CHAPTER 27

Formal-Relational Query Languages

Practice Exercises

27.1 Let the following relation schemas be given:

$$R = (A, B, C)$$
$$S = (D, E, F)$$

Let relations $r(R)$ and $s(S)$ be given. Give an expression in the tuple relational calculus that is equivalent to each of the following:

a. $\Pi_A(r)$

b. $\sigma_{B=17}(r)$

c. $r \times s$

d. $\Pi_{A,F}(\sigma_{C=D}(r \times s))$

Answer:

a. $\{t \mid \exists\, q \in r\,(q[A] = t[A])\}$

b. $\{t \mid t \in r \wedge t[B] = 17\}$

c. $\{t \mid \exists\, p \in r\, \exists\, q \in s\,(t[A] = p[A] \wedge t[B] = p[B] \wedge t[C] = p[C] \wedge t[D] = q[D] \wedge t[E] = q[E] \wedge t[F] = q[F])\}$

d. $\{t \mid \exists\, p \in r\, \exists\, q \in s\,(t[A] = p[A] \wedge t[F] = q[F] \wedge p[C] = q[D]\}$

27.2 Let $R = (A, B, C)$, and let r_1 and r_2 both be relations on schema R. Give an expression in the domain relational calculus that is equivalent to each of the following:

a. $\Pi_A(r_1)$

b. $\sigma_{B=17}(r_1)$

c. $r_1 \cup r_2$

d. $r_1 \cap r_2$

e. $r_1 - r_2$

f. $\Pi_{A,B}(r_1) \bowtie \Pi_{B,C}(r_2)$

Answer:

a. $\{ < t > \mid \exists\, p,\ q\, (< t, p, q > \in r_1)\}$

b. $\{ < a, b, c > \mid < a, b, c > \in r_1 \wedge b = 17\}$

c. $\{ < a, b, c > \mid < a, b, c > \in r_1 \vee < a, b, c > \in r_2\}$

d. $\{ < a, b, c > \mid < a, b, c > \in r_1 \wedge < a, b, c > \in r_2\}$

e. $\{ < a, b, c > \mid < a, b, c > \in r_1 \wedge < a, b, c > \notin r_2\}$

f. $\{ < a, b, c > \mid \exists\, p,\ q\, (< a, b, p > \in r_1 \wedge < q, b, c > \in r_2)\}$

27.3 Let $R = (A, B)$ and $S = (A, C)$, and let $r(R)$ and $s(S)$ be relations. Write expressions in relational algebra for each of the following queries:

a. $\{ < a > \mid \exists\, b\, (< a, b > \in r \wedge b = 7)\}$

b. $\{ < a, b, c > \mid < a, b > \in r \wedge < a, c > \in s\}$

c. $\{ < a > \mid \exists\, c\, (< a, c > \in s \wedge \exists\, b_1, b_2\, (< a, b_1 > \in r \wedge < c, b_2 > \in r \wedge b_1 > b_2))\}$

Answer:

a. $\Pi_A\, (\sigma_{B=17}(r))$

b. $r \bowtie s$

c. $\Pi_A\, (s \bowtie (\Pi_{r.A}\, (\sigma_{r.b>d.b}(\, r \times \rho_d\, (r)))))$

27.4 Consider the relational database of Figure 27.13 where the primary keys are underlined. Give an expression in tuple relational calculus for each of the following queries:

a. Find all employees who work directly for "Jones."

b. Find all cities of residence of all employees who work directly for "Jones."

c. Find the name of the manager of the manager of "Jones."

employee (*person_name*, *street*, *city*)
works (*person_name*, *company_name*, *salary*)
company (*company_name*, *city*)
manages (*person_name*, *manager_name*)

Figure 27.13 Employee database.

d. Find those employees who earn more than all employees living in the city "Mumbai."

Answer:

a. Query:

$$\{t \mid \exists\, m \in manages\,(t[person\_name] = m[person\_name] \wedge m[manager\_name] = \text{'Jones'})\}$$

b. Query:

$$\{t \mid \exists\, m \in manages\, \exists e \in employee(e[person\_name] = m[person\_name] \wedge m[manager\_name] = \text{'Jones'} \wedge t[city] = e[city])\}$$

c. Query:

$$\{t \mid \exists\, m1 \in manages\, \exists m2 \in manages(m1[manager\_name] = m2[person\_name] \wedge m1[person\_name] = \text{'Jones'} \wedge t[manager\_name] = m2[manager\_name])\}$$

d. Query:

$$\{t \mid \exists\, w1 \in works\, \neg\exists w2 \in works(w1[salary] < w2[salary] \exists e2 \in employee\,(w2[person\_name] = e2[person\_name] \wedge e2[city] = \text{'Mumbai'}))\}$$

27.5 Let $R = (A, B)$ and $S = (A, C)$, and let $r(R)$ and $s(S)$ be relations. Write expressions in Datalog for each of the following queries:

a. $\{ < a > \mid \exists\, b\,(< a, b > \in r \wedge b = 17)\}$

b. $\{ < a, b, c > \mid < a, b > \in r \wedge < a, c > \in s\}$

c. $\{ < a > \mid \exists\, c\,(< a, c > \in s \wedge \exists\, b_1, b_2\,(< a, b_1 > \in r \wedge < c, b_2 > \in r \wedge b_1 > b_2))\}$

Answer:

a. $query(X) :- r(X, 17)$

b. $query(X, Y, Z) :- r(X, Y), s(X, Z)$

c. $query(X) :- s(X, Y), r(X, Z), r(Y, W), Z > W$

27.6 Consider the relational database of Figure 27.13 where the primary keys are underlined. Give an expression in Datalog for each of the following queries:

a. Find all employees who work (directly or indirectly) under the manager "Jones."

b. Find all cities of residence of all employees who work (directly or indirectly) under the manager "Jones."

c. Find all pairs of employees who have a (direct or indirect) manager in common.

d. Find all pairs of employees who have a (direct or indirect) manager in common and are at the same number of levels of supervision below the common manager.

Answer:

a. Query:

$$query(X) :- p(X)$$
$$p(X) :- manages(X, \text{"Jones"})$$
$$p(X) :- manages(X, Y), p(Y)$$

b. Query:

$$query(X, C) :- p(X), employee(X, S, C)$$
$$p(X) :- manages(X, \text{"Jones"})$$
$$p(X) :- manages(X, Y), p(Y)$$

c. Query:

$$query(X, Y) :- p(X, W), p(Y, W)$$
$$p(X, Y) :- manages(X, Y)$$
$$p(X, Y) :- manages(X, Z), p(Z, Y)$$

d. Query:

query(X, Y) :- *p(X, Y)*
p(X, Y) :- *manages(X, Z), manages(Y, Z)*
p(X, Y) :- *manages(X, V), manages(Y, W), p(V, W)*

27.7 Describe how an arbitrary Datalog rule can be expressed as an extended relational-algebra view.

Answer:

A Datalog rule has two parts, the *head* and the *body*. The body is a comma-separated list of *literals*. A *positive literal* has the form $p(t_1, t_2, \ldots, t_n)$ where p is the name of a relation with n attributes, and $t_1, t_2, \ldots, t_n$ are either constants or variables. A *negative literal* has the form $\neg p(t_1, t_2, \ldots, t_n)$ where p has n attributes. In the case of arithmetic literals, p will be an arithmetic operator like $>$, $=$ etc.

We consider only safe rules; see Section 27.4.4 for the definition of safety of Datalog rules. Further, we assume that every variable that occurs in an arithmetic literal also occurs in a positive nonarithmetic literal.

Consider first a rule without any negative literals. To express the rule as an extended relational-algebra view, we write it as a join of all the relations referred to in the (positive) nonarithmetic literals in the body, followed by a selection. The selection condition is a conjunction obtained as follows: If $p_1(X, Y)$, $p_2(Y, Z)$ occur in the body, where p_1 is of the schema (A, B) and p_2 is of the schema (C, D), then $p_1.B = p_2.C$ should belong to the conjunction. The arithmetic literals can then be added to the condition.

As an example, the Datalog query

query(X, Y) :- *works(X, C, S1), works(Y, C, S2), S1 > S2, manages(X, Y)*

becomes the following relational-algebra expression:

$$E_1 = \sigma_{(w1.company\_name = w2.company\_name \wedge w1.salary > w2.salary \wedge manages.person\_name = w1.person\_name \wedge manages.manager\_name = w2.person\_name)} (\rho_{w1}(works) \times \rho_{w2}(works) \times manages)$$

Now suppose the given rule has negative literals. First suppose that there are no constants in the negative literals; recall that all variables in a negative literal must also occur in a positive literal. Let $\neg q(X, Y)$ be the first negative literal, and let it be of the schema (E, F). Let E_i be the relational-algebra expression obtained after all positive and arithmetic literals have been handled. To handle this negative literal, we generate the expression

$$E_j = E_i \bowtie (\Pi_{A_1 A_2}(E_i) - q)$$

where A_1 and A_2 are the attribute names of two columns in E_i which correspond to X and Y respectively.

Now let us consider constants occurring in a negative literal. Consider a negative literal of the form $\neg q(a, b, Y)$ where a and b are constants. Then, in the above expression defining E_j, we replace q with $\sigma_{A_1=a \wedge A_2=b}(q)$.

Proceeding in a similar fashion, the remaining negative literals are processed, finally resulting in an expression E_w.

Finally the desired attributes are projected out of the expression. The attributes in E_w corresponding to the variables in the head of the rule become the projection attributes.

Thus our example rule finally becomes the view:

create view *query* **as**

$$\Pi_{w1.person\_name,\, w2.person\_name}(E_2)$$

If there are multiple rules for the same predicate, the relational-algebra expression defining the view is the union of the expressions corresponding to the individual rules.

The above conversion can be extended to handle rules that satisfy some weaker forms of the safety conditions, and to some restricted cases where the variables in arithmetic predicates do not appear in a positive nonarithmetic literal.