Fall, 2007 CIS 550

Database and Information Systems

Solutions to Homework 1

Due on September 26, 2007

The first two problems concern the Penn Ebay (PBAY) System, which is represented by the following schema:

Sellers(*sellerID*:`int`, *rating*:`char`, *email*:`string`)
Items(*itemID*:`int`, *type*:`string`)
Buyers(*buyerID*:`int`, *email*:`string`, *city*:`string`, *state*:`string`)
Stock(*itemID*:`int`, *sellerID*:`int`, *startBid*:`float`, *quantity*:`int`,*endingTime*:`int`)
Purchase(*itemID*:`int`, *buyerID*:`int`, *sellerID*:`int`, *price*:`float`, *purchaseQuantity*:`int`, *bidTime*:`int`)

Problem 1 [60 points]: Express the following queries in (a) the relational algebra, (b) the tuple relational calculus, and (c) the domain relational calculus:

Note: in problems where wording proved unclear, answers correct with respect to some reasonable interpretation of the problem were accepted.

1. Find the **ID**s of items purchased for price < \$50.

   RA: $\pi_{itemID}(\sigma_{price<50}(Purchase))$
   TRC: $\{Q \mid \exists P \in Purchase(P.price < 50 \wedge P.itemID = Q.itemID)\}$
   DRC: $\{< i > \mid \exists b, s, p, u, m(< i, b, s, p, u, m > \in Purchase \wedge p < 50)\}$

2. Find the **email**s of buyers from PA who buy items with purchaseQuantity > 3.

   RA: $\pi_{email}(\sigma_{state=\text{“PA”}}(Buyers) \bowtie \pi_{buyerID}(\sigma_{purchaseQuantity>3}(Purchase)))$
   TRC: $\{Q \mid \exists B \in Buyers, \exists P \in Purchase\ (B.state = \text{“PA”} \wedge B.buyerID = P.buyerID \wedge P.purchaseQuantity > 3 \wedge Q.email = B.email)\}$
   DRC: $\{< e > \mid \exists b, c, a(< b, e, c, a > \in Buyers \wedge a = \text{“PA”} \wedge \exists i, s, p, u, m(< i, b, s, p, u, m > \in Purchase \wedge u > 3)\}$

3. Find the **ID**s of buyers who purchased items of purchaseQuantity less than 10% of the quantity provided by the same seller the buyer purchased from in the stock.

   RA: $\pi_{buyerID}(\pi_{itemID,sellerID,quantity}(Stock) \bowtie$
   $\sigma_{purchaseQuantity<0.1*quantity}(\pi_{itemID,buyerID,sellerID,purchaseQuantity}(Purchase)))$
   TRC: $\{Q \mid \exists T \in Stock, \exists P \in Purchase\ (T.itemID = P.itemID \wedge T.sellerID = P.sellerID \wedge P.purchaseQuantity < 0.1 * T.quantity \wedge Q.buyerID = P.buyerID)\}$
   DRC: $\{< b > \mid \exists i, s, q, u(\exists d, n(< i, s, d, q, n > \in Stock) \wedge \exists p, m(< i, b, s, p, u, m > \in Purchase) \wedge u < 0.1 * q)\}$

4. Find the **ID**s of buyers who purchased items with type “furniture” for over 10% of the startBid price of the items they bought.

   RA: $\pi_{buyerID}(\pi_{itemID}(\sigma_{type=\text{“furniture”}}(Items)) \bowtie$
   $\pi_{itemID,sellerID,startBid}(Stock) \bowtie \sigma_{price>\ 1.1*startBid}(\pi_{itemID,buyerID,sellerID,price}(Purchase)))$
   TRC: $\{Q \mid \exists I \in Items, \exists T \in Stock, \exists P \in Purchase\ (I.itemID = T.itemID \wedge T.itemID = P.itemID \wedge T.sellerID = P.sellerID \wedge P.price > 1.1 * T.startBid \wedge I.type = \text{“furniture”} \wedge$

$Q.buyerID = P.buyerID)\}$
DRC: $\{< b > \mid \exists i(\exists t(< i, t > \in Items \wedge t = \text{“furniture”}) \wedge \exists s, p, d(\exists q, n(< i, s, d, q, n > \in Stock) \wedge \exists u, m(< i, b, s, p, u, m > \in Purchase) \wedge (p > 1.1 * d)))\}$

5. Find the **ID**s of buyers who either always make purchases with purchaseQuantity < 5 or haven't made any purchases.

   RA: $\pi_{buyerID}(Buyers) - \pi_{buyerID}(\sigma_{purchaseQuantity \geq 5}(Purchase))$
   TRC: $\{Q \mid \exists B \in Buyers\ (\forall P \in Purchase\ (P.purchaseQuantity < 5 \vee P.buyerID \neq B.buyerID) \wedge Q.buyerID = B.buyerID)\}$
   DRC: $\{< b > \mid \exists e, c, a(< b, e, c, a > \in Buyers \wedge \neg \exists i, s, p, u, m((< i, b, s, p, u, m > \in Purchase) \wedge (u \geq 5))\}$
   (You can either have:
   DRC: $\{< b > \mid \exists e, c, a(< b, e, c, a > \in Buyers \wedge \forall i, s, p, u, m(\neg(< i, b, s, p, u, m > \in Purchase) \vee (u < 5))\}$, but remember that is is an unsafe plan.)

6. Find the types of items stocked by ≥ 2 sellers or bought by ≥ 2 buyers.

   RA: $\pi_{type}(\sigma_{s1 \neq s2}(\rho_{sellerID \rightarrow s1}(\pi_{sellerID,type}(Items \bowtie Stock)) \bowtie \rho_{sellerID \rightarrow s2}(\pi_{sellerID,type}(Items \bowtie Stock))))$
   $\cup$
   $\pi_{type}(\sigma_{b1 \neq b2}(\rho_{buyerID \rightarrow b1}(\pi_{buyerID,type}(Items \bowtie Purchase)) \bowtie \rho_{buyerID \rightarrow b2}(\pi_{buyerID,type}(Items \bowtie Purchase))))$
   TRC: $\{Q \mid (\exists P_1, P_2 \in Purchases, \exists I_1, I_2 \in Items\ (P_1.itemID = I_1.itemID \wedge P_2.itemID = I_2.itemID \wedge I_1.type = I_2.type \wedge P_1.buyerID \neq P_2.buyerID \wedge I_1.type = Q.type)) \vee (\exists S_1, S_2 \in Stocks, \exists I_1, I_2 \in Items\ (S_1.itemID = I_1.itemID \wedge S_2.itemID = I_2.itemID \wedge I_1.type = I_2.type \wedge S_1.sellerID \neq S_2.sellerID \wedge I_1.type = Q.type))\}$
   DRC: $\{< t > \mid \exists i_1, s_1, d_1, q_1, n_1, i_2, s_2, d_2, q_2, n_2(< i_1, t > \in Items \wedge < i_1, s_1, d_1, q_1, n_1 > \in Stock \wedge < i_2, t > \in Items \wedge < i_2, s_2, d_2, q_2, n_2 > \in Stock\ \wedge s_1 \neq s_2) \vee \exists i_3, b_3, s_3, p_3, u_3, m_3, i_4, b_4, s_4, p_4, u_4, m_4(< i_3, t > \in Items \wedge < i_3, b_3, s_3, p_3, u_3, m_3 > \in Purchase \wedge < i_4, t > \in Items \wedge < i_4, b_4, s_4, p_4, u_4, m_4 > \in Purchase\ \wedge b_3 \neq b_4)\}$

Problem 2 [30 points]: State in English what the following queries compute:

1. $\{Q \mid \exists P \in Purchase, \exists S \in Sellers\ (S.rating = \text{“A”} \wedge P.sellerID = S.sellerID \wedge P.purchaseQuantity = 2 \wedge Q.buyerID = P.buyerID)\}$
   IDs of buyers who have bought 2 of the same items from a seller with rating “A”.

2. $\{< e > \mid \exists i, s(\exists r(< s, r, e > \in Sellers) \wedge \exists d, q, n(< i, s, d, q, n > \in Stock \wedge (d < 20) \wedge (q = 5)) \wedge \exists b, p, u, m(< i, b, s, p, u, m > \in Purchase \wedge (p > 50)))\}$
   Emails of sellers who have 5 of the same stock items, all with start bid price < 20, and who have sold at least 1 item with price > 50.

3. $\pi_{email}(\sigma_{city=\text{“}Philadelphia\text{”}}(Buyers) \bowtie \pi_{buyid}(\sigma_{price < 2 * startBid}(\sigma_{type=\text{“}book\text{”} \wedge purchaseQuantity=2}(Items \bowtie Purchase) \bowtie Stock)))$
   Emails of the buyers living in Philadelphia who have bought 2 same books with price less than twice of the start bid price.

4. $\pi_{rating}(\pi_{s1}(\sigma_{i1 \neq i2 \wedge s1 = s2}(\rho_{itemID \rightarrow i1, sellerID \rightarrow s1}(Stock) \bowtie \rho_{itemID \rightarrow i2, sellerID \rightarrow s2}(\sigma_{quantity \geq 3}(Stock)))) \bowtie_{s1 = sellerID} Sellers)$
   Ratings of sellers with 2 different items in stock where the seller stocks at least 3 of one of those items.

Problem 3 [10 points]: Explain how Codd's points of access path dependence and indexing dependence relate to today's Java objects. (Assume the goal is to return all instances of a particular member variable of a particular object, which might be linked to by other objects.)

Java has all of the same "shortcomings," i.e., access path dependencies, as the languages of Codd's time. For instance, a programmer must know: (1) data ordering for files serialized to disk; (2) what indices are available, and what they point to (and in fact this requires a special library like BerkeleyDB); (3) how references between files are represented (one can serialize object references in a single file but not across files).