

## Database and Information Systems

### Homework 2

**September 28, 2004 - Due October 5 at 1:30 PM**

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 .cshrc 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](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.

Reminder: the SQL string datatype is `VARCHAR(length)`, and you'll need to choose an appropriate length.

**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):

Part(PartID: int, Name: string, Mfgr: string, Brand: string, Type: string, RetailPrice: float)  
 Supplier(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)

The key fields are underlined. 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  $X$ .)

Write the following queries in SQL:

1. Find the IDs of parts manufactured by Dupont.
2. Find the total quantity of parts available from American suppliers (nation name 'USA').
3. Find the total number of nations.
4.  $\{\langle n, r \rangle \mid \exists m, s \text{ s.t. } \langle n, m, r \rangle \in \text{Nation} \wedge \langle r, s \rangle \in \text{Region} \wedge s = \text{"Asia"}\}$
5. Find the IDs of suppliers whose total supply cost (i.e., sum total of all parts' supply costs times their available quantities) is the highest of all suppliers in their region.

**Problem 2 [30 points]:** Consider the following relations:

```

Student(sid: int, name: string, email: string)
  Takes(sid: int, expgrade: char, cid: int)
      Course(cid: int, sem: string)
Professor(fid: int, name: string, email: string)
  Teaches(fid: int, cid: int)
CourseOffering(cid: int, cno: string, sub: string)

```

The key fields are underlined. Foreign keys are indicated by naming. Write the SQL DDL statements required to create these relations, including appropriate versions of all primary and foreign key integrity constraints.

**Problem 3 [40 points]** Recall the schema for the PBAY system:

```

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, descr: string)
Purchases(purchaseID: int, itemID: int, custID: int, soldFor: float, quant: int)
  Customers(custID: int, address: string)

```

The key fields are underlined. Foreign keys are indicated by naming. ItemID is the vendor-specific code for an item, and typeID is its global identifier (e.g. ISBN).

Translate the following queries into SQL:

1. Find the descriptions of all items that are sold by some seller
2. Find the IDs of the sellers of items stocked in quantity  $\geq 3$
3. Find the IDs of item types sold only by sellers whose email address is either `zives@cis.upenn.edu` or `tjgreen@cis.upenn.edu`
4. For every seller of an item whose description contains the substring “hits”, list the ID of the most expensive item or items (determined by start bid) the seller stocks
5. Find the addresses of buyers who have bought at least \$10.00 worth of items. (Note that `Purchases(..., 10.0, 3)` means \$10.00 paid for each of the three items, i.e., \$30.00 total.)
6. Find the IDs of sellers who are asking a higher start bid for some item than is average for that item type
7.  $\sigma_{sellerID=10}(Sellers \bowtie Stock \bowtie Items)$