

Database and Information Systems

Homework 2

Due on October 8, 2007

For this assignment, you will want to test your queries by running them on a real database. So please begin by signing up for an Oracle account (<http://www.seas.upenn.edu/ora>), accessible from eniac. (For those who do not have eniac accounts, please email the instructor.) Then read over the Oracle setup instructions from the course web page (<http://www.seas.upenn.edu/~zives/cis550/oracle-faq.html>) and modify your eniac .profile file as directed. Also read over the Oracle guide referenced from the course web pages (http://www.cs.wisc.edu/~dbbook/openAccess/thirdEdition/Oracle/user_guide/oracle_guide.html). Finally, download **hw2.sql** to your eniac account, launch Oracle (using the command **sql**), and then **start hw2** to create some sample tables for Problems 1 and 3. These will only be sparsely populated to test your solutions, you may need to INSERT more VALUES into the tables.

Note 1: The SQL string datatype is **VARCHAR(*length*)**, and you'll need to choose an appropriate length.

Note 2: The key fields are underlined in schema. Foreign keys are indicated by naming. (In other words, if x is the key of relation X , then each appearance of x outside of X is a foreign key referencing X).

Problem 1 [30 points]: Consider the following schema based on the TPC-H benchmark (which you'll hear more about later on in the course):

Parts(partID: int, name: string, mfg: string, brand: string, type: string, retailprice: float)

Suppliers(suppID: int, name: string, address: string, nationID: int, phone: string, acctbal: float)

PartSupp(partID: int, suppID: int, availqty: int, supplycost: float)

Nation(nationID: int, name: string, regionID: int)

Region(regionID: int, name: string)

Write the following queries in SQL:

1. Find the IDs of parts available from a single supplier with quantity > 500.
2. Find the IDs of parts available with quantity > 500 considering all suppliers.
3. Find total number of suppliers in each region.

4. Find the IDs of parts which are supplied by suppliers from different nations.
5. $\{\langle n \rangle \mid \exists s, a, t, h, b, p, e, m, r, y, i, v, c (\langle s, n, a, t, h, b \rangle \in \text{supplier} \wedge \langle p, s, v, c \rangle \in \text{partsupp} \wedge \langle p, e, m, r, y, i \rangle \in \text{parts} \wedge y = 'generic' \wedge v > 500)\}$

Problem 2[30 points]: Consider the following Inverted Index schema, useful for finding words within text documents with URLs:

Word(wordid: int, wordname: string)
 DocumentURL(docid: int, url: string)
 Occurs(wordid: int, docid: int)

Write the SQL DDL statements to create these relations, including all primary and foreign key integrity constraints.

Problem 3[40 points]: Use the schema from Homework 1's PBAY system:

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)
 Purchases(itemID: int, buyerID: int, sellerID: int, price: float, purchaseQuantity: int, bidTime: int)

Write the following queries in SQL (from Problem 1 and Problem 2 of Homework 1):

1. Find the **IDs** of items purchased for price < \$50.
2. Find the **emails** of buyers from PA who buy items with purchaseQuantity > 3.
3. Find the **IDs** of buyers who purchased items from of purchaseQuantity less than 10% of the quantity provided by the same seller the buyer purchase from in the stock.
4. Find the **IDs** of buyers who purchased items with type "furniture" for over 10% of the startBid price of the items they bought.
5. Find the **IDs** of buyers who either always make purchases with purchaseQuantity < 5 or haven't made any purchases.
6. Find the types of items stocked by ≥ 2 sellers or bought by ≥ 2 buyers.
7. $\{Q \mid \exists P \in \text{Purchase}, \exists S \in \text{Sellers} (S.\text{rating} = \text{"A"} \wedge P.\text{sellerID} = S.\text{sellerID} \wedge Q.\text{buyerID} = P.\text{buyerID} \wedge P.\text{purchaseQuantity} = 2)\}$
8. $\{\langle e \rangle \mid \exists i, s (\exists r (\langle s, r, e \rangle \in \text{Sellers}) \wedge \exists d, q, n (\langle i, s, d, q, n \rangle \in \text{Stock} \wedge (d < 20) \wedge (q = 5)) \wedge \exists b, p, u, m (\langle i, b, s, p, u, m \rangle \in \text{Purchase} \wedge (p > 50)))\}$

9. $\pi_{email}(\sigma_{city="Philadelphia"}(Buyers) \bowtie \pi_{buyid}(\sigma_{price < 2 * startBid}(\sigma_{type="book"} \wedge purchaseQuantity=2 (Items \bowtie Purchase) \bowtie Stock)))$
10. $\pi_{rating}(\pi_{s1}(\sigma_{i1 \neq i2 \wedge s1=s2}(\rho_{itemID \rightarrow i1, sellerID \rightarrow s1}(Stock) \bowtie \rho_{itemID \rightarrow i2, sellerID \rightarrow s2}(\sigma_{quantity \geq 3}(Stock)))) \bowtie Sellers)$