Name:

Please note that much of this assignment is devoted to getting familiar with the class accounts and submission procedure; thus, it is much simpler than the ones that will follow. You should submit (1) a hard-copy of this assignment with your answers filled in and (2) an electronic file as detailed below. You are welcome (and encouraged) to use any resources (e.g., Web sites) 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* how and why it works.

For this homework, questions marked with ⋆ are required for COS 580 students only. 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 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. COS 580 students must answer all 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.)

Some questions will use the database instance depicted in Figure 1. The two title rows of each table indicate the attribute names and types. The `Trees` relation indicates, for each tree, its common name, botanical name, type, typical trunk diameter (in inches), typical height (in feet), and minimum and maximum zones. The zones refer to regions of the U.S. where the trees are likely to grow well. The `Places` relation indicates, for each place, the city, state, population, zone (as above), subzone (a finer subdivision of zones), and average minimum temperature.

1. (1 pt.) Write your name in the space provided above.

2. (2 pts.) Read the material on the class Web site. Sign your name here to indicate that you have read this material:

3. (2 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.)

Account	User Name	Password Changed?
Gandalf		
PostgreSQL		

Trees

name	botname	ttype	dia	height	minz	maxz
varchar(50)	varchar(50)	varchar(25)	float	float	int	int
White Pine	Pinus strobus	coniferous	30.0	90.0	3	8
Pitch Pine	Pinus rigida	coniferous	18.0	35.0	5	7
Bigtooth Aspen	Populus grandidentata	deciduous	15.0	70.0	3	5
Quaking Aspen	Populus tremuloides	deciduous	13.0	67.5	1	8

Places

city	state	pop	zone	subzone	mintemp
varchar(20)	varchar(20)	integer	integer	char(1)	float
Orono	Maine	9112	5	b	-15.0
Bangor	Maine	31473	5	a	-15.1
Bar Harbor	Maine	4820	5	b	-14.0
Caribou	Maine	8312	4	a	-25.4
Van Buren	Maine	2631	3	a	-35.6
Tucson	Arizona	486699	8	a	39.0

For notational convenience in relational algebra, we shall abbreviate as follows:

Trees(name, botname, ttype, dia, height, minz, maxz) $T(N, B, T, D, H, M, X)$
Places(city, state, pop, zone, subzone, mintemp) $P(C, S, P, Z, Y, L)$

Figure 1: **Sample database instance**

4. (9 pts.) Indicate the result of evaluating each of the following relational algebra expressions on the database instance of Figure 1.

   (a) $\pi_Z P$

   (b) $\pi_{NB}\sigma_{T=\text{`coniferous'}}T \times \pi_{TD}\sigma_{T=\text{`coniferous'}}T \times \pi_{HM}\sigma_{T=\text{`coniferous'}}T$

2

(c) $\sigma_{Y=\text{'a'}} \pi_{CSY} (P \bowtie \sigma_{P>10000} P)$

5. (9 pts.) Provide relational algebra queries that return the information requested in each part below. Reminder: As in all query-writing questions, your answers should work on *all* database instances, not only the instance of Figure 1.

(a) The common names of trees with typical height greater than 66 feet.

(b) A single list of all tree names, both common names and botanical names, sorted in ascending order of names.

(c) The places (city, state) that are within the range of the Bigtooth Aspen. That is, the places whose zone is no smaller than the Bigtooth Aspen's minimum zone and no larger than its maximum zone.

6. (9 pts.) Indicate the result of evaluating each of the following SQL queries on the database instance of Figure 1.

(a)
```
select zone from Places;
```

(b)
```
select distinct name, height*12
from Trees, Places
where minz <= zone and zone <= maxz and city like 'Ba%';
```

(c)
```
select distinct name, height*12
from Trees
where minz <= some
      (select zone from Places where city like 'Ba%')
```

```
    and maxz >= some
        (select zone from Places where city like 'Ba%');
```

7. (9 pts.) For each part below, write a SQL query that is equivalent to the query in the corresponding part of Question 5.

   (a)

   (b)

   (c)

8. (10 pts.) Write SQL statements to create the database instance consisting of the tables Trees and Places depicted in Figure 1. In addition to creating the tables, your answer must populate them with the instance data.

9. (10 pts.) Read about the *script*, *col*, *tar*, *gzip*, *top*, *ps*, 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 is the default file name used by *script*?

(b) Indicate how we may use *col* to remove backspace-generating control-characters from a file, without converting blanks to tabs.

(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!

(d) Explain the difference between the files `a1.tar` and `a2.tar` resulting from the following actions:

```
mkdir /tmp/d; echo foo > /tmp/d/f1; echo bar > /tmp/d/f2
tar cf a1.tar /tmp/d
cd /tmp; tar cf a2.tar d
```

(e) Suppose you notice that the CPU usage on a machine is very high. Indicate how you may determine the process responsible for that usage and the owner of that process. Assuming that the owner is you, indicate how you would terminate the process without access to that process's user interface, if any. Provide at least two ways of terminating the errant process and explain the differences between them.

10. (10 pts.) Execute the statements in your answer to Question 8 in PostgreSQL using the *psql* program. Using Question 9 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`.)

Indicate the size of your file, in bytes: _____

11. (10 pts.) Execute the statements in your answer to Question 7 in *psql*. As in Question 10, capture your psql session in a *plain-text ASCII* file called `query-log.txt`.

Indicate the size of your file, in bytes: _____

12. (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 `/tmp/Doe-hw01` that contains the files `dbgen-log.txt` and `query-log.txt` from Questions 10 and 11:

    cd /tmp; 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.

Provide the name and size (in bytes) of the file you submitted:
name: _____ ; size: _____

13. (4 pts.) Provide a SQL query for trees and pairs of places such that the first place's zone, but not the second's, lies in the tree's zone range. More precisely, the desired output is a set of tuples of the form (t, c_1, s_1, c_2, s_2) where t is the common name of a tree, c_1 and s_2 are the city and state of the first place, p_1, and, similarly, c_2 and s_2 is the city and state of the second place, p_2. We require that p_1's zone is within t's zone range and that p_2's zone is outside t's zone range. (We say a zone z is within zone range $[x, y]$ if $x \leq z \leq y$.)

14. (5 pts.) Provide a relational algebra query that is equivalent to the query in Question 13.

15. (15 pts.) ⋆ For each relational algebra expression in Question 4, provide an equivalent SQL query in the corresponding part below. Although you do not need to (but may) provide a formal proof of equivalence, you must explain why your answer is correct.

(a)

(b)

(c)

16. (30 pts.) ⋆ Prove or disprove each of the following for $S(I, J, K)$ and $T(J, K, L)$, with all attributes of integer type:

(a) $\pi_{IJK}(S \bowtie T) \equiv ? S$

(b) $\pi_{IJK}(S \cup (S \bowtie T)) \equiv ? S$

(c) $\pi_{IJ}S \bowtie \pi_K S \equiv ? S$

(d) $\pi_{IJ} S \bowtie \pi_{JK} S \equiv ? S$

(e) $(S \bowtie S \cup S \bowtie T) - T \equiv ?S$

(f) $\pi_{IJ}\sigma_{I<J}(S \bowtie T) \equiv ? \pi_{IJ}(\sigma_{I<J}\pi_{IJ}S \bowtie T)$