

Exam Databricks-Certified-Professional-Data-Engineer Preview | Reliable Databricks-Certified-Professional-Data-Engineer Exam Question



Purpose of this Exam Guide
The purpose of this exam guide is to give you an overview of the exam and what is covered on the exam to help you determine your exam readiness. This document will get updated anytime there are any changes to an exam (and when those changes will take effect on an exam) so that you can be prepared. This version covers the currently live version as of December 1, 2023. Please check back two weeks before you take your exam to make sure you have the most current version.

It is well known, to get the general respect of the community needs to be achieved by acquiring knowledge, and a harvest. Society will never welcome lazy people, and luck will never come to those who do not. We must continue to pursue own life value, such as get the test Databricks-Certified-Professional-Data-Engineer Certification, not only to meet what we have now, but also to constantly challenge and try something new and meaningful.

Databricks Certified Professional Data Engineer exam consists of multiple-choice questions and is conducted online. Databricks-Certified-Professional-Data-Engineer exam is intended to measure the candidate's proficiency in various areas, such as Spark architecture, Spark programming, data processing, data analysis, and data modeling. Databricks-Certified-Professional-Data-Engineer exam also tests the candidate's ability to optimize Spark performance and troubleshoot Spark applications. It is recommended that individuals who plan to take Databricks-Certified-Professional-Data-Engineer exam have at least two years of hands-on experience in big data technologies and Apache Spark.

The Databricks Databricks-Certified-Professional-Data-Engineer Exam consists of multiple-choice questions and hands-on tasks that test the candidate's practical knowledge of Databricks. Databricks-Certified-Professional-Data-Engineer exam covers a wide range of topics such as data engineering, data processing, ETL, data modeling, data warehousing, data governance, and data security. Databricks-Certified-Professional-Data-Engineer exam is designed to evaluate the candidate's ability to design and implement scalable data pipelines using Databricks.

>> Exam Databricks-Certified-Professional-Data-Engineer Preview <<

Free PDF 2026 Newest Databricks Exam Databricks-Certified-Professional-Data-Engineer Preview

Many candidates worry that after a long-time review of Databricks-Certified-Professional-Data-Engineer, they may still fail the exam due to inadaptation of the test model. So our PassTestking will provide a exam simulation for you to experience the real exam model before real exam. Databricks-Certified-Professional-Data-Engineer exam simulation software is full of questions, which will improve your ability to face the exam after you exercise them. Besides, the detailed answers analysis provided by our professionals will make you be more confidence to Pass Databricks-Certified-Professional-Data-Engineer Exam.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q189-Q194):

NEW QUESTION # 189

You are currently looking at a table that contains data from an e-commerce platform, each row contains a list of items(Item number) that were present in the cart, when the customer makes a change to the cart the entire information is saved as a separate list and appended to an existing list for the duration of the customer session, to identify all the items customer bought you have to make a unique list of items, you were asked to create a unique item's list that was added to the cart by the user, fill in the blanks of below query by choosing the appropriate higher-order function?

Note: See below sample data and expected output.

Schema: cartId INT, items Array<INT>

Fill in the blanks:

Fill in the blanks:

SELECT cartId, _(items) FROM carts

- A. ARRAY_DISTINCT, FLATTEN
- B. ARRAY_DISTINCT, ARRAY_UNION
- C. ARRAY_DISTINCT, ARRAY_FLATTEN
- D. ARRAY_UNION, ARRAY_DISTINCT
- E. FLATTEN, ARRAY_DISTINCT

Answer: A

Explanation:

Explanation

FLATTEN -> Transforms an array of arrays into a single array.

ARRAY_DISTINCT -> The function returns an array of the same type as the input argument where all duplicate values have been removed.

Table Description automatically generated

NEW QUESTION # 190

The data engineering team maintains the following code:

Assuming that this code produces logically correct results and the data in the source tables has been de-duplicated and validated, which statement describes what will occur when this code is executed?

- A. No computation will occur until enriched_itemized_orders_by_account is queried; upon query materialization, results will be calculated using the current valid version of data in each of the three tables referenced in the join logic.
- B. min("sale_date").alias("first_transaction_date"): Determines the earliest sale date for the customer.
- C. sum("sale_total").alias("lifetime_value"): Calculates the total sales amount (lifetime value) for the customer.
- D. max("sale_date").alias("last_transaction_date"): Determines the latest sale date for the customer.
- E. countDistinct("order_id").alias("total_orders"): Counts the number of unique orders placed by the customer.
- F. **The enriched_itemized_orders_by_account table will be overwritten using the current valid version of data in each of the three tables referenced in the join logic.**
- G. An incremental job will detect if new rows have been written to any of the source tables; if new rows are detected, all results will be recalculated and used to overwrite the enriched_itemized_orders_by_account table.
- H. An incremental job will leverage information in the state store to identify unjoined rows in the source tables and write these rows to the enriched_itemized_orders_by_account table.
- I. mean("sale_total").alias("average_sales"): Calculates the average sale amount for the customer.
- J. A batch job will update the enriched_itemized_orders_by_account table, replacing only those rows that have different values than the current version of the table, using accountID as the primary key.

Answer: F

Explanation:

Writes Data to gold_customer_lifetime_sales_summary Table:

The .write.mode("overwrite").table("gold_customer_lifetime_sales_summary") command writes the aggregated data to the gold_customer_lifetime_sales_summary table.

The mode("overwrite") specifies that the existing data in the gold_customer_lifetime_sales_summary table will be completely replaced by the new aggregated data.

Conclusion:

When this code is executed, it reads all records from the silver_customer_sales table, performs the specified aggregations grouped by customer_id, and then overwrites the entire gold_customer_lifetime_sales_summary table with the aggregated results. Therefore, option D accurately describes this process: "The gold_customer_lifetime_sales_summary table will be overwritten by aggregated values calculated from all records in the silver_customer_sales table as a batch job." Explanation:

The provided PySpark code performs the following operations:

Reads Data from silver_customer_sales Table:

The code starts by accessing the silver_customer_sales table using the spark.table method.

Groups Data by customer_id:

The .groupByKey("customer_id") function groups the data based on the customer_id column.

Aggregates Data:

The .agg() function computes several aggregate metrics for each customer_id:

Reference:

PySpark DataFrame groupBy

PySpark Basics

NEW QUESTION # 191

A Delta Lake table representing metadata about content posts from users has the following schema:

user_id LONG, post_text STRING, post_id STRING, longitude FLOAT, latitude FLOAT, post_time TIMESTAMP, date DATE

This table is partitioned by the date column. A query is run with the following filter:

longitude < 20 & longitude > -20

Which statement describes how data will be filtered?

- A. No file skipping will occur because the optimizer does not know the relationship between the partition column and the longitude.
- B. **Statistics in the Delta Log will be used to identify data files that might include records in the filtered range.**
- C. The Delta Engine will use row-level statistics in the transaction log to identify the files that meet the filter criteria.
- D. Statistics in the Delta Log will be used to identify partitions that might include files in the filtered range.
- E. The Delta Engine will scan the parquet file footers to identify each row that meets the filter criteria.

Answer: B

Explanation:

Explanation

This is the correct answer because it describes how data will be filtered when a query is run with the following filter: longitude < 20 & longitude > -20. The query is run on a Delta Lake table that has the following schema:

user_id LONG, post_text STRING, post_id STRING, longitude FLOAT, latitude FLOAT, post_time TIMESTAMP, date DATE

This table is partitioned by the date column. When a query is run on a partitioned Delta Lake table, Delta Lake uses statistics in the Delta Log to identify data files that might include records in the filtered range. The statistics include information such as min and max values for each column in each data file. By using these statistics, Delta Lake can skip reading data files that do not match the filter condition, which can improve query performance and reduce I/O costs. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Data skipping" section.

NEW QUESTION # 192

Which of the following commands can be used to query a delta table?

- A. 1.%sql
2.Select * from table_name
- B. 1.%python
2.execute.sql("select * from table")
- C. **Both A & B**
(Correct)
- D. 1.%python
2.delta.sql("select * from table")
- E. 1.%python
2.spark.sql("select * from table_name")

Answer: C

Explanation:

Explanation

The answer is both options A and B

Options C and D are incorrect because there is no command in Spark called execute.sql or delta.sql

NEW QUESTION # 193

Which of the following two options are supported in identifying the arrival of new files, and incremental data from Cloud object storage using Auto Loader?

- A. Checkpointing and Write ahead logging
- B. File hashing, Dynamic file lookup
- C. Directory listing, File notification
- D. Writing ahead logging, read head logging
- E. Checking pointing, watermarking

Answer: C

Explanation:

Explanation

The answer is A, Directory listing, File notifications

Directory listing: Auto Loader identifies new files by listing the input directory.

File notification: Auto Loader can automatically set up a notification service and queue service that subscribe to file events from the input directory.

Choosing between file notification and directory listing modes | Databricks on AWS

NEW QUESTION # 194

.....

With the development of the times, the pace of the society is getting faster and faster. If we don't try to improve our value, we're likely to be eliminated by society. Under the circumstances, we must find ways to prove our abilities. For example, getting the Databricks-Certified-Professional-Data-Engineer Certification is a good way. If we had it, the chances of getting a good job would be greatly improved. And our Databricks-Certified-Professional-Data-Engineer exam braindumps are the tool to help you get the Databricks-Certified-Professional-Data-Engineer certification.

Reliable Databricks-Certified-Professional-Data-Engineer Exam Question:

<https://www.passtestking.com/Databricks/Databricks-Certified-Professional-Data-Engineer-practice-exam-dumps.html>

- Study Databricks-Certified-Professional-Data-Engineer Center Testking Databricks-Certified-Professional-Data-Engineer Learning Materials Valid Real Databricks-Certified-Professional-Data-Engineer Exam Go to website (www.practicevce.com) open and search for ➡ Databricks-Certified-Professional-Data-Engineer to download for free Reliable Databricks-Certified-Professional-Data-Engineer Study Notes
- Pass Guaranteed Databricks - Valid Exam Databricks-Certified-Professional-Data-Engineer Preview Search for 《 Databricks-Certified-Professional-Data-Engineer 》 and easily obtain a free download on (www.pdfvce.com) Reliable Databricks-Certified-Professional-Data-Engineer Exam Tutorial
- Quiz Databricks - Databricks-Certified-Professional-Data-Engineer - Databricks Certified Professional Data Engineer Exam -Professional Exam Preview Search for 【 Databricks-Certified-Professional-Data-Engineer 】 and easily obtain a free download on [www.pdfdumps.com] Key Databricks-Certified-Professional-Data-Engineer Concepts
- Exam Databricks-Certified-Professional-Data-Engineer Preview 100% Pass | The Best Reliable Databricks Certified Professional Data Engineer Exam Exam Question Pass for sure Easily obtain 【 Databricks-Certified-Professional-Data-Engineer 】 for free download through 《 www.pdfvce.com 》 Valid Databricks-Certified-Professional-Data-Engineer Exam Papers
- Databricks-Certified-Professional-Data-Engineer Valid Study Materials Reliable Databricks-Certified-Professional-Data-Engineer Test Sample Databricks-Certified-Professional-Data-Engineer Exam Fees Search for [Databricks-Certified-Professional-Data-Engineer] on ➤ www.easy4engine.com immediately to obtain a free download Databricks-Certified-Professional-Data-Engineer Latest Learning Material
- Practice Databricks-Certified-Professional-Data-Engineer Engine Reliable Databricks-Certified-Professional-Data-Engineer Exam Tutorial Latest Databricks-Certified-Professional-Data-Engineer Learning Materials Easily obtain “ Databricks-Certified-Professional-Data-Engineer ” for free download through ➡ www.pdfvce.com Reliable Databricks-Certified-Professional-Data-Engineer Test Question
- Exam Databricks-Certified-Professional-Data-Engineer Preview - 2026 Databricks First-grade Reliable Databricks-Certified-Professional-Data-Engineer Exam Question Pass Guaranteed Simply search for ➡ Databricks-Certified-Professional-Data-Engineer for free download on “ www.validtorrent.com ” Valid Databricks-Certified-Professional-Data-Engineer Exam Papers
- Pass Guaranteed Databricks - Valid Exam Databricks-Certified-Professional-Data-Engineer Preview Open website ➡ www.pdfvce.com and search for ➡ Databricks-Certified-Professional-Data-Engineer for free download Valid Real Databricks-Certified-Professional-Data-Engineer Exam
- Databricks-Certified-Professional-Data-Engineer Cert Exam Databricks-Certified-Professional-Data-Engineer Torrent Reliable Databricks-Certified-Professional-Data-Engineer Test Question Search for ➡ Databricks-Certified-Professional-Data-Engineer and obtain a free download on ➡ www.torrentvce.com Study Databricks-

Certified-Professional-Data-Engineer Center