
1. PostgreSQL (Banco de Dados)

Questão 01

Consider the given table "a". Which PostgreSQL statement produces the following single row? 1, 2, 3, 4, 5, 1, 5, 6, 1

Table a		
id	name	quantity
1	Rice	32
2	Beans	34
3	Tomatoes	12
4	Onion	112
5	Potatoes	222
1	Rice	31
5	Potatoes	1
6	Candy	12
1	Rice	354

- A) SELECT GROUP_CONCAT(CAST(id as varchar), ', ') FROM a;
- B) SELECT AGREGATE(CAST(id from varchar), ', ') FROM a;
- C) SELECT STRING_CONCAT(CAST(id as varchar), ', ') FROM a;
- D) SELECT STRING_JOIN(CAST(id from varchar), ', ') FROM a;
- E) SELECT STRING_AGG(cast(id as varchar), ', ') FROM a;

Questão 02

Consider a PostgreSQL table "a" created with the following SQL statement: `CREATE TABLE a (n VARCHAR(100));`

How many lines will be available in a table after executing the following transaction?

SQL

BEGIN;

INSERT 'A';

SAVEPOINT x1;

INSERT 'B';

SAVEPOINT x2;

INSERT 'C';

ROLLBACK TO x1;

INSERT 'D';

COMMIT;

Choose the correct option

- A) 0
- B) 1
- C) 2
- D) 3
- E) 4

Questão 03

A table is created using the following PostgreSQL statement: `CREATE TABLE product (id INT, name VARCHAR(100), type VARCHAR(100));`

What is the correct statement to retrieve the quantity of products with different names?

- A) `SELECT NUMBER (Produto) FROM Estoque;`
- B) `SELECT SUM (Produto) FROM Estoque;`
- C) `SELECT COUNT (DISTINCT name) FROM product;`
- D) `SELECT TOTAL (EXCLUSIVE name) FROM product;`
- E) `SELECT RESUME (name) FROM product;`

Questão 04

Which PostgreSQL statement has a valid syntax?

- A) `DELETE FROM table_event WHERE event_date = '13-AUG-2013';`
- B) `CREATE TABLE table2 (id integer, x text, y numeric);`
- C) `INSERT INTO t DEFAULT VALUE;`
- D) `SELECT a, b, c FROM t ORDER FOR a AFTER b;`

Questão 05

What does the following PostgreSQL function do?

SQL

`CREATE OR REPLACE FUNCTION x(integer, integer) RETURNS integer`

`AS $$`

`DECLARE retval integer;`

`BEGIN`

`SELECT $1 + $2 INTO retval;`

`RETURN retval;`

`END;`

`$$`

`LANGUAGE plpgsql;`

- A) It returns the sum of two columns of a data table
- B) It returns one of the parameters, i.e., it's a prime number
- C) It returns two columns from two tables
- D) It finds the constants \$1 and \$2 in a database table
- E) It returns the sum of the two numbers passed as parameter

2. Google BigQuery

Questão 06

Data ingested from sensors using Google BigQuery now requires extraction. How many methods are available to accomplish this task using BigQuery's data extraction capabilities?

- A) 1
- B) 2
- C) 3
- D) 4

Questão 07

What is the primary benefit of using BigQuery's columnar storage format?

- A) BigQuery is a fully-managed relational database management system
- B) BigQuery is a NoSQL database for real-time data processing
- C) BigQuery is a serverless, fully managed, and scalable data warehouse for analytics
- D) BigQuery supports Online Transaction Processing (OLTP) workloads

Questão 08

You want to query data across multiple tables using SQL in Google BigQuery. Which query syntax should be used to accomplish this task?

- A) Select * From `bigquery-public-data.demo_data.gsod`
- B) Select * From `'bigquery-public-data.demo_data.gsod'`
- C) Select * From `*bigquery-public-data.demo_data.gsod*`
- D) Select * From `'bigquery-public-data.demo_data.gsod*'`

Questão 09

A company intends to optimize query performance and reduce costs while migrating its on-premises data warehouse to Google BigQuery. As the cloud architect, you must recommend strategies to improve query performance while minimizing expenses. Which strategy should you adopt?

- A) Configure BigQuery slots in advance to ensure dedicated resources.
- B) Leverage BigQuery's caching mechanism to store and reuse results.
- C) Use BigQuery's partitioning feature to divide data into smaller subsets based on data size.
- D) Implement materialized views in BigQuery to store precomputed results.

3. Estatística e Análise de Dados

Questão 10

Sarah finds that the first three principal components (PCA) explain 70% of the total variance in the data. What can Sarah infer?

- A) The first three variables in the dataset explain 70% of the total variance.
- B) Three variables can explain 70% of the variance, but PCA does not indicate which ones.
- C) The first three principal components represent three original variables.
- D) The first three principal components capture 70% of the information and are combinations of the original variables.

Questão 11

In a logistic regression model, the coefficient for a certain factor is 1.7. What is the correct interpretation?

- A) For every additional year of age, the log odds increase by 1.7.
- B) For individuals with the factor, the log odds are 1.7 times higher than without it.
- C) For individuals with the factor, the log odds increase by 1.7, holding age constant.
- D) The probability of developing the condition is 1.7 times higher with the factor.

Aqui estão as perguntas da seção de Análise Estatística formatadas conforme solicitado:

4. Análise Estatística

Questão 12

Sarah, a data scientist, is working with a large dataset containing information about customer behaviors at a supermarket. The dataset includes 20 variables, such as the amount of money spent on different product categories, shopping frequency, and time spent in the supermarket. Sarah decides to use Principal Component Analysis (PCA) to identify the key patterns in the data and reduce the complexity of her subsequent predictive models. After running PCA, Sarah finds that the first three principal components explain 70% of the total variance in the data. What can Sarah infer from this finding?

- A) The first three variables in the dataset explain 70% of the total variance.
- B) Three variables can explain 70% of the variance in customer behavior, but PCA does not indicate which ones.
- C) The first three principal components represent three original variables that explain 70% of the total variance.
- D) The first three principal components capture 70% of the information in the original data and are combinations of the original variables.

Questão 13

Dr. Reed, a medical researcher, is conducting a study on the effects of a certain factor and age on the likelihood of a health condition. He uses a logistic regression model with a certain factor (presence/absence) and age as the independent variables, and the development of the health condition (yes/no) as the dependent variable. The coefficient for the certain factor in his model is 1.7. Based on the coefficient for the certain factor in Dr. Reed's logistic regression model, which is the correct interpretation?

- A) For every additional year of age, the log odds of developing the health condition increase by 1.7, holding the certain factor constant.
 - B) For individuals with the certain factor, the log odds of developing the health condition are 1.7 times higher than for individuals without the certain factor, holding age constant.
 - C) For individuals with the certain factor, the log odds of developing the health condition increase by 1.7, holding age constant.
 - D) The probability of an individual with the certain factor developing the health condition is 1.7 times higher than an individual without the certain factor, holding age constant.
-

4. Experiência e Deployment (Dissertativas)

Questão 14

Describe your experience creating and deploying a machine learning model from scratch:

- How did you approach the problem and what techniques did you use?
- What tools did you use?
- How did you prepare the model for production (daily inferences)?
- How did you deploy it in a cloud environment (AWS, GCP, Azure)?

Questão 15

Tell me about a predictive model you built end-to-end, from raw data to production. What was the problem, which algorithms did you choose, and how did you ensure reliability after deployment?

Seção: Git and GitHub

Questão 16

Jane uses Git for version control in her project. After several commits, she wants to review the history of her changes to ensure everything has been documented correctly. Which Git command should she use?

- A) git log
- B) git history
- C) git show
- D) git status

Questão 17

Kylie must move her code from the stage to commit to the repo. She must provide a message along with each commit to track each commit and progress as she keeps adding code to the repo. Which code should Kylie use to achieve this functionality?

- A) git commit -m "Completed the first stage."
- B) git commit message "Completed the first stage."
- C) git commit -msg "Completed the first stage."
- D) git commit git -m "Completed the first stage."

Questão 18

The changes in your local Git must be pushed to GitHub. You have committed the changes using `git commit -a -m committing`. Which option will you choose to push the changes in the branch to its default remote origin?

- A) git push github
- B) git push remote
- C) git push origin
- D) Any of the options

Questão 19

Ronaldo wants to specify the branch on which his workflow should run by default. Which keyword should he use in the workflow configuration?

- A) default-branch
- B) base-branch
- C) branch
- D) main-branch

Questão 20

Diana wants to run a workflow only when a pull request is opened. Which event should she specify in the workflow configuration?

- A) opened
- B) pull-request
- C) on-pull-request
- D) pull request

5. Coding (Programação)

Questão 21

Strings: Classify & Count Characters

Analyze the input string and count how many characters belong to:

1. **Letters:** A-Z, a-z.
2. **Digits:** 0-9.
3. **Special characters:** spaces, punctuation, symbols.
 - **Sample Input:** Admin@123!
 - **Sample Output:** Letters: 5, Digits: 3, Special Characters: 2.

```
def classifyCharacters (str):
```

```
    #write your Logic here:
```

```
    #this is default OUTPUT. You can change it.
```

```
    print("-404")
```

```
#INPUT [uncomment & modify if required]
```

```
str = input()
```

```
#OUTPUT [uncomment & modify if required]
```

```
classifyCharacters (str)
```