Fall, 2005 CIS 550

Database and Information Systems

Solutions to Homework 1

Due on September 27, 2005

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`, *address*:`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 with **startBid** price $< \$10$

   RA: $\pi_{itemID}(\sigma_{startBid<10}(Stock))$
   TRC: $\{P \mid \exists S \in Stock(S.startBid < 10 \wedge P.itemID = S.itemID)\}$
   DRC: $\{< i > \mid \exists s, b, q, e(< i, s, b, q, e > \in Stock \wedge b < 10)\}$

2. Find the **email**s of sellers with rating 'A' who have items in stock with **startBid** price $< \$10$.

   RA: $\pi_{email}(\sigma_{rating='A'}(Sellers) \bowtie \pi_{sellerID}(\sigma_{startBid<10}(Stock)))$
   TRC: $\{P \mid \exists S \in Sellers, \exists T \in Stock\ (S.rating = \text{“A”} \wedge S.sellerID = T.sellerID \wedge T.startBid < 10 \wedge P.email = S.email)\}$
   DRC: $\{< m > \mid \exists s, r(< s, r, m > \in Sellers \wedge (r = \text{“A”} \wedge (\exists i, b, q, e(< i, s, b, q, e > \in Stock \wedge b < 10))))\}$

3. Find the **ID**s of buyers who purchased items for over 10% beyond the **startBid** price of the items they bought.

   RA: $\pi_{buyerID}(\pi_{itemID,startBid}(Stock) \bowtie \sigma_{price>1.1*startBid}(\pi_{itemID,buyid,price}(Purchase)))$
   TRC: $\{P \mid \exists T \in Stock, \exists U \in Purchase\ (T.itemID = U.itemID \wedge U.price > 1.1 * T.startBid \wedge U.buyerID = P.buyerID)\}$
   DRC: $\{< b > \mid \exists i, p, d(\exists s, q(< i, s, d, q, e > \in Stock) \wedge \exists s, u, m(< i, b, s, p, u, m > \in Purchase)) \wedge p > 1.1 * d\}$

4. Find the **ID**s of buyers who purchase items with type “book” and with a **bidTime** within 5 minutes of the **endingTime**. (clarify: endingTime - bidTime will return minutes)

   RA: $\pi_{buyerID}(\pi_{itemID}(\sigma_{type=\text{“book”}}(Items)) \bowtie$
   $\pi_{sellerID,itemID,endingTime}(Stock) \bowtie \sigma_{(endingTime-bidTime)<5}(\pi_{sellerID,itemID,buyerID,bidTime}(Purchase)))$
   TRC: $\{P \mid \exists I \in Items, \exists T \in Stock, \exists U \in Purchase\ (I.itemID = T.itemID \wedge T.itemID = U.itemID \wedge T.sellerID = U.sellerID \wedge (U.endingTime - T.bidTime < 5) \wedge I.type = \text{“book”} \wedge U.buyerID = P.buyerID\}$
   DRC: $\{< b > \mid \exists t(< i, t > \in Items \wedge t = \text{“book”}) \wedge (\exists d, e(< i, s, d, q, e > \in Stock) \wedge \exists p, u(< i, s, b, p, u, m > \in Purchase \wedge (e - m) < 5))\}$

5. Find the **ID**s of buyers who either always make purchases with price $<$ \$10 or haven't make any purchase.
   RA: $\pi_{buyerID}(Buyers) - \pi_{buyerID}(\sigma_{price \geq 10}(Purchase))$
   TRC: $\{Q \mid \exists B \in Buyers\ (\forall P \in Purchases\ (P.price < 10 \vee P.buyerID \neq B.buyerID) \wedge B.buyerID = Q.buyerID)\}$
   DRC: $\{< b > | \exists l, a (< b, l, a > \in Buyers \wedge \forall i, s, p, u, m (< i, s, b, p, u, m > \notin Purchase \vee p < 10))\}$

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: $\{< t > | \exists i_1, s_1, d_1, q_1, e_1, i_2, s_2, d_2, q_2, e_2 (< i_1, t > \in Items \wedge < i_1, s_1, d_1, q_1, e_1 > \in Stock \wedge < i_2, t > \in Items \wedge < i_2, s_2, d_2, q_2, e_2 > \in Stock \wedge s_1 \neq s_2) \vee \exists i_3, s_3, b_3, p_3, u_3, m_3, i_4, s_4, b_4, p_4, u_4, m_4 (< i_3, t > \in Items \wedge < i_3, s_3, b_3, p_3, u_3, m_3 > \in Purchase \wedge < i_4, t > \in Items \wedge < i_4, s_4, b_4, p_4, u_4, m_4 > \in Purchase \wedge b_3 \neq b_4)\}$

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

1. $\pi_{buyerID}(\sigma_{purchaseQuantity=2}(Purchase) \bowtie \pi_{sellerID}(\sigma_{rating='A'}(Sellers)))$
   IDs of buyers who have bought 2 of the same items from a seller with rating "A".

2. $\pi_{email}((\pi_{sellerID}(\sigma_{quantity=4 \wedge startbid<10}(Stock)) \bowtie \pi_{sellerID}(\sigma_{price>20}(Purchase))) \bowtie Sellers)$
   Emails of sellers who have 4 of the same stock items, all with start bid price < 10, and who have sold at least 1 item with price > 20.

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

4. $\pi_{rating}(\pi_{s1}(\sigma_{i1 \neq i2 \bowtie s1=s2}(\rho_{itemID \rightarrow i1, sellerID \rightarrow s1}(Stock) \bowtie \rho_{itemID \rightarrow i2, sellerID \rightarrow s2}(\sigma_{quantity \geq 3}(Stock)))) \bowtie Sellers)$
   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 the principles of data independence, as argued for by Codd, and their significance.

The cornerstone of Codd's ideas of data independence is the separation of the data to be processed from the means of processing it, in order to enable greater flexibility and extensibility. In essence, the logical data model is a representation of the data independent of its order, indices, and access methods or paths. These concepts are the ones upon which today's database field is built.