Snowflake DSA-C02

Snowflake SnowPro Advanced - Data Scientist Certification Questions & Answers

Get Instant Access to Vital Exam Acing Materials | Study Guide | Sample Questions | Practice Test

DSA-C02
Snowflake Certified SnowPro Advanced - Data Scientist
65 Questions Exam – Duration of 115 minutes

Table of Contents:

Discover More about the DSA-C02 Certification

Are you interested in passing the Snowflake DSA-C02 exam? First discover, who benefits from the DSA-C02 certification. The DSA-C02 is suitable for a candidate if he wants to learn about Advance. Passing the DSA-C02 exam earns you the Snowflake Certified SnowPro Advanced - Data Scientist title.

While preparing for the DSA-C02 exam, many candidates struggle to get the necessary materials. But do not worry; your struggling days are over. The DSA-C02 PDF contains some of the most valuable preparation tips and the details and instant access to useful DSA-C02 study materials **just at one click.**

Snowflake DSA-C02 SnowPro Advanced - Data Scientist Certification Details:

Exam Name	Snowflake SnowPro Advanced - Data Scientist
Exam Code	DSA-C02
Exam Price	$375 USD
Duration	115 minutes
Number of Questions	65
Passing Score	750 + Scaled Scoring from 0 - 1000
Recommended Training / Books	**Data Science Training** **DSA-C02: SnowPro Advanced: Data Scientist Exam Study Guide**
Schedule Exam	**PEARSON VUE**
Sample Questions	**Snowflake DSA-C02 Sample Questions**
Recommended Practice	**Snowflake Certified SnowPro Advanced - Data Scientist Practice Test**

DSA-C02 Syllabus:

Section	Objectives	Weight
Data Science Concepts	- Define machine learning concepts for data science workloads. • Machine Learning - Supervised learning - Unsupervised learning	**15-20%**

Section	Objectives	Weight
	- Outline machine learning problem types. • Supervised Learning 1. Structured Data - Linear regression - Binary classification - Multi-class classification - Time-series forecasting 2. Unstructured Data - Image classification - Segmentation • Unsupervised Learning - Clustering - Association models - Summarize the machine learning lifecycle. • Data collection • Data visualization and exploration • Feature engineering • Training models • Model deployment • Model monitoring and evaluation (e.g., model explainability, precision, recall, accuracy, confusion matrix) • Model versioning - Define statistical concepts for data science. • Normal versus skewed distributions (e.g., mean, outliers) • Central limit theorem • Z and T tests • Bootstrapping • Confidence intervals	
Data Pipelining	- Enrich data by consuming data sharing sources. • Snowflake Marketplace • Direct Sharing • Shared database considerations - Build a data science pipeline. • Automation of data transformation with streams and tasks • Python User-Defined Functions (UDFs)	**15-20%**

Section	Objectives	Weight
	• Python User-Defined Table Functions (UDTFs) • Python stored procedures • Integration with machine learning platforms (e.g., connectors, ML partners, etc.)	
Data Preparation and Feature Engineering	- Prepare and clean data in Snowflake. • Use Snowpark for Python and SQL - Aggregate - Joins - Identify critical data - Remove duplicates - Remove irrelevant fields - Handle missing values - Data type casting - Sampling data - Perform exploratory data analysis in Snowflake. • Snowpark and SQL - Identify initial patterns (i.e., data profiling) - Connect external machine learning platforms and/or notebooks (e.g., Jupyter) • Use Snowflake native statistical functions to analyze and calculate descriptive data statistics. - Window Functions - MIN/MAX/AVG/STDEV - VARIANCE - TOPn - Approximation/High Performing function • Linear Regression - Find the slope and intercept - Verify the dependencies on dependent and independent variables - Perform feature engineering on Snowflake data. • Preprocessing - Scaling data - Encoding - Normalization • Data Transformations - Data Frames (i.e, Pandas, Snowpark) - Derived features (e.g., average spend) • Binarizing data - Binning continuous data into intervals - Label encoding - One hot encoding	30–35%

Section	Objectives	Weight
	- Visualize and interpret the data to present a business case. • Statistical summaries - Snowsight with SQL - Streamlit - Interpret open-source graph libraries - Identify data outliers • Common types of visualization formats - Bar charts - Scatterplots - Heat maps	
Model Development	- Connect data science tools directly to data in Snowflake. • Connecting Python to Snowflake - Snowpark - Python connector with Pandas support - Spark connector • Snowflake Best Practices - One platform, one copy of data, many workloads - Enrich datasets using the Snowflake Marketplace - External tables - External functions - Zero-copy cloning for training snapshots - Data governance - Train a data science model. • Hyperparameter tuning • Optimization metric selection (e.g., log loss, AUC, RMSE) • Partitioning - Cross validation - Train validation hold-out • Down/Up-sampling • Training with Python stored procedures • Training outside Snowflake through external functions • Training with Python User-Defined Table Functions (UDTFs) - Validate a data science model. • ROC curve/confusion matrix - Calculate the expected payout of the model • Regression problems	**15-20%**

Section	Objectives	Weight
	• Residuals plot - Interpret graphics with context • Model metrics - Interpret a model. • Feature impact • Partial dependence plots • Confidence intervals	

Model Deployment	- Move a data science model into production.
	
	• Use an external hosted model
	  - External functions
	  - Pre-built models
	• Deploy a model in Snowflake
	  - Vectorized/Scalar Python User Defined Functions (UDFs)
	  - Pre-built models
	  - Storing predictions
	  - Stage commands
	- Determine the effectiveness of a model and retrain if necessary.
	
	• Metrics for model evaluation
	  1. Data drift /Model decay
	  - Data distribution comparisons
	  -> Do the data making predictions look similar to the training data?
	  -> Do the same data points give the same predictions once a model is deployed?
	• Area under the curve
	• Accuracy, precision, recall
	• User defined functions (UDFs)
	- Outline model lifecycle and validation tools.
	
	• Streams and tasks
	• Metadata tagging
	• Model versioning with partner tools
	• Automation of model retraining
		15-20%

Broaden Your Knowledge with Snowflake DSA-C02 Sample Questions:

Question: 1

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

a) METADATA$ACTION
b) METADATA$FILE_ID
c) METADATA$ISUPDATE
d) METADATA$DELETE
e) METADATA$ROW_ID

Answer: a, c, e

Question: 2

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

a) Data sharing is only possible between Snowflake accounts.
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) 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.

Answer: a

Question: 3

In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?

a) no change
b) by intercept
c) by its slope
d) by 1

Answer: c

Question: 4

Performance metrics are a part of every machine learning pipeline, Which ones are not the performance metrics used in the Machine learning?

a) AU-ROC
b) Root Mean Squared Error (RMSE)
c) AUM
d) R (R-Squared)

Answer: c

Question: 5

You previously trained a model using a training dataset. You want to detect any data drift in the new data collected since the model was trained. What should you do?

a) Retrained your training dataset after correcting data outliers & no need to introduce new data.
b) Create a new dataset using the new data and a timestamp column and create a data drift monitor that uses the training dataset as a baseline and the new dataset as a target.
c) Create a new version of the dataset using only the new data and retrain the model.
d) Add the new data to the existing dataset and enable Application Insights for the service where the model is deployed.

Answer: b

Question: 6

How do you handle missing or corrupted data in a dataset?

a) Drop missing rows or columns
b) Replace missing values with mean/median/mode
c) Assign a unique category to missing values
d) All of the above

Answer: d

Question: 7

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

Answer: a

Question: 8

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

a) Unsupervised
b) Reinforcement Learning
c) Instructor Learning
d) Supervised
e) Regression Learning

Answer: d

Question: 9

Skewness of Normal distribution is _____.

a) Negative
b) 0
c) Positive
d) Undefined

Answer: b

Question: 10

Consider a data frame df with 10 rows and index ['r1', 'r2', 'r3', 'row4', 'row5', 'row6', 'r7', 'r8', 'r9', 'row10']. What does the expression g = df.groupby(df.index.str.len()) do?

a) Groups df based on index values
b) Groups df based on length of each index value
c) Groups df based on index strings
d) Data frames cannot be grouped by index values. Hence it results in Error.

Answer: d

Avail the Study Guide to Pass Snowflake DSA-C02 SnowPro Advanced - Data Scientist Exam:

- Find out about the DSA-C02 syllabus topics. Visiting the official site offers an idea about the exam structure and other important study resources. Going through the syllabus topics help to plan the exam in an organized manner.

- Once you are done exploring the **DSA-C02 syllabus**, it is time to plan for studying and covering the syllabus topics from the core. Chalk out the

best plan for yourself to cover each part of the syllabus in a hassle-free manner.

- A study schedule helps you to stay calm throughout your exam preparation. It should contain your materials and thoughts like study hours, number of topics for daily studying mentioned on it. The best bet to clear the exam is to follow your schedule rigorously.
- The candidate should not miss out on the scope to learn from the DSA-C02 training. Joining the Snowflake provided training for DSA-C02 exam helps a candidate to strengthen his practical knowledge base from the certification.
- Learning about the probable questions and gaining knowledge regarding the exam structure helps a lot. Go through the **DSA-C02 sample questions** and boost your knowledge
- Make yourself a pro through online practicing the syllabus topics. DSA-C02 practice tests would guide you on your strengths and weaknesses regarding the syllabus topics. Through rigorous practicing, you can improve the weaker sections too. Learn well about time management during exam and become confident gradually with practice tests.

Career Benefits:

- Passing the DSA-C02 exam, helps a candidate to prosper highly in his career. Having the certification on the resume adds to the candidate's benefit and helps to get the best opportunities.

Here Is the Trusted Practice Test for the DSA-C02 Certification

VMExam.Com is here with all the necessary details regarding the DSA-C02 exam. We provide authentic practice tests for the DSA-C02 exam. What do you gain from these practice tests? You get to experience the real exam-like questions made by industry experts and get a scope to improve your performance in the actual exam. Rely on VMExam.Com for rigorous, unlimited two-month attempts on the **DSA-C02 practice tests**, and gradually build your confidence. Rigorous practice made many aspirants successful and made their journey easy towards grabbing the Snowflake Certified SnowPro Advanced - Data Scientist.

Start Online practice of DSA-C02 Exam by visiting URL
https://www.vmexam.com/snowflake/dsa-c02-snowflake-snowpro-advanced-data-scientist