Name: _____

This exam is open book, open notes, but there can be no sharing of any material. You can use the Internet, but only as a library. If you are not sure if something is allowed, check with me.

COS 480 students must answer all questions that are not marked with a ⋆ in **40 minutes**. The points for each question are indicated in parentheses next to the question number. Questions marked with a ⋆ may also be answered, for extra credit.

COS 580 students must answer all questions, including those marked with a ⋆, in **60 minutes**. Each question is worth 2/3 times the points indicated in parentheses.

Some questions will use the database instance depicted below. The three title rows of each table indicate the names, types, and abbreviated names of attributes. The `Trees` (or `T`) 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` (or `P`) relation indicates, for each place, the city, state, population, zone (as above), subzone (a finer subdivision of zones), and average minimum temperature.

Your answers to questions that ask for queries should work for all instances of databases conforming to the given schema, not only the one depicted below.

These tables are repeated on the last page of the test. (You may detach that page and use it for reference. There is no need to reattach it.)

Trees T

name varchar(50) N	botname varchar(50) B	ttype varchar(25) T	dia float D	height float H	minz int M	maxz int X
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 P

city varchar(20) C	state varchar(20) S	pop integer P	zone integer Z	subzone char(1) Y	mintemp float L
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

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

2. (8 + 3 + 8 pts.) Consider the relation $R(A, B, C, D)$ with the following basis of functional dependencies:

$$\begin{array}{lrcl} f_1: & A & \rightarrow & B \\ f_2: & B & \rightarrow & C \\ f_3: & CD & \rightarrow & A \end{array}$$

   (a) List all keys of R. Justify your answer.

   (b) Indicate which, if any, of the functional dependencies are BCNF violations.

(c) Decompose R, as needed, to obtain a BCNF schema. Show all intermediate steps, including the functional dependencies that are used for decomposition, the set of projected dependencies, and the BCNF violations.

3. (10 pts.) Write SQL queries as directed below.

   (a) Write a SQL query to find the total population of all cities in each state, excluding states with a total population less than five million. The output should consist of tuples of the form (s, p) where p is the total population of all cities in state s and where p is at least five million. The result should be sorted in descending order of total population.

(b) A tree is said to grow well in a place if the place's zone is no smaller than the tree's minimum zone and no larger than the tree's maximum zone. Write a SQL query to find the height of the tallest tree that *does not* grow well in each place, where a place is a city-state combination. The output should consist of tuples of the form (c, s, h) where h is the height of the tallest tree that does not grow well in city c of state s. The result should be sorted in ascending order of states and, as a secondary sort, in ascending order of cities.

4. (10 pts.) For each part below, use the extended bag algebra to write a query that is equivalent to the SQL query in the corresponding part of Question 3.

   (a) [Hint: Use $\mathcal{T}_{-A}$ to sort in descending order of A]

   (b)

5. (4 pts.) Answer the following questions based on the paper by Codd in the readings.

   (a) Explain the *connection trap* using an example that is different from the one in the paper.

(b) Is a connection trap possible in a relational database? Justify your answer.

ONLY ⋆ QUESTIONS BELOW THIS POINT

6. (16 pts.) ⋆

   (a) Write a SQL query to find the common name of the tallest trees that grow well in the most populous cities of each state. Growing well is interpreted as in Question 3b. The output should consist of tuples of the form (s, n) where s is a state and n is the common name of a tree t that grows well in a city c in state s such that (1) no city in s has a population greater than that of c and (2) no tree that grows well in c has a height greater than that of t. [Hint: It may be useful to use views.]

(b) Use the extended bag algebra to write a query that is equivalent to the SQL query of Question 6a.

7. (4 pts.) ⋆ Answer the following questions based on the paper by Graefe in the readings.

   (a) Explain the *Halloween problem* using an example that is different from the one in the paper.

(b) Can the Halloween problem occur in a database system that is based on an access-oriented model, such as the network model described in class or an object-oriented database without a high-level query language? Justify your answer.

Scratch page

Material here will not be graded. You may detach and discard this page.

Trees T

name varchar(50) N	botname varchar(50) B	ttype varchar(25) T	dia float D	height float H	minz int M	maxz int X
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 P

city varchar(20) C	state varchar(20) S	pop integer P	zone integer Z	subzone char(1) Y	mintemp float L
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