

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 encouraged to use available resources, such as 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.

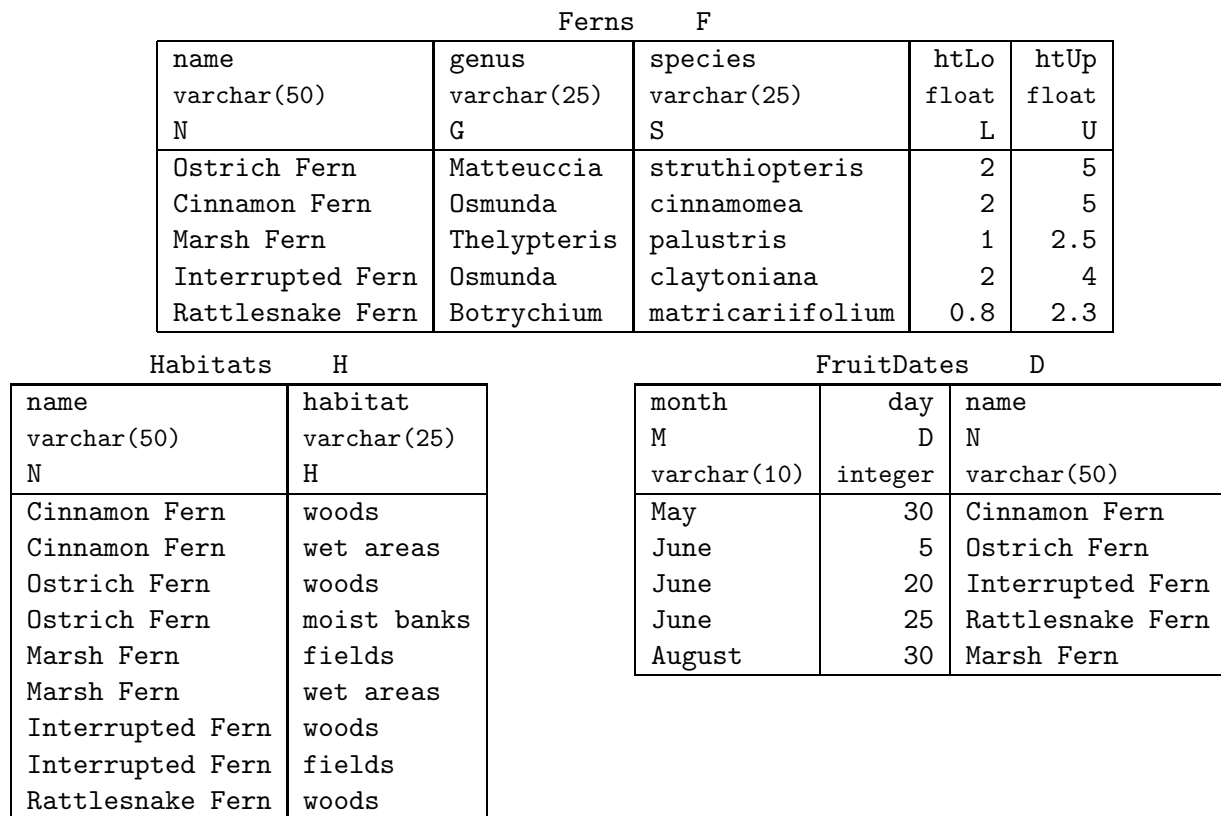


Figure 1: A database of ferns.

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 \star 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, which records some information on ferns.¹ 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 **Ferns** table records the common name, genus, and species of each fern, along with the lower and upper bounds of its typical height range, in feet. The **Habitats** table records the typical habitats for the listed ferns. The **FruitDates** table records the expected fruiting dates (month and day) of the listed ferns.

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.)

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

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

(a) $\pi_H H$

¹The data are derived from (1) George Henry Tilton. *The Fern Lover's Companion: A Guide for the Northeastern States and Canada*. Project Gutenberg, <http://www.gutenberg.org/ebooks/11365>, February 2004 and (2) Anne C. Hallowell and Barbara G. Hallowell. *Fern Finder*. Nature Study Guild Publishers, Rochester, New York, 2nd edition, 2001.

(b) π_{LUF}

(c) $\pi_{GSLU} \sigma_{H='woods'} (\pi_H H \times \pi_{LUF} \times \pi_{GS} \sigma_{G='Osmunda'} F)$

5. (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 names of ferns that grow in the woods.

(b) The genera² of ferns that fruit in the woods in June.

(c) The common names and height ranges of ferns that grow in the woods but not in fields.

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

(a) `select htLo from Ferns where htUp > 3;`

²plural of genus.

(b) `select distinct F.name, F.htLo
from Ferns F, Habitats H
where H.habitat = 'woods' and H.habitat <> 'fields'
order by F.name;`

(c) `(select habitat from Habitats)
except
(select habitat from Habitats where name = 'Ostrich Fern');`

7. (12 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. (4 pts.) Write SQL statements to create the database instance consisting of the tables **Ferns**, **Habitats**, and **FruitDates** 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) How may we invoke *script* to run a specified program instead of the default shell?

(b) What does the following command do?

```
col -b < infile > 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 plants
echo fernsRus > plants/ferns.txt
cd plants
tar cf - ferns.txt | gzip > a.tgz
cd ..
tar cf - plants | gzip > b.tgz
```


- (e) Suppose a buggy process is repeatedly spawning child processes and causing CPU and memory overload. Explain how this situation may be remedied using some of the tools listed earlier.

10. (3 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. (3 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 `/foo/Doe-hw01` that contains the files `dbgen-log.txt` and `query-log.txt` from Questions 10 and 11. (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.

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

name: _____ ; size: _____

13. (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)

17. (10 pts.) ★ Write a SQL query that is equivalent to the query of Question 16 without using any grouping or aggregation features of SQL.

18. (30 pts.) \star Prove or disprove each of the following for $S(I, J, K)$ and $T(J, K, L)$, with all attributes of integer type:

(a) $S - (S - T) \stackrel{?}{\equiv} T - (T - S)$

(b) $\pi_{JKL}(\pi_I S \times \rho_{S'(J')} \pi_J S \times T) \stackrel{?}{\equiv} T$

$$(c) \sigma_{I \leq J} S \cup \sigma_{I > J} S \equiv ? S$$

$$(d) \pi_{JKL}(\sigma_{J < K} S \bowtie T) \stackrel{?}{\equiv} \pi_{JKL} \sigma_{J < K} T$$

$$(e) \rho_{R(I,J,K)}(S \cap T) \bowtie \rho_{R(I,J,K)}(S - T) \stackrel{?}{\equiv} \rho_{R(I,J,K)}(S \cap T) - \rho_{R(I,J,K)}(S \cup T)$$