Name: _____

Please answer the following questions within the space provided on the following pages. Should you need more space, you can use scratch paper, but clearly label on the scratch paper what problem it corresponds to. While you are not required to explain your queries, comments may help me to understand what you were trying to do and thus increase the likelihood of partial credit should something go wrong. If you get entirely stuck somewhere, explain in words as much as possible what you would try.

This is a pen and paper exam, and thus computers and internet capable devices are prohibited. You are free to use a prepared 3x5 inch index card with handwritten notes on a single side should you desire. If you have any confusion about question intention or wording, please do not hesitate to ask!

Please restrict yourself on this exam to only SQL keywords that we have discussed in class and which have shown up on homework!

Your work must be your own on this exam, and under no conditions should you discuss the exam or ask questions to anyone but myself. Failure to abide by these rules will be considered a breach of Willamette's Honor Code and will result in penalties as set forth by Willamette's academic honesty policy.

Please sign and date the below lines to indicate that you have read and understand these instructions and agree to abide by them. *Failure to abide by the rules will result in a 0 on the test.* Good luck!!

_____ _____

Signature Date

Question:	1	2	3	4	5	6	7	8	9	10	Total
Points:	5	8	12	14	4	2	6	4	10	0	65
Score:											

(5) 1. A mysterious table (named `mysterious_table`) has the following query run on it:

```
SELECT
  MAX(salmon) * trout AS col_a,
  trout + CAST(halibut AS int) - snapper AS col_b,
  cod / (snapper % 10) AS col_c
FROM mysterious_table
WHERE halibut LIKE '%0'
  AND cod::TEXT ILIKE '__0.__'
```

and returns a table with the following form:

Column Name	Data Type
col_a	DOUBLE PRECISION
col_b	INTEGER
col_c	NUMERIC

Determine *as much* information as you can about the columns comprising `mysterious_table`, and explain how you arrived at your conclusions. You can safely assume that any conversions made were necessary and that any filter statements would result it at least 1 row being shown.

Solution:

2. You have the below delimited file:

```
name:color:radius:price:amount in package
basketball:orange:24:11.97:1
bowling ball:teal:12:115:1
ping pong ball:white:0.79:9.68:12
tennis ball:green:6.86:4.99:3
```

(3) (a) Write out a command to create a table that will hold this information, including appropriate data types.

Solution:

(3) (b) Write out a command to import the data from the delimited file into your above created table. You can assume the file is located at `/data/balls.csv`.

Solution:

(2) (c) After importing the data, you realize that your table is still missing the information for volleyballs, which are white, have a radius of 21cm, and cost around $45 for a single ball. Write a command to add this information to the end of your table.

Solution:

3. You have a table named `fridge_items` in your database that looks like below.

`name` *TEXT*	`type` *TEXT*	`weight` *FLOAT*	`price` *NUMERIC(10,2)*	`fluid_vol` *FLOAT*
Milk	dairy	`NULL`	3.50	128
Ketchup	condiment	`NULL`	4.00	32
Cheese	dairy	32	14.00	`NULL`
Apples	fruit	8	1.00	`NULL`
Salami	meat	16	7.00	`NULL`
Chicken	meat	64	12.00	`NULL`
Yogurt	dairy	`NULL`	3.50	32

Use it to determine the output of the below queries, **including all column names and type**.

(3) (a)
```
SELECT DISTINCT type, fluid_vol
FROM fridge_items
WHERE price < 10
ORDER BY type, fluid_vol
```

Solution:

(3) (b)
```
SELECT SUM(weight) AS sum
FROM fridge_items
WHERE name LIKE '%s%'
```

Solution:

(3) (c)
```
SELECT COUNT(f2.price) AS count
FROM fridge_items AS f1
LEFT JOIN fridge_items AS f2
  ON f1.weight = f2.fluid_vol
```

Solution:

(3) (d)
```
SELECT
  percentile_cont(0.5) WITHIN GROUP
      (ORDER BY price) AS mdprice
FROM fridge_items
WHERE fluid_vol IS NOT NULL OR price % 1 > 0
```

Solution:

4. Suppose you are keeping golf statistics in a database comprised of three tables: one name `players`, one named `rounds`, and the last named `events`. The schema and a description of the contents of each table is below:

players

Name	Type	Description
`pid`	`TEXT`	A unique player ID number
`full_name`	`TEXT`	The full players name, first then last
`birthdate`	`DATE`	The date of the players birth

rounds

Name	Type	Description
`pid`	`TEXT`	The player ID of the participating player
`eid`	`INT`	The unique event ID number
`round_num`	`INT`	What round this is at the given event
`strokes`	`INT`	The total number of strokes across 18 holes that the player took in that round
`date_played`	`DATE`	The date the round was played

events

Name	Type	Description
`eid`	`SERIAL`	A unique event ID number
`event_name`	`TEXT`	The full event name
`par_score`	`INT`	The par number of strokes across all 18 holes for the event course

Every time a player plays another round at a given event, or plays at another event, another row is added to the table `rounds` table, with necessary additions to the `players` or `events` table as needed (if a new player or a new event). Any `pid` or `eid` in the `rounds` table will have a corresponding value in the `players` or `events` tables, respectively. In case you are unfamiliar, a reminder that in golf the object is to have the fewest number of strokes, and that a frequent goal is to "shoot below par", which is to score a total number of strokes fewer than the given par number of strokes.

Using these tables, construct queries that would answer the following questions.

(3) (a) What different events did the player with pid `007` play at in the year 2021?

Solution:

(3) (b) What is the average number of strokes of all players with first name "Beth" who played a round of golf on the 4th of July, 2022?

Solution:

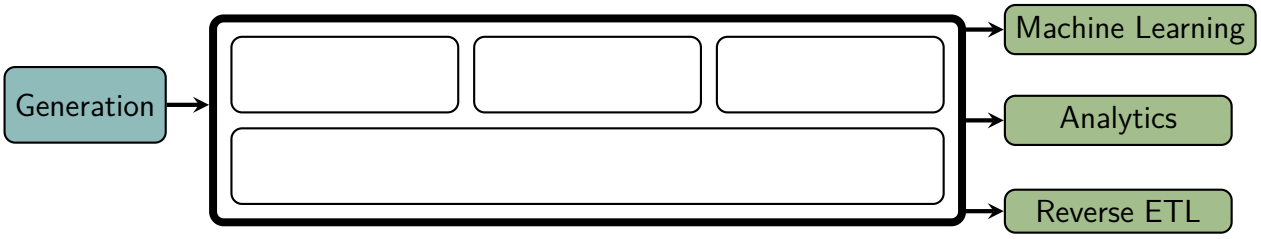
(4) (c) What is the name of the youngest player to shoot below par in the second round at the event named `'Augusta 2021'`? In the case of a tie, take the one that comes first in the alphabet.

Solution:

(4) (d) What player won the first round of the `'Augusta 2021'` event by shooting the lowest score?

Solution:

(4) 5. Fill in the names of the components/steps which make up the Data Engineering Lifecycle in the image below.



(2) 6. You have a set of data which contains predominantly one-to-many relationships and where outside relationships are rare. What type of storage model would likely be most ideal?

A. A document model

B. A runway model

C. A relational model

D. A graph model

(6) 7. Use the folder tree to the lower left to construct a BASH command that would answer the questions to the below right. For each question you can assume that you are currently located in the `class` folder.

```
/
├── documents
│   ├── class
│   ├── testing
│   └── fun
├── images
│   ├── screenshots
│   └── wallpapers
├── videos
├── applications
└── src
    ├── R
    └── Python
```

(a) You want to move the file `awesome.py` from the `testing` to the `Python` folder *using only relative paths*.

Solution:

(b) You want to get the number of files and folders in your `wallpapers` folder. You can use either relative or absolute paths.

Solution:

(c) You have written or downloaded a cool program called `diagnostics` which, when run, prints to the screen a bunch of diagnostic information about the state of your system. You'd like to save this information in a file called `diagnostic_info.txt` in the `fun` folder.

Solution:

(4) 8. On your laptop, you have defined a `.ssh/config` file with the following contents:

```
Host home
    User henry
    HostName 56.211.23.1

Host office
    User hmartin
    HostName www.hmartin.willamette.edu
    Port 39870
```

All remotes are running an SSH server, and SSH keys have been set up and shared between each remote and your laptop. Use this information to write out BASH commands that could be run to achieve the following tasks.

(a) You are currently on your laptop, and desire to copy the file called `important.txt` from the current directory on your laptop to the home directory of your office computer, keeping the filename the same. A reminder that a user's "home directory" is the directory that they initially end up in upon logging in.

Solution:

(b) Suppose you find yourself at the office one day without your laptop. From your *office* computer, which does not currently have a `.ssh/config` setup (but does have SSH installed), how could you copy the file `todo.md` from your home server's home directory to your office computer's home directory?

Solution:

(10) 9. **Database Admin:** Choose **one** of the following prompts to discuss in 4-6 sentences. Circle the prompt please so that it is clear which you are responding to.

(a) Your superior has recently become frustrated about how long it is taking different analysis programs to run on large datasets. How would you discuss with them about what scaling options exist, and what the costs and benefits of different scaling options might be?

(b) Your superior has recently become frustrated by large amounts of downtime on your main database pipeline owing to failing parts and some costly user mistakes. How would you discuss with them various areas and ways to improve maintenance policies and avoid these issues in the future?

(4 (bonus)) 10. Many databases are said to support varying level of ACID characteristics. What do the letters of ACID stand for, and describe what each is referring to in a sentence.

Solution: