

Name: _____

This midterm is **open book, open notes**, but there can be no sharing of any material. Some questions in this midterm use the following tables (which we have encountered before in the last quiz): Packages, which holds information on package IDs, weights, and the account to which each package is to be charged, and Accounts, which holds the phone number and address for each account. (The type of each column is indicated immediately below the column name.) In relational algebra expressions, we use $P(I, W, A)$ and $A(A, P, L, C)$ as short-hand for the schemas of the relations Packages and Accounts, respectively. As always, when asked for relational algebra or SQL expressions of queries, you must provide answers that work for all possible database instances, not just the example instance depicted below. For questions on E-R and ODL modeling, use the variants described in the textbook and in class.

Packages

ID varchar(10)	Weight integer	Acct varchar(8)
MD101	1	19087
MD135	2	20948
DC001	5	93498
VA1098	20	19087

Accounts

Acct varchar(8)	Phone varchar(16)	Address varchar(100)	PostCode varchar(10)
19087	124.2693	111 First St	20901
20948	394.3975	304 Misty Dr	20742
93498	879.3941	384 Frosty Rd	20742
98491	179.3343	385 Oak St	20999

- (1 pt) Write your name in the space provided above.
- (4 pts) Write a relational algebra query for the postal code of the account 93498.
- (5 pts) Answer the following questions (related to the programming assignments) in brief:
 - In an Unix shell script, how does one specify that the program `/bin/bash` is to be used for interpreting the script?
 - Why does the following HTML fragment not validate as HTML 4.0 at the W3C validator service (<http://validator.w3.org/>)? How can this problem be fixed without changing the intent of the fragment?

`Foo Bar`

4. (5 pts) Write a relational algebra query that produces a binary relation containing pairs of accounts that are in the same postal code. Each pair should be listed only once; that is, the relation should not contain both (a, b) and (b, a) for any accounts a and b .

5. (5 pts) Write a relational algebra query for the accounts that have shipped exactly one

package.

6. (5 pts) Write a SQL query that is equivalent to the query in Question 4. You may use any method (mechanical or other).

7. (5 pts) Write a SQL query that is equivalent to the query in Question 5. You may use

any method (mechanical or other).

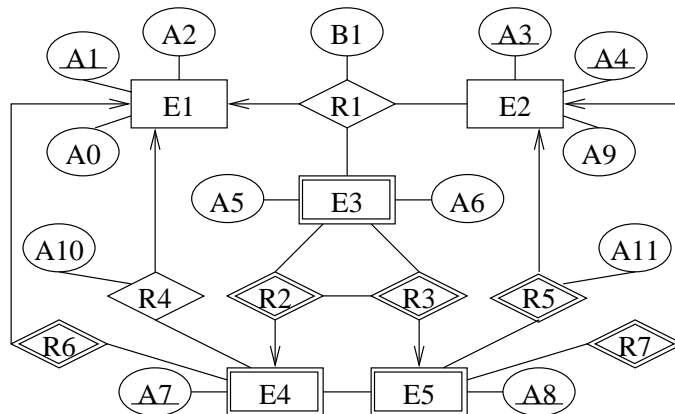
8. (5 pts) Write a SQL query that produces a frequency distribution of package weights. The output should be a table with columns **Weight** and **Frequency** and contain rows (w, f) such that there are f packages weighing w . The table should skip weights with frequency less than 10 and should be sorted by ascending weights.

9. (5 pts) The query in Question 8 is likely to be too fine-grained for most applications since it produces a count for every distinct weight value in the Packages table. We wish to fix this problem in the usual way by producing frequencies for weight ranges instead of for individual weights.

Write a SQL query that produces a table with columns **WeightLow**, **WeightHigh**, and **Frequency**, containing rows (l, h, f) such that there are f packages with weight in the range $[l, h)$ (i.e., with weight w such that $l \leq w < h$). The ranges should be chosen such that, for any row (l, h, f) , l is an integral multiple of 10 and $h = l + 10$. Empty

ranges (those with frequency 0) should be omitted from the output, which should be sorted in ascending order of the ranges. (Hint: The SQL2 standard permits arithmetic expressions in the group by and having clauses and includes a `floor` scalar function with the obvious semantics.)

10. (5 pts) There are some (zero or more) *syntactic* problems in the following E-R diagram. Indicate the simplest possible fixes for these problems (if any) by marking the figure appropriately. When other questions refer to this diagram, use your fixed version of the diagram.



11. (5 pts) In the E-R diagram of Question 10, what is the primary key of E4? Indicate

the entity set that is the source of each attribute in E4's primary key. Repeat this question for entity set E3.

12. (5 pts) Translate the E-R diagram of Question 10 into an equivalent one that does not use attributes on relationships. (Depict the resulting E-R diagram below.)

13. (5 pts) Consider an ODL schema consisting of two classes, A and B , with a relationship R from A to B and its inverse S . (The classes have no data or method members.)

(a) What are the arities of R and S ?

(b) List the possibilities for the multiplicities of R and S .

(c) For each possibility in your answer to Question 13b, present the ODL declarations for A and B .