text{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000))$

sid is a primary key for the Student relation $\forall s_1 \in \text{Student}(s_1.sid = s_2.sid \rightarrow s_1.sname = s_2.sname)$

TRC constraint:

There is a student whose name is Eric

$$\exists s \in \text{Student}(s.sname = \text{Eric})$$

TRC constraint using $\neq \emptyset$ set predicate:

```
\{s.sid, s.sname \mid \mathtt{Student}(s) \land s.sname = \mathtt{Eric})\} 
eq \emptyset
```

TRC constraint:

There is a student whose name is Eric

$$\exists s \in \text{Student}(s.sname = \text{Eric})$$

TRC constraint using $\neq \emptyset$ set predicate:

```
\{1 \mid \text{Student}(s) \land s.sname = \text{Eric}\} \neq \emptyset
```

TRC constraint:

There is a student whose name is Eric and who bought a book.

```
\exists s \in \text{Student}(s.sname = \text{Eric} \land \exists t \in \text{Buys}(t.sid = s.sid))
```

TRC constraint:

There is a student whose name is Eric and who bought a book.

```
\exists s \in \text{Student}, t \in \text{Buys}(s.sname = \text{Eric} \land t.sid = s.sid)
```

There is a student who only bought books with bookno \neq 1000 $\exists s \in \text{Student}(\forall t \in \text{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000)$

Equivalently,

 $\exists s \in \text{Student}(\neg \exists t \in \text{Buys} \neg (t.sid = s.sid \rightarrow t.bookno \neq 1000))$

Equivalently,

 $\exists s \in \text{Student}(\neg \exists t \in \text{Buys}(t.sid = s.sid \land \neg(t.bookno \neq 1000)))$

Equivalently,

 $\exists s \in \text{Student}(\neg \exists t \in \text{Buys}(t.sid = s.sid \land t.bookno = 1000))$

There is a student who only bought books with bookno \neq 1000

```
\exists s \in \text{Student}(\forall t \in \text{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000)
\{1 \mid \text{Student}(s) \land \\ \neg \{1 \mid t \in \text{Buys}(t.sid = s.sid \land t.bookno = 1000)\} \neq \emptyset\} \neq \emptyset
```

sid is a primary key for the Student relation

```
\forall s_1 \in \text{Student}, s_2 \in \text{Student}(s_1.sid = s_2.sid \rightarrow s_1.sname = s_2.sname)
```

Equivalently,

```
\neg \exists s_1 \in \text{Student}, s_2 \in \text{Student}(s_1.sid = s_2.sid \land \neg(s_1.sname = s_2.sname))
```

Equivalently,

```
\neg \exists s_1 \in \text{Student}, s_2 \in \text{Student}(s_1.sid = s_2.sid \land s_1.sname \neq s_2.sname)
```

Each student bought at least two books.

```
\forall s \in \texttt{Student}(\exists t_1 \in \texttt{Buys}, t_2 \in \texttt{Buys}(t_1.sid = s.sid \land t_2.sid = s.sid \land t_1.bookno \neq t_2.bookno))
```

Equivalently,

```
\neg \exists s \in \text{Student} \neg (\exists t_1 \in \text{Buys}, t_2 \in \text{Buys}(t_1.sid = s.sid \land t_2.sid = s.sid \land t_1.bookno \neq t_2.bookno))
```

TRC queries

A *TRC query* over a database schema **R** is a set-abstraction of the form

$$\{(t_{j_1}.A_1,\ldots,t_{j_m}.A_m)|F(R_1,\ldots,R_n;t_1,\ldots,t_k)\}$$

where

- F is a TRC formula in $\mathcal{F}(\mathbf{R})$,
- $\{t_{i_1},\ldots,t_{i_m}\}\subseteq \mathbf{free}(F)$, and
- **③** for $i \in [1, m]$, $A_i \in att(t_{j_i})$.

We will call this the query F.

The schema associated with F is (A_1, \ldots, A_m) , and denote it by att(F).

TRC queries (Examples)

```
 \{(s.id, s.name) \mid \texttt{Student}(s)\}   \{(s.sid) \mid \texttt{Student}(s) \land s.sname = \texttt{Eric}\}   \{() \mid \exists s \in \texttt{Student}(s.sname = \texttt{Eric})\}^3   \{(s.sid, t.bookno) \mid \texttt{Student}(s) \land s.sname = \texttt{Eric} \land \\ \texttt{Buys}(t) \land t.sid = s.sid)\}   \{(s.sid) \mid \texttt{Student}(s) \land \forall t \in \texttt{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000))\}
```

³() denotes the empty tuple, i.e., the tuple without components

Find the sid and name of each student.

```
\{(s.sid, s.sname) \mid Student(s)\}
```

```
select s.sid, s.sname
from Student s;
```

Find the sid of each student whose name is Eric.

```
\{(s.sid) \mid \mathtt{Student}(s) \land s.sname = \mathsf{Eric}\}
```

```
select s.sid
from Student s
where s.sname = 'Eric';
```

Witness if there is a student whose name is Eric.

```
\{() \mid \exists s \in \text{Student}(s.sname = \text{Eric})\}
```

```
select row()
from Student s
where s.sname = 'Eric';
```

Find the (s,b) pairs of each student with sid s and name Eric who bought a book with bookno *b*.

```
\{(s.sid, t.bookno) \mid \texttt{Student}(s) \land s.sname = \textbf{Eric} \land \\ \texttt{Buys}(t) \land t.sid = s.sid\}
```

```
select s.sid, t.bookno
from Student s, Buys t
where s.sname = 'Eric' and
    t.sid = s.sid;
```

TRC queries expressed as SQL queries using IN predicate

Find the sid of each student whose name is Eric and who bought a book.

```
\{(s.\textit{sid}) \mid \texttt{Student}(s) \land s.\textit{sname} = \textbf{Eric} \land \exists \ t \in \texttt{Buys}(t.\textit{sid} = s.\textit{sid})\}
```

Equivalently,

```
\{(\textit{s.sid}) \mid \texttt{Student}(\textit{s}) \land \textit{s.sname} = \textbf{Eric} \land \textit{s.sid} \in \{\textit{t.sid} \mid \texttt{Buys}(\textit{t})\}\}
```

Find the sid of each student who only bought books except that with bookno 1000.

```
\{(s.sid) \mid \texttt{Student}(s) \land \forall \ t \in \texttt{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000))\}
```

Or, equivalently

```
\{(\textit{s.sid}) \mid \texttt{Student}(\textit{s}) \land \neg \exists \textit{t} \in \texttt{Buys}(\textit{t.sid} = \textit{s.sid} \land \textit{t.bookno} = \texttt{1000})\}
```

TRC queries expressed as SQL queries using NOT IN predicate

Find the sid of each student who only bought books except that with bookno 1000.

```
\{(s.sid) \mid \mathtt{Student}(s) \land \forall \ t \in \mathtt{Buys}(t.sid = s.sid \rightarrow t.bookno \neq 1000))\}
```

SQL query:

Or, using NOT IN set membership predicate

Logical Equivalences involving formulas F, G, and H

¬ ¬ F	=		Double Negation
$\neg (F \land G)$	=	$\neg F \lor \neg G$	De Morgan for ∧
$\neg (F \lor G)$	=	$\neg F \land \neg G$	De Morgan for ∨
$F \wedge (G \vee H)$	=	$(F \wedge G) \vee (F \wedge H)$	Distribution of ∧ over ∨
$F \lor (G \land H)$	=	$(F \lor G) \land (F \lor H)$	Distribution of ∨ over ∧
$F \wedge F$	=	F	Idempotence of ∧
$F \vee F$	=	F	Idempotence of ∨
$F \wedge (F \vee G)$	=	F	Absorption Law for ∧
$F \lor (F \land G)$	=	F	Absorption Law for ∨
$F \wedge G$	=	$G \wedge F$	Commutativity of A
$F \vee G$	=	$G \lor F$	Commutativity of ∨
$F \wedge (G \wedge H)$	=	$(F \wedge G) \wedge H$	Associativity of ∧
$F \lor (G \lor H)$	=	$(F \lor G) \lor H$	Associativity of ∨
$F \rightarrow G$	=	$\neg F \lor G$	Conditional Law
$F \rightarrow G$	=	$\neg G \rightarrow \neg F$	Contrapositive Law
$\neg (F \rightarrow G)$	=	$F \wedge \neg G$	Abjunction Law
$F \to (F \land G)$	=	F o G	Absorption Law of Conditional
$(F \land G) \rightarrow H$	=	F o (G o H)	Exportation Law of Conditional
$(F \land G) \rightarrow H$	=	$(F \rightarrow H) \lor (G \rightarrow H)$	Left Distribution of \land over \rightarrow
		$(F \to H) \land (G \to H)$	Left Distribution of \lor over \rightarrow
$F \rightarrow (G \land H)$	=	$(F \rightarrow G) \land (F \rightarrow H)$	Right Distribution of \land over \rightarrow
$F \rightarrow (G \lor H)$	=	$(F \rightarrow G) \lor (F \rightarrow H)$	Right Distribution of \vee over \rightarrow

Logical Equivalences involving formulas F and G and truth formulas true and false

¬ true	=	false	
_ ¬ false	=	true	
$F \wedge \neg F$	=	false	Contradiction Law for ∧
$F \lor \neg F$	≡	true	Tautology Law for ∨
F ∧ true	=	F	Identity Law for ∧
F ∨ false	≡	F	Identity Law for ∨
F ∧ false	=	false	Domination Law for ∧
$F \vee \text{true}$	=	true	Domination Law for ∨
$false \to F$	=	true	Left Domination Law for \rightarrow
true o F	=	F	Right Identity Law for \rightarrow
$F \rightarrow true$	=	true	Right Domination Law for \rightarrow
$F \rightarrow false$	=	$\neg F$	Left Identity Law for \rightarrow
$F \rightarrow F$	=	true	Domination Law for \rightarrow
$((F \to G) \to F) \to F$	=	true	Pierce's Law

Logical equivalences for formulas with quantifiers

$\neg \exists t F(t)$ $\neg \forall t F(t)$	=	$\forall t \neg F(t)$ $\exists t \neg F(t)$	De Morgan for ∃ De Morgan for ∀
$\exists t F(t)$	=	$\neg \forall t \neg F(t)$	20 morganion v
$\forall t F(t)$	=	$\neg \exists t \neg F(t)$	
$\exists t (F(t) \lor G(t))$	=	$(\exists t F(t)) \vee (\exists t G(t))$	Distribution of ∃ over ∨
$\forall t (F(t) \land G(t))$	=	$(\forall t F(t)) \wedge (\forall t G(t))$	Distribution of ∀ over ∧
$\exists t (F \land G(t))$	=	$F \wedge (\exists t G(t))$	if t is not free in F
$\forall t (F \vee G(t))$	=	$F \vee (\forall t G(t))$	if t is not free in F
∃t F	=	F	if t is not free in F
∀t F	=	F	if t is not free in F
$\exists t (F(t) \land G(t))$	=	$\neg \forall t (F(t) \rightarrow \neg G(t))$	
$\forall t (F(t) \rightarrow G(t))$	=	$\neg \exists t (F(t) \land \neg G(t))$	
$\exists t_1 \exists t_2 F(t_1, t_2)$	=	$\exists t_2 \exists t_1 F(t_1, t_2)$	
$\forall t_1 \ \forall t_2 \ F(t_1, t_2)$	=	$\forall t_2 \ \forall t_1 \ F(t_1, t_2)$	