Fall, 2004 CIS 550

Database and Information Systems

Solutions to Homework 1

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

Sellers(*sellerID*:`int`, *rating*:`char[2]`, *email*:`string`)
Items(*itemID*:`int`, *typeID*:`int`)
Stock(*itemID*:`int`, *startBid*:`float`, *qty*:`int`)
SoldBy(*itemID*:`int`, *sellerID*:`int`)
Description(*itemID*:`int`, *desc*:`string`)
Purchases(*purchaseID*:`int`, *itemID*:`int`, *custID*:`int`, *soldFor*:`float`, *quant*:`int`)
Customers(*custID*:`int`, *address*:`string`)

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 sellers with rating "A+";

   RA: $\pi_{sellerID}(\sigma_{rating='A+'}(Sellers))$
   DRC: $\{\langle s\rangle \mid \exists r, e(\langle s, r, e\rangle \in Sellers \wedge r = \text{'A+'})\}$

2. Find the IDs of sellers with the IDs of items they sell;

   RA: $SoldBy$
   DRC: $\{\langle s, i\rangle \mid \langle s, i\rangle \in SoldBy\}$

3. Find the IDs of customers who bought at least 2 of the same item, or who bought the entire stock of an item;

   RA: $\pi_{custID}(\sigma_{qty\geq 2}(Purchases)) \cup \pi_{custID}(\sigma_{qty=1}(Stock \bowtie Purchases))$
   DRC: $\{\langle s\rangle \mid \exists p, i, c, q(\langle p, i, c, s, q\rangle \in Purchases \wedge (q \geq 2 \vee \exists b, t(\langle i, b, t\rangle \in Stock \wedge t = 1)))\}$

4. Find the IDs of customers who paid no more than $100 for any item;

   RA: $\pi_{custID}(Customers) - \pi_{custID}(\sigma_{soldFor>100}(Purchases))$
   DRC: $\{\langle c\rangle \mid \exists a(\langle c, a\rangle \in Customers \wedge \forall p, i, s, q(\langle p, i, c, s, q\rangle \notin Purchases \vee s \leq 100))\}$

5. Find the IDs of item types bought by at least 3 customers;

   RA: $\pi_{typeID}(\sigma_{c_1\neq c_2 \wedge c_2 \neq c_3}(\rho_{c_1/custID}(\pi_{typeID,custID}(Purchases \bowtie Items)) \bowtie \rho_{c_2/custID}(\pi_{typeID,custID}(Purchases \bowtie Items)) \bowtie \rho_{c_3/custID}(\pi_{typeID,custID}(Purchases \bowtie Items))$
   DRC: $\{\langle t\rangle \mid \exists p_1, p_2, p_3, i_1, i_2, i_3, c_1, c_2, c_3, s_1, s_2, s_3, q_1, q_2, q_3(\langle p_1, i_1, c_1, s_1, q_1\rangle \in Purchases \wedge \langle p_2, i_2, c_2, s_2, q_2\rangle \in Purchases \wedge \langle p_3, i_3, c_3, s_3, q_3\rangle \in Purchases \wedge \langle i_1, t\rangle \in Items \wedge \langle i_2, t\rangle \in Items \wedge \langle i_3, t\rangle \in Items \wedge c_1 \neq c_2 \wedge c_2 \neq c_3)\}$

6. Find the IDs of sellers who sell an item (type) for less than at least one other seller.

   RA: $\pi_{sellerID}(\sigma_{s_1\neq s_2 \wedge b_1<b_2}(\rho_{s_1/sellerID,b_1/startBid}(\pi_{typeID,sellerID,startBid}(Items \bowtie Stock \bowtie SoldBy)) \bowtie \rho_{s_2/sellerID,b_2/startBid}(\pi_{typeID,sellerID,startBid}(Items \bowtie Stock \bowtie SoldBy))))$
   DRC: $\{\langle s_1\rangle \mid \exists i_1, b_1, q_1, s_2, i_2, b_2, q_2, t(\langle s_1, i_1\rangle \in SoldBy \wedge \langle s_2, i_2\rangle \in SoldBy \wedge \langle i_1, t\rangle \in Items \wedge \langle i_2, t\rangle \in Items \wedge \langle i_1, b_1, q_1\rangle \in Stock \wedge \langle i_2, b_2, q_2\rangle \in Stock \wedge s_1 \neq s_2 \wedge b_1 < b_2)\}$

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

1. $\pi_{desc}(\pi_{itemID}(\sigma_{startBid>1000}(Stock) \bowtie \sigma_{soldFor<2000}(Purchases)) \bowtie Description)$

   Descriptions of items whose starting bid was greater than \$1000 and which sold for less than \$2000.

2. $\pi_{desc}(\sigma_{soldFor<2000}(\sigma_{startBid>1000}(Stock \bowtie Purchases \bowtie Description)))$

   Same as previous.

3. $\pi_{email}((\pi_{sellerID}(\sigma_{startBid<10}(Stock) \bowtie SoldBy) \cap$
   $\pi_{sellerID}(\sigma_{startBid>1000}(Stock) \bowtie SoldBy)) \bowtie Sellers)$

   Email addresses of sellers selling both an item with start bid less than \$10 and an item with start bid more than \$1000.

4. $\pi_{email}((\pi_{sellerID}(\sigma_{startBid<10}(Stock) \bowtie SoldBy) \cup$
   $\pi_{sellerID}(\sigma_{startBid>1000}(Stock) \bowtie SoldBy)) \bowtie Sellers)$

   Email addresses of sellers selling either an item with start bid less than \$10 or an item with start bid more than \$1000.

5. $\pi_{typeID}((\pi_{sellerID}(\sigma_{startBid<10}(\sigma_{qty>100}(Stock)) \bowtie SoldBy) \cap$
   $\pi_{sellerID}(\sigma_{startBid>1000}(\sigma_{qty<3}(Stock)) \bowtie SoldBy)) \bowtie SoldBy \bowtie Items)$

   Item type IDs of all items sold by sellers selling some item with start bid less than \$10 and quantity greater than 100 and some item with start bid more than \$1000 and quantity less than 3.

Problem 3 [10 points]: What is a data model, and why was it an important innovation?

A *data model* is an abstract representation of the data components and relationships — a collection of high-level description constructs that hide many low-level storage details. This has many benefits. Decoupling an application from the details of the storage implementation allows an administrator to modify the storage implementation without breaking the application. By accessing the data via queries in a declarative query language, the DBMS is free to choose a most efficient execution plan. Moreover, the data model and declarative query language allow the user to think about the data at a higher level.