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

2. Find the emails of buyers from PA who buy items with purchaseQuantity > 3.
   RA: \( \pi_{\text{email}}(\sigma_{\text{state} = \text{PA}}(\text{Buyers}) \bowtie \pi_{\text{buyerID}}(\sigma_{\text{purchaseQuantity}>3}(\text{Purchase})) \)
   TRC: \{Q \mid \exists B \in \text{Buyers}, \exists P \in \text{Purchase} (B.state = \text{PA} \land B.buyerID = P.buyerID \land P.purchaseQuantity > 3 \land Q.email = B.email)\}
   DRC: \{< e > \mid \exists b, c, a(< b, c, a \in \text{Buyers} \land a = \text{PA} \land \exists i, s, p, u, m(< i, b, s, p, u, m \in \text{Purchase} \land u > 3))\}

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

4. Find the IDs of buyers who purchased items with type “furniture” for over 10% of the startBid price of the items they bought.
   RA: \( \pi_{\text{buyerID}}(\pi_{\text{itemID}}(\sigma_{\text{type} = \text{furniture}}(\text{Items})) \bowtie \pi_{\text{itemID}, \text{sellerID}, \text{startBid}}(\text{Stock}) \bowtie \sigma_{\text{price}>1.1*\text{startBid}}(\pi_{\text{itemID}, \text{buyerID}, \text{sellerID}, \text{price}}(\text{Purchase}))) \)
   TRC: \{Q \mid \exists I \in \text{Items}, \exists T \in \text{Stock}, \exists P \in \text{Purchase} (I.itemID = T.itemID \land T.itemID = P.itemID \land T.sellerID = P.sellerID \land P.price > 1.1 * T.startBid \land I.type = \text{“furniture”} \land
Problem 2 [30 points]: State in English what the following queries compute:

1. \( \{ P \mid \exists P \in Purchase \land P.sellerID = S.sellerID \land P.purchaseQuantity = 2 \land Q.buyerID = Q.buyerID \} \)

IDs of buyers who have bought 2 of the same items from a seller with rating “A”.

2. \( \{ < s > \mid \exists s, r, e \in Sellers \land \exists d, q, n \in Stock \land (d < 20) \land (q = 5) \land \exists b, s, p, u, m \in Purchase \land (p > 50) \} \)

Emails of sellers who have 5 or haven’t made any purchases.

3. \( \pi_{email}(\sigma_{city="Philadelphia"}(Buyers)) \bowtie \pi_{buyerID}(\sigma_{price<2\bowt startBid}(\sigma_{type=\"book\"}(Items \bowt Purchase) \bowt Stock))) \)

Emails of the buyers living in Philadelphia who have bought 2 books with price less than twice of the start bid price.

4. \( \pi_{rating}(\sigma_{s1 \neq s2}(\sigma_{itemID \neq i1, sellerID \neq s1}(Stock) \bowtie \rho_{itemID \neq i2, sellerID \neq s2}(\sigma_{quantity \geq 3}(Stock)))) \bowtie b1 \bowt Seller ID Sellers) \)

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).