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 | \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} \land \langle p, s, v, c \rangle \in \text{partsupp} \land \langle p, m, r, y, i \rangle \in \text{parts} \land y = 'generic' \land 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 | \exists P \in \text{Purchase}, \exists S \in \text{Sellers} (S.rating = "A" \land P.sellerID = S.sellerID \land Q.buyerID = P.buyerID \land Q.purchaseQuantity = 2)}
8. \{< e > | \exists i, s(\exists r(\langle s, r, e \rangle \in \text{Sellers}) \land \exists q, n(\langle i, s, d, q, n \rangle \in \text{Stock} \land (d < 20) \land (q = 5)) \land \exists b, p, u, m(\langle i, b, s, p, u, m \rangle \in \text{Purchase} \land (p > 50)))\}
9. \( \pi_{email}(\sigma_{\text{city}=\text{Philadelphia}}(Buyers) \bowtie \pi_{\text{buyid}}(\sigma_{\text{price} < 2 * \text{startBid}}(\sigma_{\text{type} = \text{book}} \land \text{purchaseQuantity} = 2)(Items \bowtie \text{Purchase}) \bowtie \text{Stock}))) \)

10. \( \pi_{\text{rating}}(\pi_{s1}(\sigma_{i1 \neq i2 \land s1 = s2}(\rho_{\text{itemID} \rightarrow i1, \text{sellerID} \rightarrow s1}(Stock) \bowtie \rho_{\text{itemID} \rightarrow i2, \text{sellerID} \rightarrow s2}(\sigma_{\text{quantity} \geq 3}(Stock)))) \bowtie \text{Sellers})) \)