

Name: _____

Please note that much of the programming portion of this assignment is devoted to getting familiar with the class accounts and submission procedure; thus, it is much simpler than the programming assignments that will follow. You should submit

1. a hard-copy of this assignment with your answers filled in, and
2. an electronic package that contains the source files for your work on the programming questions, by following the submission procedure described below and in class.

You are welcome to use any inanimate resources (e.g., books, Web sites, publicly available code) to help you with your work. However, *all such help must be clearly noted* in your submissions. Further, no matter what you use, *you must be able to explain, in detail, how it works*. You may be called upon to explain your homework in person. Refer to the class policy for details, and ask for clarifications if you are unsure if something is allowed.

COS 480 Students earn points as indicated for each question below. The non-★ questions sum to 100, the full score. Any points earned on ★ questions, potentially another 50 points, are simply added to the score. If the resulting score is greater than 100, the remaining points are extra credit. A ★ mark on a question in this homework does *not* mean a similar question is not applicable to COS 480 students in future homeworks or exams. Therefore, COS 480 students are encouraged to answer the ★ questions for extra credit on this homework as well.

COS 580 Students must answer all questions, including ★ questions, to yield a raw score (maximum 150) that will be scaled by 2/3 to obtain the score (maximum 100) on the homework. COS 580 students who wish to work for extra credit should contact me for a suitable assignment.

1. (1 pt.) Write your name in the space provided above.
2. (1 pts.) Read the material on the class Web site. Sign your name here to indicate that you have read this material: _____
3. (1 pts.) Change the passwords on your Gandalf (Unix) and PostgreSQL accounts. Fill in the following information. (Change your passwords and note *yes* in the last column. Do *not* write your old or new passwords here.)

Account	User Name	Password Changed?
Gandalf	_____	_____
PostgreSQL	_____	_____

Some questions use the food and nutrition database instance depicted below. (It is based on a real example; see <http://ndb.eip10.org/>.) The three title rows of each table indicate attribute names, their data types, and their abbreviated names for use in relational algebra expressions. The abbreviated name for each table is listed next to its name. The **Foods** table records the names and food groups of several common food items, along with the percentage, by weight, of the inedible portions (refuse). The **Nutrients** table records the standardized name (tagname), common name, and units for several nutrients. The **Compositions** table records the amounts of nutrients in 100 g of various foods.

Foods F		
food	fgroup	refuse
varchar(50)	varchar(25)	float
F	G	R
Nuts, butternuts, dried	Nut and Seed Products	73
Bamboo shoots, raw	Vegetables and Vegetable Products	71
Nuts, pilinuts-canarytree, dried	Nut and Seed Products	81
Durian, raw or frozen	Fruits and Fruit Juices	68

Nutrients N		
tagname	description	units
T	D	U
varchar(10)	varchar(25)	varchar(10)
WATER	Water	g
ENERC_KCAL	Energy	kcal
PROCNT	Protein	g
FAT	Total lipid (fat)	g
ASH	Ash	g
CHOCDF	Carbohydrate, by difference	g

Compositions C		
food	ntag	amt
varchar(50)	varchar(10)	float
F	N	A
Durian, raw or frozen	WATER	64.990
Durian, raw or frozen	ENERC_KCAL	147.000
Durian, raw or frozen	PROCNT	1.470
Durian, raw or frozen	FAT	5.330
Durian, raw or frozen	CHOCDF	27.090
Nuts, butternuts, dried	WATER	3.340
Nuts, butternuts, dried	ENERC_KCAL	612.000
Nuts, butternuts, dried	PROCNT	24.900
Nuts, butternuts, dried	FAT	56.980
Nuts, butternuts, dried	ASH	2.730
Nuts, butternuts, dried	CHOCDF	12.050

Figure 1: A food and nutrition database.

4. (1 pt.) Package and submit your solutions to the programming questions. (See Question 17.) *After* uploading your submission to the FTP server, complete the following:
File name: _____ Size, in bytes: _____
5. (9 pts.) Indicate the result of evaluating each of the following relational algebra expressions on the database instance of Figure 1.

(a) $\pi_F \sigma_{A>50} C$

(b) $(\pi_{GF} \cup \pi_{DN}) \times (\pi_{UN} - \pi_{NC})$

(c) $\pi_{FA} \sigma_{G=\text{Nut and Seed Products}} (F \bowtie N \bowtie C)$

6. (12 pts.) Provide relational algebra queries that return the information requested in each part below, **using only the six basic operators**. (Do not use derived operators or syntactic sugar.) Reminder: As in all query-writing questions, your answers should work on *all* database instances, not only the instance of Figure 1.

(a) The amount of each nutrient, with units, in 100 g of raw bamboo shoots.

(b) The names of foods in the “nuts and seed products” food group that contain more than 20 g of protein per 100 g. Your query must not include any tag names such as PROCNT.

(c) The names of foods that either have more than 72% of refuse or have more than 60 g of water per 100 g (or both).

7. (9 pts.) Indicate the result of evaluating each of the following SQL queries on the database instance of Figure 1.

(a) `select food from Compositions order by amt desc;`

(b) `select distinct F.name, N.description
from Foods F, Nutrients N, Compositions C
where F.food like '%nut%' or F.fgroup like '%nut%' and N.units = 'g'
order by F.name, N.description desc;`

```
(c)  select food
      from Foods
      where food not in
        (select food
         from Compositions
         where amt < 2);
```

8. (12 pts.) For each part below, write a SQL query that is equivalent to the query in the corresponding part of Question 6.

(a)

(b)

(c)

9. (5 pts.) Write SQL statements to create the database instance in depicted in Figure 1, page 2. In addition to creating the tables, your answer must populate them with the instance data.

10. (18 pts.) Read about the *script*, *col*, *tar*, *gzip*, *top*, *ps*, *tee*, and *kill* Unix utilities by using the *man* command on Gandalf. (Type “`man script`” at the shell prompt, for example.) Answer the following in brief:

(a) What happens if we type “`script well`” at the shell prompt, followed by typing “`viRETohmm?ESC:w hRET:q`” (where `RET` and `ESC` denote the *Enter/Return* and *Escape* keys), followed by typing “`exit`”?

(b) What does the following command do, assuming a file named `infile` exists in the working directory?

```
cat infile | col -b | tee outfile
```

(c) What is the result of executing the command `tar cf *.sql` in a directory that contains files `a.sql` and `b.sql` (only)? Explain the result briefly. **Caution:** If you try it out, be sure that you do so in a directory that contains only junk files. Really, be careful! [Hint: This question is identical to one in last session’s class; it’s that important!]

(d) Explain the difference between the files `a.tgz` and `b.tgz` resulting from the following actions:

```
mkdir food
echo celery carrot beet > /food/veg.txt
tar cf - food | gzip > b.tgz
cd food
tar cf - food.txt | gzip > a.tgz
```

- (e) Explain the result of executing the following shell script. **Caution:** Do not blindly execute it on a public machine!

```
#!/bin/sh
i=1
while :
do
    sleep $i &
    i='expr $i + 1'
done
```

- (f) Describe an effective method to terminate the script of Question 10e, along with all the processes it may have generated before termination.

11. (15 pts.) For each relational algebra expression in Question 5, provide an equivalent SQL query in the corresponding part below. Although you do not need to provide a formal proof of equivalence, you must *explain why your answer is correct*.

(a)

(b)

(c)

12. (10 pts.) ★ Write a relational algebra query for the names of foods that have more calories than the calories in the food in their category that has the most fat *using only the six basic operators, without any derived operators or syntactic sugar.*

13. (10 pts.) ★ Write a SQL query that is equivalent to the query of Question 12 *without using any grouping or aggregation features of SQL.*

14. (30 pts.) ★ Prove or disprove each of the following for $R(I, J, K, L)$ and $S(J, K, M)$, with all attributes of integer type:

$$(a) S \cap \pi_{IJK} R \stackrel{?}{\equiv} S - (\pi_{IJK} R - S)$$

$$(b) \pi_{IJ} R \bowtie \pi_{KL} R \stackrel{?}{\equiv} R$$

$$(c) \pi_{IJK} R \bowtie \pi_{KL} R \stackrel{?}{\equiv} R$$

$$(d) \pi_{JKM}(R \bowtie S) \stackrel{?}{\equiv} S - (\pi_{JK} S - \pi_{JK} R) \bowtie S$$

[additional space for answering the earlier question]

15. (3 pts.) Execute the statements in your answer to Question 9 in PostgreSQL using the *psql* program. Using Question 10 as a hint, capture your psql session in a *plain-text ASCII* file called `dbgen-log.txt`.

For this and the next question, please check carefully that your file is a plain-text ASCII file. For example, check whether your file can be viewed properly using the command “`more dbgen-log.txt`” on *gandalf*.

Indicate the size of your file, in bytes: _____

16. (3 pts.) Execute the statements in your answer to Question 8 in *psql*. As in Question 15, capture your psql session in a *plain-text ASCII* file called `query-log.txt`.

Indicate the size of your file, in bytes: _____

17. (10 pts.) You should generate a gzip-compressed tar archive file called *M-hw01-N.tgz*, replacing *M* with your last-name and replacing *N* with an arbitrary 4-digit integer (e.g., `Doe-hw01-4242.tgz`). The execution of the following sequence of commands (replacing `Doe-hw01-4242.tgz` with the name of your file) should result in the creation of a directory `/foo/Doe-hw01` that contains the files `dbgen-log.txt` and `query-log.txt` from Questions 15 and 16. Here `/foo` is an arbitrary directory that will be specified when testing.

```
cd /foo; gzip -dc Doe-hw01-4242.tgz | tar xf -
```

Please check carefully that your file satisfies these requirements. Upload your file using anonymous FTP (`anonymous` as the user name and your email address as the password) to the FTP server `gandalf.umcs.maine.edu` in directory `/incoming/cs/cos480/`. If you need to upload an updated version of your submission for any reason, you can follow this procedure again using a different four-digit integer in the file name. (If you try using the same file name as your earlier submission, the upload will likely fail.) You will not be able to list the FTP upload directory (standard security setup), so pay attention to the diagnostic messages from your FTP program. If the messages indicate success, your file will have been uploaded. You must upload the file before you submit your hard-copy homework.