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. Some questions are marked with a $\star$. The points for the non-$\star$ questions sum to 60, while those for the $\star$ questions sum to 30.

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

COS 580 students must answer all questions, including those marked with a $\star$, in 75 minutes. Each question is worth 2/3 times the points indicated in parentheses.

Several questions on this exam use the database instance suggested below. A row in the `PTides` table represents a predicted tide and lists the location, time, kind (high or low), and height of the prediction. A row in the `DockSched` table represents a scheduled arrival of a boat at a dock and lists the harbor, boat's name, pilot's name, scheduled arrival time, and boat's length. 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.)

PTides

location varchar(20)	ptime timestamp	kind char(1)	height float
Blue Hill	2005-10-13 01:27	L	0.29
Blue Hill	2005-10-13 07:42	H	9.80
Blue Hill	2005-10-13 13:47	L	1.00
Blue Hill	2005-10-13 19:59	H	10.98
Eastport	2005-10-13 01:25	L	0.77
Eastport	2005-10-13 07:31	H	17.60

DockSched

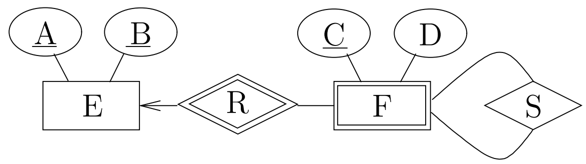
harbor varchar(20)	boat varchar(20)	pilot varchar(20)	dtime timestamp	blength integer
Blue Hill	Why Knot	Knotting	2005-10-13 08:00	14
Blue Hill	Why Knot	Knotting	2005-10-14 08:00	14
Blue Hill	Phair Game	Phair	2005-10-13 08:10	10
Castine	Phair Game	Phair	2005-10-13 08:30	10

For notational convenience in relational algebra, we use the following abbreviations:

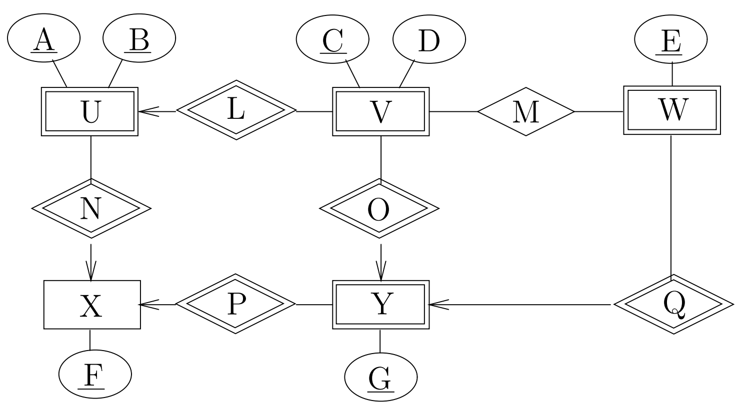
`PTides(location, ptime, kind, height)`	$P(L, P, K, H)$
`DockSched(harbor, boat, pilot, dtime, blength)`	$D(H, B, P, D, L)$

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

2. (7 pts.) Generate a relational schema that best corresponds to the following E-R diagram. Underline key attributes.



3. (7 pts.) What is the key of V in the following E-R diagram?



4. (5 pts.) Write a SQL query to find times at which Blue Hill is predicted to have a low tide of height less than 0.5 or a high tide of height greater than 10. The output should consist of tuples of the form (p, k, h) where p, k, and h are, respectively, the predicted time, the kind, and the height of a qualifying tide at Blue Hill.

5. (5 pts.) Write a relational algebra query that is equivalent to the query of Question 4.

6. (5 pts.) Write a Datalog query that is equivalent to the query of Question 4.

7. (5 pts.) Consider the relation $R(A, B, C, D, E, F, G)$ with the following basis of dependencies:

$$\begin{aligned} AB &\rightarrow CE & (1)\\ C &\rightarrow B & (2)\\ CD &\rightarrow AB & (3)\\ F &\rightarrow AG & (4)\\ EG &\rightarrow D & (5) \end{aligned}$$

Compute $\{A, B, D\}^+$ and $\{B, F\}^+$.

8. (10 pts.) Determine all keys of the relation R of Question 7. Justify your answer.

9. (15 pts.) Render the schema of Question 7 in BCNF. *For each decomposition step, you must clearly list (1) the dependency used for the decomposition and (2) a basis of the projected dependencies for the resulting relations.* Summarize the final schema (list all its relations) and explain why it is in BCNF.

10. (5 pts.) $\star$ We say that a pilot p_1 *meets* another pilot p_2 at a harbor h, and write $meets(p_1, p_2, h)$, if p_1 is scheduled to dock at h at time t_1, p_2 is scheduled to dock at h at time t_2, and t_1 lies in the one-hour long interval ending at t_2 (i.e., $t_1 \in [t_2 - 1{:}00, t_2]$). Write a Datalog program for the predicate *meets*. Note that, in general, $meets(p_1, p_2, h)$ does not imply $meets(p_2, p_1, h)$. You may use the syntax t `+ 01:00` to denote the result of adding one hour to timestamp t.

11. (5 pts.) $\star$ We say a pilot p *connects to* a pilot q, and write $connects(p, q)$, if there exists a sequence of pilots $p_1, p_2, \ldots, p_k$, with $k > 1$, such that $p_1 = p$, $p_k = q$ and, for each i in $[1, k-1]$, there is some h_i such that $meets(p_i, p_{i+1}, h_i)$. Write a Datalog program for the predicate *connects*.

12. (10 pts.) $\star$ Suppose there is a possibility that all dockings scheduled for Blue Hill within one hour of a low tide may be canceled. Write a Datalog query to find pairs of pilots who may lose connections if the cancellations occur. That is, write a Datalog query for a predicate *connloss* where *connloss*(p, q) iff *connects*(p, q) is true but will not be true if the cancellations occur. You are not permitted to modify the `DockSched` table in any way (since the cancellations are possible but not guaranteed).

13. $(2+3$ pts.) $\star$ The following questions are based on Graefe's paper in the reading list.

   (a) It is possible to use a non-clustering sparse index? If not, why? If so, how?

   (b) Provide two non-isomorphic merge-trees for 25 runs with a maximum fan-in of 7. Which one is preferable? Why?

14. (5 pts.) $\star$ Prove or disprove: A relation R that has a singleton basis of functional dependencies $\{f\}$ is in BCNF if and only if $\mathrm{LHS}(f) \cup \mathrm{RHS}(f) = \mathrm{attr}(R)$.

Scratch page

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

PTides

location varchar(20)	ptime timestamp	kind char(1)	height float
Blue Hill	2005-10-13 01:27	L	0.29
Blue Hill	2005-10-13 07:42	H	9.80
Blue Hill	2005-10-13 13:47	L	1.00
Blue Hill	2005-10-13 19:59	H	10.98
Eastport	2005-10-13 01:25	L	0.77
Eastport	2005-10-13 07:31	H	17.60

DockSched

harbor varchar(20)	boat varchar(20)	pilot varchar(20)	dtime timestamp	blength integer
Blue Hill	Why Knot	Knotting	2005-10-13 08:00	14
Blue Hill	Why Knot	Knotting	2005-10-14 08:00	14
Blue Hill	Phair Game	Phair	2005-10-13 08:10	10
Castine	Phair Game	Phair	2005-10-13 08:30	10

For notational convenience in relational algebra, we use the following abbreviations:

`PTides(location, ptime, kind, height)`	$P(L, P, K, H)$
`DockSched(harbor, boat, pilot, dtime, blength)`	$D(H, B, P, D, L)$