DSA-C02^Q&As

SnowPro Advanced: Data Scientist Certification

Pass Snowflake DSA-C02 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/dsa-c02.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Snowflake Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

A Data Scientist as data providers require to allow consumers to access all databases and database objects in a share by granting a single privilege on shared databases. Which one is incorrect SnowSQL command used by her while doing this task?

Assuming:

A database named product_db exists with a schema named product_agg and a table named Item_agg.

The database, schema, and table will be shared with two accounts named xy12345 and yz23456.

1.USE ROLE accountadmin;

2.CREATE DIRECT SHARE product_s;

3.GRANT USAGE ON DATABASE product_db TO SHARE product_s;

4.GRANT USAGE ON SCHEMA product_db. product_agg TO SHARE product_s;

5.GRANT SELECT ON TABLE sales_db. product_agg.Item_agg TO SHARE product_s; 6.SHOW GRANTS TO SHARE product_s;

7.ALTER SHARE product_s ADD ACCOUNTS=xy12345, yz23456;

8.SHOW GRANTS OF SHARE product_s;

A. GRANT USAGE ON DATABASE product_db TO SHARE product_s;

B. CREATE DIRECT SHARE product_s;

C. GRANT SELECT ON TABLE sales_db. product_agg.Item_agg TO SHARE product_s;

D. ALTER SHARE product_s ADD ACCOUNTS=xy12345, yz23456;

Correct Answer: C

Explanation:

CREATE SHARE product_s is the correct Snowsql command to create Share object.

Rest are correct ones.

https://docs.snowflake.com/en/user-guide/data-sharing-provider#creating-a-share-using-sql

QUESTION 2

Which method is used for detecting data outliers in Machine learning?

A. Scaler

B. Z-Score

C. BOXI

D. CMIYC

Correct Answer: B

Explanation:

What are outliers?

Outliers are the values that look different from the other values in the data. Below is a plot high-lighting the outliers in `red\\' and outliers can be seen in both the extremes of data.

Reasons for outliers in data

Errors during data entry or a faulty measuring device (a faulty sensor may result in extreme readings).

Natural occurrence (salaries of junior level employees vs C-level employees) Problems caused by outliers

Outliers in the data may causes problems during model fitting (esp. linear models). Outliers may inflate the error metrics which give higher weights to large errors (example, mean squared error, RMSE).

Z-score method is of the method for detecting outliers. This methodis generally used when a variable\\' distribution looks close to Gaussian. Z-score is the number of standard deviations a value of a variable is away from the variable\\' mean.

Z-Score = (X-mean) / Standard deviation

IQR method , Box plots are some more example of methods used to detect data outliers in Data science.

QUESTION 3

All Snowpark ML modeling and preprocessing classes are in the _____ namespace?

A. snowpark.ml.modeling

B. snowflake.sklearn.modeling

C. snowflake.scikit.modeling

D. snowflake.ml.modeling

Correct Answer: D

Explanation:

All Snowpark ML modeling and preprocessing classes are in the snowflake.ml.modeling namespace. The Snowpark ML modules have the same name as the corresponding module from the sklearn namespace. For example, the Snowpark

ML module corresponding to sklearn.calibration is snow-flake.ml.modeling.calibration. The xgboost and lightgbm modules correspond to snowflake.ml.modeling.xgboost and snow-flake.ml.modeling.lightgbm, respectively.

Not all of the classes from scikit-learn are supported in Snowpark ML.

QUESTION 4

Which of the following Snowflake parameter can be used to Automatically Suspend Tasks which are running Data science pipelines after specified Failed Runs?

A. SUSPEND_TASK

B. SUSPEND_TASK_AUTO_NUM_FAILURES

C. SUSPEND_TASK_AFTER_NUM_FAILURES

D. There is none as such available.

Correct Answer: C

Explanation:

Automatically Suspend Tasks After Failed Runs

Optionally suspend tasks automatically after a specified number of consecutive runs that either fail or time out. This feature can reduce costs by suspending tasks that consume Snowflake credits but fail to run to completion. Failed task runs

include runs in which the SQL code in the task body either produces a user error or times out. Task runs that are skipped, canceled, or that fail due to a sys-tem error are considered indeterminate and are not included in the count of failed

task runs.

Set the SUSPEND_TASK_AFTER_NUM_FAILURES = num parameter on a standalone task or the root task in a DAG. When the parameter is set to a value greater than 0, the following behavior applies to runs of the standalone task or DAG:

Standalone tasks are automatically suspended after the specified number of consecutive task runs either fail or time out.

The root task is automatically suspended after the run of any single task in a DAG fails or times out the specified number of times in consecutive runs. The parameter can be set when creating a task (using CREATE TASK) or later (using

ALTER TASK). The setting applies to tasks that rely on either Snowflake-managed compute resources (i.e. serverless compute model) or user-managed compute resources (i.e. a virtual warehouse).

The SUSPEND_TASK_AFTER_NUM_FAILURES parameter can also be set at the account, database, or schema level. The setting applies to all standalone or root tasks contained in the modified object. Note that explicitly setting the

parameter at a lower (i.e. more granular) level overrides the parameter value set at a higher level.

QUESTION 5

Which of the following is a Python-based web application framework for visualizing data and analyzing results in a more efficient and flexible way?

A. StreamBI

B. Streamlit

C. Streamsets

D. Rapter

Correct Answer: B

Explanation:

Streamlit is a Python-based web application framework for visualizing data and analyzing results in a more efficient and flexible way. It is an open source library that assists data scientists and academics to develop Machine Learning (ML)

visualization dashboards in a short period of time. We can build and deploy powerful data applications with just a few lines of code.

Why Streamlit?

Currently, real-world applications are in high demand and developers are developing new libraries and frameworks to make on-the-go dashboards easier to build and deploy. Streamlit is a library that reduces your dashboard development

time from days to hours. Following are some reasons to choose the Streamlit:

It is a free and open-source library.

Installing Streamlit is as simple as installing any other python package It is easy to learn because you won\'t need any web development experience, only a basic under-standing of Python is enough to build a data application. It is compatible

with almost all machine learning frameworks, including Tensorflow and Pytorch, Scikit-learn, and visualization libraries such as Seaborn, Altair, Plotly, and many others.

[Latest DSA-C02 Dumps](...)   [DSA-C02 PDF Dumps](...)   [DSA-C02 Exam Questions](...)