



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.pass4itsure.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**

Which are the following additional Metadata columns Stream contains that could be used for creating Efficient Data science Pipelines and helps in transforming only the New/Modified data only? Choose 3.

- A. METADATA\$ACTION
- B. METADATA\$FILE_ID
- C. METADATA\$ISUPDATE
- D. METADATA\$DELETE
- E. METADATA\$ROW_ID

Correct Answer: ACE

Explanation:

A stream stores an offset for the source object and not any actual table columns or data. When queried, a stream accesses and returns the historic data in the same shape as the source object (i.e. the same column names and ordering) with

the following additional columns:

METADATA\$ACTION

Indicates the DML operation (INSERT, DELETE) recorded.

METADATA\$ISUPDATE

Indicates whether the operation was part of an UPDATE statement. Updates to rows in the source object are represented as a pair of DELETE and INSERT records in the stream with a metadata column METADATA\$ISUPDATE values set to

TRUE. Note that streams record the differences between two offsets. If a row is added and then updated in the current offset, the delta change is a new row. The METADATA\$ISUPDATE row records a FALSE value.

METADATA\$ROW_ID

Specifies the unique and immutable ID for the row, which can be used to track changes to specific rows over time.

QUESTION 2

Mark the Incorrect understanding of Data Scientist about Streams? Choose 2.

- A. Streams on views support both local views and views shared using Snowflake Secure Data Sharing, including secure views.
- B. Streams can track changes in materialized views.
- C. Streams itself does not contain any table data.



D. Streams do not support repeatable read isolation.

Correct Answer: BD

Explanation: Streams on views support both local views and views shared using Snowflake Secure Data Sharing, including secure views. Currently, streams cannot track changes in materialized views. stream itself does not contain any table data. A stream only stores an offset for the source object and returns CDC records by leveraging the versioning history for the source object. When the first stream for a table is created, several hidden columns are added to the source table and begin storing change tracking metadata. These columns consume a small amount of storage. The CDC records returned when querying a stream rely on a combination of the offset stored in the stream and the change tracking metadata stored in the table. Note that for streams on views, change tracking must be enabled explicitly for the view and underlying tables to add the hidden columns to these tables. Streams support repeatable read isolation. In repeatable read mode, multiple SQL statements within a transaction see the same set of records in a stream. This differs from the read committed mode supported for tables, in which statements see any changes made by previous statements executed within the same transaction, even though those changes are not yet committed. The delta records returned by streams in a transaction is the range from the current position of the stream until the transaction start time. The stream position advances to the transaction start time if the transaction commits; otherwise it stays at the same position.

QUESTION 3

There are a couple of different types of classification tasks in machine learning, Choose the Correct Classification which best categorized the below Application Tasks in Machine learning?

To detect whether email is spam or not

To determine whether or not a patient has a certain disease in medicine.

To determine whether or not quality specifications were met when it comes to QA (Quality Assurance).

A. Multi-Label Classification

B. Multi-Class Classification

C. Binary Classification

D. Logistic Regression

Correct Answer: C

Explanation: The Supervised Machine Learning algorithm can be broadly classified into Regression and Classification Algorithms. In Regression algorithms, we have predicted the output for continuous values, but to predict the categorical values, we need Classification algorithms.

What is the Classification Algorithm?

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then

classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc. Classes can be called as targets/labels or categories. Unlike regression, the output variable of Classification is a

category, not a value, such as "Green or Blue", "fruit or animal", etc. Since the Classification algorithm is a Supervised



learning technique, hence it takes labeled input data, which means it contains input with the corresponding output.

In classification algorithm, a discrete output function(y) is mapped to input variable(x).

$y=f(x)$, where y = categorical output

The best example of an ML classification algorithm is Email Spam Detector. The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the

categorical data. The algorithm which implements the classification on a dataset is known as a classifier.

There are two types of Classifications:

Binary Classifier: If the classification problem has only two possible outcomes, then it is called as Binary Classifier.

Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc. Multi-class Classifier: If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.

Example: Classifications of types of crops, Classification of types of music. Binary classification in deep learning refers to the type of classification where we have two class labels - one normal and one abnormal. Some examples of binary

classification use:

To detect whether email is spam or not

To determine whether or not a patient has a certain disease in medicine.

To determine whether or not quality specifications were met when it comes to QA (Quality Assurance).

For example, the normal class label would be that a patient has the disease, and the abnormal class label would be that they do not, or vice-versa. As is with every other type of classification, it is only as good as the binary classification

dataset that it has ?or, in other words, the more training and data it has, the better it is.

QUESTION 4

Which one is not the types of Feature Engineering Transformation?

- A. Scaling
- B. Encoding
- C. Aggregation
- D. Normalization

Correct Answer: C

Explanation:

What is Feature Engineering?

Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available



data to build more accurate and efficient machine learning models.

The success of machine learning models heavily depends on the quality of the features used to train them. Feature engineering involves a set of techniques that enable us to create new features by combining or transforming the existing

ones. These techniques help to highlight the most important patterns and relationships in the data, which in turn helps the machine learning model to learn from the data more effectively.

What is a Feature?

In the context of machine learning, a feature (also known as a variable or attribute) is an individual measurable property or characteristic of a data point that is used as input for a machine learning algorithm. Features can be numerical,

categorical, or text-based, and they represent different aspects of the data that are relevant to the problem at hand. For example, in a dataset of housing prices, features could include the number of bedrooms, the square footage, the location,

and the age of the property. In a dataset of customer demographics, features could include age, gender, income level, and occupation. The choice and quality of features are critical in machine learning, as they can greatly impact the accuracy and performance of the model.

Why do we Engineer Features?

We engineer features to improve the performance of machine learning models by providing them with relevant and informative input data. Raw data may contain noise, irrelevant information, or missing values, which can lead to inaccurate or

biased model predictions. By engineering features, we can extract meaningful information from the raw data, create new variables that capture important patterns and relationships, and transform the data into a more suitable format for

machine learning algorithms. Feature engineering can also help in addressing issues such as overfitting, underfitting, and high dimensionality. For example, by reducing the number of features, we can prevent the model from becoming too

complex or overfitting to the training data. By selecting the most relevant features, we can improve the model's accuracy and interpretability. In addition, feature engineering is a crucial step in preparing data for analysis and decision-making

in various fields, such as finance, healthcare, marketing, and social sciences. It can help uncover hidden insights, identify trends and patterns, and support data-driven decision-making.

We engineer features for various reasons, and some of the main reasons include:

Improve User Experience: The primary reason we engineer features is to enhance the user experience of a product or service. By adding new features, we can make the product more intuitive, efficient, and user-friendly, which can increase

user satisfaction and engagement. **Competitive Advantage:** Another reason we engineer features is to gain a competitive advantage in the marketplace. By offering unique and innovative features, we can differentiate our product from

competitors and attract more customers. **Meet Customer Needs:** We engineer features to meet the evolving needs of customers. By analyzing user feedback, market trends, and customer behavior, we can identify areas where new features

could enhance the product's value and meet customer needs. **Increase Revenue:** Features can also be engineered to



generate more revenue. For example, a new feature that streamlines the checkout process can increase sales, or a feature

that provides additional functionality could lead to more upsells or cross-sells. Future-Proofing: Engineering features can also be done to future-proof a product or service. By anticipating future trends and potential customer needs, we can

develop features that ensure the product remains relevant and useful in the long term.

Processes Involved in Feature Engineering

Feature engineering in Machine learning consists of mainly 5 processes: Feature Creation, Feature Transformation, Feature Extraction, Feature Selection, and Feature Scaling. It is an iterative process that requires experimentation and

testing to find the best combination of features for a given problem. The success of a machine learning model largely depends on the quality of the features used in the model.

Feature Transformation

Feature Transformation is the process of transforming the features into a more suitable representation for the machine learning model. This is done to ensure that the model can effectively learn from the data.

Types of Feature Transformation:

Normalization: Rescaling the features to have a similar range, such as between 0 and 1, to prevent some features from dominating others.

Scaling: Rescaling the features to have a similar scale, such as having a standard deviation of 1, to make sure the model considers all features equally. Encoding: Transforming categorical features into a numerical representation. Examples

are one-hot encoding and label encoding.

Transformation: Transforming the features using mathematical operations to change the distribution or scale of the features. Examples are logarithmic, square root, and reciprocal transformations.

QUESTION 5

Which command manually triggers a single run of a scheduled task (either a standalone task or the root task in a DAG) independent of the schedule defined for the task?

- A. RUN TASK
- B. CALL TASK
- C. EXECUTE TASK
- D. RUN ROOT TASK

Correct Answer: C

Explanation: The EXECUTE TASK command manually triggers a single run of a scheduled task (either a standalone task or the root task in a DAG) independent of the schedule defined for the task. A successful run of a root task triggers a cascading run of child tasks in the DAG as their precedent task completes, as though the root task had run on its defined schedule. This SQL command is useful for testing new or modified standalone tasks and DAGs before you enable them to execute SQL code in production. Call this SQL command directly in scripts or in stored procedures. In



addition, this command supports integrating tasks in external data pipelines. Any third-party services that can authenticate into your Snowflake account and authorize SQL actions can execute the EXECUTE TASK command to run tasks.

QUESTION 6

Which of the following method is used for multiclass classification?

- A. one vs rest
- B. loocv
- C. all vs one
- D. one vs another

Correct Answer: A

Explanation: Binary vs. Multi-Class Classification Classification problems are common in machine learning. In most cases, developers prefer using a supervised machine-learning approach to predict class labels for a given dataset. Unlike regression, classification involves designing the classifier model and training it to input and categorize the test dataset. For that, you can divide the dataset into either binary or multi-class modules. As the name suggests, binary classification involves solving a problem with only two class labels. This makes it easy to filter the data, apply classification algorithms, and train the model to predict outcomes. On the other hand, multi-class classification is applicable when there are more than two class labels in the input train data. The technique enables developers to categorize the test data into multiple binary class labels. That said, while binary classification requires only one classifier model, the one used in the multi-class approach depends on the classification technique. Below are the two models of the multi-class classification algorithm. One-Vs-Rest Classification Model for Multi-Class Classification Also known as one-vs-all, the one-vs-rest model is a defined heuristic method that leverages a binary classification algorithm for multi-class classifications. The technique involves splitting a multi-class dataset into multiple sets of binary problems. Following this, a binary classifier is trained to handle each binary classification model with the most confident one making predictions. For instance, with a multi-class classification problem with red, green, and blue datasets, binary classification can be categorized as follows: Problem one: red vs. green/blue Problem two: blue vs. green/red Problem three: green vs. blue/red The only challenge of using this model is that you should create a model for every class. The three classes require three models from the above datasets, which can be challenging for large sets of data with million rows, slow models, such as neural networks and datasets with a significant number of classes. The one-vs-rest approach requires individual models to prognosticate the probability-like score. The class index with the largest score is then used to predict a class. As such, it is commonly used for classification algorithms that can naturally predict scores or numerical class membership such as perceptron and logistic regression.

QUESTION 7

Data providers add Snowflake objects (databases, schemas, tables, secure views, etc.) to a share using. Which of the following options? Choose 2.

- A. Grant privileges on objects to a share via Account role.
- B. Grant privileges on objects directly to a share.
- C. Grant privileges on objects to a share via a database role.
- D. Grant privileges on objects to a share via a third-party role.



Correct Answer: BC

Explanation: What is a Share?

Shares are named Snowflake objects that encapsulate all of the information required to share a database.

Data providers add Snowflake objects (databases, schemas, tables, secure views, etc.) to a share using either or both of the following options:

Option 1: Grant privileges on objects to a share via a database role. Option 2: Grant privileges on objects directly to a share. You choose which accounts can consume data from the share by adding the accounts to the share.

After a database is created (in a consumer account) from a share, all the shared objects are accessible to users in the consumer account. Shares are secure, configurable, and controlled completely by the provider account:

New objects added to a share become immediately available to all consumers, providing real-time access to shared data.

Access to a share (or any of the objects in a share) can be revoked at any time.

QUESTION 8

All aggregate functions except _____ ignore null values in their input collection

- A. Count(attribute)
- B. Count(*)
- C. Avg D. Sum

Correct Answer: B

Explanation: Count(*)

* is used to select all values including null.

QUESTION 9

As Data Scientist looking out to use Reader account, Which ones are the correct considerations about Reader Accounts for Third-Party Access?

- A. Reader accounts (formerly known as "read-only accounts") provide a quick, easy, and cost-effective way to share data without requiring the consumer to become a Snowflake customer.
- B. Each reader account belongs to the provider account that created it.
- C. Users in a reader account can query data that has been shared with the reader account, but cannot perform any of the DML tasks that are allowed in a full account, such as data loading, insert, update, and similar data manipulation operations.
- D. Data sharing is only possible between Snowflake accounts.

Correct Answer: D



Explanation:

Data sharing is only supported between Snowflake accounts. As a data provider, you might want to share data with a consumer who does not already have a Snowflake account or is not ready to become a licensed Snowflake customer. To

facilitate sharing data with these consumers, you can create reader accounts. Reader accounts (formerly known as "read-only accounts") provide a quick, easy, and cost-effective way to share data without requiring the consumer to become

a Snowflake customer.

Each reader account belongs to the provider account that created it. As a provider, you use shares to share databases with reader accounts; however, a reader account can only consume data from the provider account that created it. So,

Data Sharing is possible between Snowflake and Non-snowflake accounts via Reader Account.

QUESTION 10

Mark the Incorrect statements regarding MIN / MAX Functions?

- A. NULL values are skipped unless all the records are NULL
- B. NULL values are ignored unless all the records are NULL, in which case a NULL value is returned
- C. The data type of the returned value is the same as the data type of the input values
- D. For compatibility with other systems, the DISTINCT keyword can be specified as an argument for MIN or MAX, but it does not have any effect

Correct Answer: B

Explanation:

NULL values are ignored unless all the records are NULL, in which case a NULL value is returned

QUESTION 11

Which one is incorrect understanding about Providers of Direct share?

- A. A data provider is any Snowflake account that creates shares and makes them available to other Snowflake accounts to consume.
- B. As a data provider, you share a database with one or more Snowflake accounts.
- C. You can create as many shares as you want, and add as many accounts to a share as you want.
- D. If you want to provide a share to many accounts, you can do the same via Direct Share.

Correct Answer: D

Explanation:



If you want to provide a share to many accounts, you might want to use a listing or a data ex-change.

QUESTION 12

Which type of Machine learning Data Scientist generally used for solving classification and regression problems?

- A. Supervised
- B. Unsupervised
- C. Reinforcement Learning
- D. Instructor Learning
- E. Regression Learning

Correct Answer: A

Explanation:

Supervised Learning

Overview:

Supervised learning is a type of machine learning that uses labeled data to train machine learning models. In labeled data, the output is already known. The model just needs to map the inputs to the respective outputs.

Algorithms:

Some of the most popularly used supervised learning algorithms are:

Linear Regression

Logistic Regression

Support Vector Machine

K Nearest Neighbor

Decision Tree

Random Forest

Naive Bayes

Working:

Supervised learning algorithms take labelled inputs and map them to the known outputs, which means you already know the target variable. Supervised Learning methods need external supervision to train machine learning models. Hence,

the name supervised. They need guidance and additional information to return the desired result.

Applications:



Supervised learning algorithms are generally used for solving classification and regression problems.

Few of the top supervised learning applications are weather prediction, sales forecasting, stock price analysis.

QUESTION 13

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 method is generally used when a variable's distribution looks close to Gaussian. Z-score is the number of standard deviations a value of a variable is away from the variable's mean.

$$Z\text{-Score} = (X - \text{mean}) / \text{Standard deviation}$$

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

QUESTION 14

What Can Snowflake Data Scientist do in the Snowflake Marketplace as Provider? Choose all apply.

- A. Publish listings for free-to-use datasets to generate interest and new opportunities among the Snowflake customer base.
- B. Publish listings for datasets that can be customized for the consumer.
- C. Share live datasets securely and in real-time without creating copies of the data or imposing data integration tasks on the consumer.



D. Eliminate the costs of building and maintaining APIs and data pipelines to deliver data to customers.

Correct Answer: ABCD

Explanation:

All are correct!

About the Snowflake Marketplace

You can use the Snowflake Marketplace to discover and access third-party data and services, as well as market your own data products across the Snowflake Data Cloud. As a data provider, you can use listings on the Snowflake

Marketplace to share curated data offerings with many consumers simultaneously, rather than maintain sharing relationships with each individual consumer. With Paid Listings, you can also charge for your data products.

As a consumer, you might use the data provided on the Snowflake Marketplace to explore and access the following:

Historical data for research, forecasting, and machine learning. Up-to-date streaming data, such as current weather and traffic conditions. Specialized identity data for understanding subscribers and audience targets.

New insights from unexpected sources of data.

The Snowflake Marketplace is available globally to all non-VPS Snowflake accounts hosted on Amazon Web Services, Google Cloud Platform, and Microsoft Azure, with the exception of Microsoft Azure Government. Support for Microsoft

Azure Government is planned.

QUESTION 15

Secure Data Sharing do not let you share which of the following selected objects in a database in your account with other Snowflake accounts?

- A. Sequences
- B. Tables
- C. External tables
- D. Secure UDFs

Correct Answer: A

Explanation: Secure Data Sharing lets you share selected objects in a database in your account with other Snowflake accounts. You can share the following Snowflake database objects: Tables External tables Secure views Secure materialized views Secure UDFs Snowflake enables the sharing of databases through shares, which are created by data providers and "imported" by data consumers.

[Latest DSA-C02 Dumps](#)

[DSA-C02 PDF Dumps](#)

[DSA-C02 Braindumps](#)