

Topic Overview

The content that we have covered on this first test is fairly tightly focused, but here is a breakdown of the general topics you should be prepared for:

- Creating tables
 - Choosing proper data types
 - Importing and Exporting from/to CSV
 - Creating from a **SELECT** query
- Making selections
 - Choosing unique entries
 - Choosing desired columns
 - Aliasing desired column names
 - Filtering properly using **WHERE** and boolean operators
 - Sorting
 - Limiting
- Calculations
 - Data type of common operation outputs
 - Column operations
 - Basic arithmetic operations
 - Applying common, built-in functions
 - Aggregates
 - Simple aggregates like **avg()**, **sum()**, etc
 - Order dependent aggregates like **percentile_cont()** and **mode()**

Question Types

Questions will fall into 3-4 main divisions, of which I will include examples of each later in the study guide.

Qualitative: In general, these wouldn't deal with direct values in a table, but are more conceptual in understanding what a particular piece of SQL is doing.

- Given a general table and query, describe what the output would look like, or what properties it might have.
- Given a desired output, what properties might the query or initial table have needed to possess?
- Given a table and desired output, what would the query need to look like?

Quantitative: These will deal more directly with sample data in a table.

- For this particular query with this tabular data, what would the output be? (These will naturally be with small and simple tables, as you won't have a computer to aid you.)

Example Questions

1. You have a particular table in your database called `inventory` that follows the below schema and has at least one row of data.

Column Name	Data Type
<code>id</code>	<code>SERIAL</code>
<code>name</code>	<code>VARCHAR(20)</code>
<code>weight</code>	<code>REAL</code>
<code>price</code>	<code>NUMERIC(5,2)</code>
<code>stock</code>	<code>INT</code>

You then run the following query:

```
SELECT COUNT(weight) / COUNT(*) * 100::REAL  
FROM inventory;
```

- (a) How many columns are returned in the output?
- A. 0
 - B. 1
 - C. 5
 - D. Impossible to tell
- (b) How many rows are returned in the output?
- A. 0
 - B. 1
 - C. The same as the number of rows in the `id` column
 - D. Impossible to tell
- (c) For each column that is returned, what would be its corresponding data type?
- (d) In a sentence or two, describe what this query is doing. I'm looking less for a line by line description of what is happening, and more an overall description of what the query is trying to achieve.

2. Without any information about the table called `mystery`, you run the below query:

```
SELECT
  dim1 * dim2 * dim3 AS volume,
  |/(score::DECIMAL + 10) AS metric
FROM mystery
WHERE best_by + '3 days 10 minutes' < sold
ORDER BY score::DECIMAL
```

where any type conversions were **necessary** (not optional). The resulting table has the form:

Column Name	Data Type
<code>volume</code>	<code>NUMERIC</code>
<code>metric</code>	<code>DOUBLE PRECISION</code>

Write as *much detail as you can* about what you know about the table `mystery` from just this query and its results.

3. Suppose I wanted to import the below CSV file (saved at `C:\Data\important.csv`) into a Postgresql database. Write out the necessary commands to create the table and import the data.

```
id,name,p1,p2,p3,total,submitted
1,Bill,7,8,2,17,2022-01-25 18:00
2,Nancy,7,7,7,21,2022-01-26 15:15
3,Jacob,5,10,5.6,20.6,2022-01-25 23:47
4,Sebastian,9.5,10,10,29.5,2022-01-29 19:34
```

4. You have a table named `special` in your database, that looks as can be seen below:

<code>id</code>	<code>name</code>	<code>cola</code>	<code>colb</code>	<code>colc</code>
<i>SERIAL</i>	<i>TEXT</i>	<i>INT</i>	<i>NUMERIC(4,2)</i>	<i>INT</i>
1	Bob	3	4.50	9
2	Bob	2	2.00	5
3	Bob	NULL	4.10	4
4	Bob	5	12.40	10
5	Bob	8	NULL	7

(a) What would be the output of the below query?

```
SELECT
  name,
  colb / (colc / cola) AS o1,
  2 * colc + colb AS o2
FROM special
WHERE colb IS NOT NULL
ORDER BY o1
```

(b) What would be the output of the below query?

```
SELECT
  min(colc - cola) AS mind,
  percentile_disc(0.5) WITHIN GROUP (ORDER BY name) AS midname,
  sum(colb + colc) AS summy
FROM special
WHERE id % 2 > 0;
```

5. You have the table (named `teachers`) of teachers in your local area with a schema given below, where I have also added a quick description of each column.

Column Name	Data Type	Description
<code>id</code>	<code>SERIAL</code>	Unique identifying integer
<code>name</code>	<code>TEXT</code>	Full name of the teacher
<code>pronoun</code>	<code>VARCHAR(4)</code>	Preferred pronouns: "she", "he" or "they"
<code>grade</code>	<code>INT</code>	Grade level taught. Kindergarden is 0.
<code>yr_exp</code>	<code>INT</code>	Years of teaching experience
<code>salary</code>	<code>NUMERIC(8,2)</code>	Yearly salary in US dollars
<code>peak_deg</code>	<code>VARCHAR(3)</code>	Peak degree obtained: HS,BS/BA,MS,PhD

Write out queries to answer the following questions.

- (a) What is the average salary of high school (grades 9-12) teachers with graduate degrees?

- (b) What Ms. or Mrs. Johnson has been teaching for the longest?