

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 . 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. (11 pts.) Indicate the result of evaluating each of the following SQL queries.

(a)

```
select min(height) as MinHt
from PTides
where location like '%l%'
group by kind
order by MinHt;
```

(b)

```
select D1.pilot as A, D2.pilot as B
from DockSched D1, DockSched D2
where D1.pilot <> D2.pilot and D1.harbor = D2.harbor
order by A, B;
```

(c)

```
select min(P1.height) as M
from PTides P1, PTides P2
where P1.kind = 'H'
group by P2.kind
having max(P2.height) < 10
order by M;
```

3. (12 pts.) Indicate the result of evaluating each of the following expressions in the *extended bag algebra* as discussed in class. (Recall the shorthand notation for table- and attribute-names introduced earlier.)

(a) $\pi_K \sigma_{H>0.5} P$

(b) $\tau_{LK'} \gamma_{LK} P$

(c) $\tau_B \gamma_{B, \text{sum}(L') \rightarrow X} (\pi_B D \times \rho_{D'(B'L')} \pi_{BL} D)$

4. (18 pts.) Write SQL queries as directed below.

(a) Write a SQL query to find the mean low-tide height for each location. The output should consist of tuples of the form (n, l) , where l is the mean low-tide height for location n , sorted in ascending order of l .

(b) Write a SQL query to find the names of pilots who are scheduled to dock in Blue Hill harbor during a *predicted low-tide interval*. A predicted low-tide interval is defined to be a two-hour interval of time centered on a low-tide time as predicted by PTides. You may assume the syntax $x + 01:00$ for adding one hour to the timestamp x (and a similar syntax for subtraction).

- (c) Write a SQL query to find the mean low-tide height and mean high-tide height for each location. The output should consist of tuples of the form (n, l, h) , where l and h are the mean low- and high- tide heights (respectively) for location n . The result should be sorted in ascending order of mean low tide heights. For full credit, write a query that uses only one select clause (no subqueries). You may write a query that uses more than one select clause for partial credit.

5. (18 pts.) For each part below, write a query in the *extended bag algebra* that is equivalent to the SQL query in the corresponding part of Question 4.

(a)

(b)

(c)

6. (5 pts.; 1 + 2 + 2) ★ Answer the following briefly based on the assigned reading.

(a) How do iterator-based systems typically optimize for common subexpressions?

(b) A typical drawing of a left-deep tree can be converted to one of a right-deep tree by simply flipping the diagram along a vertical axis. What, then, is the difference between a left-deep plan and a right-deep plan, other than visual representation?

(c) What is the number of distinct left-deep plans with leaf nodes drawn from the set $\{1, 2, 3, \dots, n\}$? Justify your answer.

7. (10 pts.) ★ Write a SQL query that returns the predicted time (or times, in case of ties) of the lowest low tide (in the database) for Blue Hill, *without using any aggregation functions*. (There is no partial credit for answers that use an aggregation function!)

8. (15 pts.) ★ The basic selection operator discussed in class permits only *simple predicates* of the form aCb as *selection predicates* in the subscript, where a is an attribute name, C is a simple comparison operator drawn from the set $\{=, \neq, <, >, \leq, \geq\}$, and b is either an attribute name or a literal. Three extensions to the selection operator are proposed below. For each extension, indicate whether queries using the extended form can be rewritten as equivalent queries that do not use the extended form (i.e., that use only the basic relational algebra). Justify your answers.

(a) Selection predicates may contain conjunctions of simple predicates
(e.g., $\sigma_{A=5 \wedge B<3 \wedge D>7}R$).

(b) Selection predicates may contain disjunctions of simple predicates
(e.g., $\sigma_{A=5 \vee B<3 \vee D>7}R$).

(c) Selection predicates may contain arbitrarily parenthesized conjunctions and disjunctions of simple predicates,

(e.g., $\sigma_{(A=5 \vee B<3) \wedge (A=6 \wedge B<2) \vee ((D>7 \wedge A<4) \vee (D>8 \wedge (C=1 \vee C=2)))} R$).

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) $D(H, B, P, D, L)$