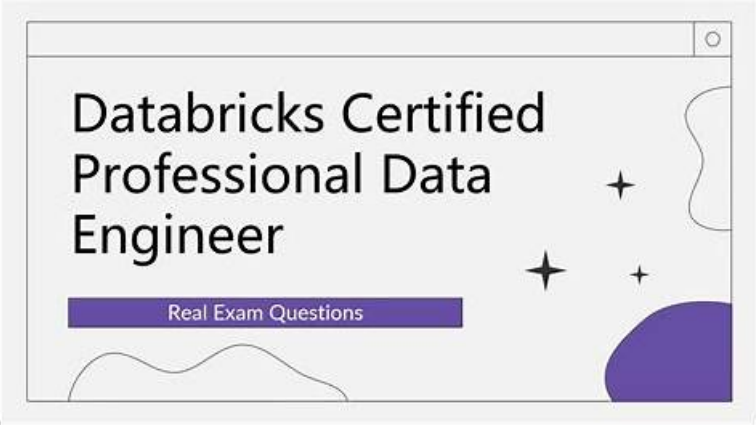
Real Databricks Databricks-Certified-Professional-Data-Engineer Exam Questions -The Greatest Shortcut Towards Success



Our website is the first choice among IT workers, especially the ones who are going to take Databricks-Certified-Professional-Data-Engineer certification exam in their first try. It is well known that getting certified by Databricks-Certified-Professional-Data-Engineer real exam is a guaranteed way to succeed with IT careers. We are here to provide you the high quality Databricks-Certified-Professional-Data-Engineer Braindumps Pdf for the preparation of the actual test and ensure you get maximum results with less effort.

The Databricks Databricks-Certified-Professional-Data-Engineer Exam covers a wide range of topics, including data architecture, data modeling, data integration, data processing, and data analytics. Databricks-Certified-Professional-Data-Engineer exam consists of both theoretical and practical components, which test the candidate's ability to apply their knowledge to real-world scenarios. The practical component requires candidates to complete a series of hands-on exercises using Databricks notebooks, which are used to build, test, and optimize data pipelines.

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

Accurate Exam Databricks-Certified-Professional-Data-Engineer Testking & Leading Offer in Qualification Exams & Free PDF Databricks-Certified-Professional-Data-Engineer: Databricks Certified Professional Data Engineer Exam

They work closely and check all Databricks Databricks-Certified-Professional-Data-Engineer exam practice test questions step by step and ensure the top standard of Databricks-Certified-Professional-Data-Engineer exam questions all the time. So rest assured that with the Databricks-Certified-Professional-Data-Engineer exam dumps you will get everything that you need to prepare and pass the Databricks Databricks-Certified-Professional-Data-Engineer Certification Exam with good scores. Countless Databricks Certified Professional Data Engineer Exam exam candidates have passed their Databricks-Certified-Professional-Data-Engineer exam and they all got help from real and updated Databricks Databricks-Certified-Professional-Data-Engineer exam questions. You can also be the next successful candidate for the Databricks-Certified-Professional-Data-Engineer certification exam.

Databricks-Certified-Professional-Data-Engineer exam consists of multiple-choice questions and hands-on, real-world scenarios that test the candidate's ability to design, build, and deploy data pipelines on Databricks. Databricks-Certified-Professional-Data-Engineer Exam covers various topics, including data engineering concepts, Databricks architecture, data processing using Spark, and data integration with other systems. Databricks Certified Professional Data Engineer Exam certification program provides a comprehensive learning experience that prepares candidates to become skilled data engineers and provides them with a competitive edge in the job market.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q139-Q144):

NEW QUESTION # 139
The data engineering team maintains the following code:

```
import pyspark.sql.functions as F

(spark.table("silver_customer_sales")
  .groupBy("customer_id")
  .agg(
    F.min("sale_date").alias("first_transaction_date"),
    F.max("sale_date").alias("last_transaction_date"),
    F.mean("sale_total").alias("average_sales"),
    F.countDistinct("order_id").alias("total_orders"),
    F.sum("sale_total").alias("lifetime_value")
  ).write
  .mode("overwrite")
  .table("gold_customer_lifetime_sales_summary")
)
```


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

- A. 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.
- B. An incremental job will detect if new rows have been written to the silver_customer_sales table; if new rows are detected, all aggregates will be recalculated and used to overwrite the gold_customer_lifetime_sales_summary table.
- C. An incremental job will leverage running information in the state store to update aggregate values in the gold_customer_lifetime_sales_summary table.
- D. The silver_customer_sales table will be overwritten by aggregated values calculated from all records in the gold_customer_lifetime_sales_summary table as a batch job.
- E. A batch job will update the gold_customer_lifetime_sales_summary table, replacing only those rows that have different values than the current version of the table, using customer_id as the primary key.

Answer: A

Explanation:
This code is using the pyspark.sql.functions library to group the silver_customer_sales table by customer_id and then aggregate the data using the minimum sale date, maximum sale total, and sum of distinct order ids. The resulting aggregated data is then written to the gold_customer_lifetime_sales_summary table, overwriting any existing data in that table. This is a batch job that does not use any incremental or streaming logic, and does not perform any merge or update operations. Therefore, the code will overwrite the gold table with the aggregated values from the silver table every time it is executed. Reference:
https://docs.databricks.com/spark/latest/dataframes-datasets/introduction-to-dataframes-python.html
https://docs.databricks.com/spark/latest/dataframes-datasets/transforming-data-with-dataframes.html
https://docs.databricks.com/spark/latest/dataframes-datasets/aggregating-data-with-dataframes.html

NEW QUESTION # 140
Which statement describes Delta Lake optimized writes?

- A. An asynchronous job runs after the write completes to detect if files could be further compacted; yes, an OPTIMIZE job is executed toward a default of 1 GB.
- B. Before a job cluster terminates, OPTIMIZE is executed on all tables modified during the most recent job.
- C. A shuffle occurs prior to writing to try to group data together resulting in fewer files instead of each executor writing multiple files based on directory partitions.

- D. Optimized writes logical partitions instead of directory partitions partition boundaries are only represented in metadata fewer small files are written.

Answer: C

Explanation:
Delta Lake optimized writes involve a shuffle operation before writing out data to the Delta table. The shuffle operation groups data by partition keys, which can lead to a reduction in the number of output files and potentially larger files, instead of multiple smaller files. This approach can significantly reduce the total number of files in the table, improve read performance by reducing the metadata overhead, and optimize the table storage layout, especially for workloads with many small files.
References:
* Databricks documentation on Delta Lake performance tuning:
https://docs.databricks.com/delta/optimizations/auto-optimize.html

NEW QUESTION # 141
Spill occurs as a result of executing various wide transformations. However, diagnosing spill requires one to proactively look for key indicators.
Where in the Spark UI are two of the primary indicators that a partition is spilling to disk?

- A. Driver's and Executor's log files
- B. Stage's detail screen and Executor's files
- C. Executor's detail screen and Executor's log files
- D. Stage's detail screen and Query's detail screen

Answer: D

Explanation:
In Apache Spark's UI, indicators of data spilling to disk during the execution of wide transformations can be found in the Stage's detail screen and the Query's detail screen. These screens provide detailed metrics about each stage of a Spark job, including information about memory usage and spill data. If a task is spilling data to disk, it indicates that the data being processed exceeds the available memory, causing Spark to spill data to disk to free up memory. This is an important performance metric as excessive spill can significantly slow down the processing.
Reference:
Apache Spark Monitoring and Instrumentation: Spark Monitoring Guide
Spark UI Explained: Spark UI Documentation

NEW QUESTION # 142
You are currently working on a production job failure with a job set up in job clusters due to a data issue, what cluster do you need to start to investigate and analyze the data?

- A. Databricks SQL Endpoint can be used to investigate the issue
- B. A Job cluster can be used to analyze the problem
- C. All-purpose cluster/ interactive cluster is the recommended way to run commands and view the data.
- D. Existing job cluster can be used to investigate the issue

Answer: C

Explanation:
Explanation
Answer is All-purpose cluster/ interactive cluster is the recommended way to run commands and view the data.
A job cluster can not provide a way for a user to interact with a notebook once the job is submitted, but an Interactive cluster allows to you display data, view visualizations write or edit quries, which makes it a perfect fit to investigate and analyze the data.

NEW QUESTION # 143
You are working on a marketing team request to identify customers with the same information between two tables CUSTOMERS_2021 and CUSTOMERS_2020 each table contains 25 columns with the same schema, You are looking to identify rows that match between two tables across all columns, which of the following can be used to perform in SQL

- A. 1.SELECT * FROM CUSTOMERS_2021
  2. INTERSECT
  3.SELECT * FROM CUSTOMERS_2020
- B. 1.SELECT * FROM CUSTOMERS_2021
  2.EXCEPT
  3.SELECT * FROM CUSTOMERS_2020
- C. 1.SELECT * FROM CUSTOMERS_2021
  2. UNION
  3.SELECT * FROM CUSTOMERS_2020
- D. 1.SELECT * FROM CUSTOMERS_2021 C1
  2.INNER JOIN CUSTOMERS_2020 C2
  3.ON C1.CUSTOMER_ID = C2.CUSTOMER_ID
- E. 1.SELECT * FROM CUSTOMERS_2021
  2. UNION ALL
  3.SELECT * FROM CUSTOMERS_2020

Answer: A

Explanation:
Explanation
Answer is,
1.SELECT * FROM CUSTOMERS_2021
2. INTERSECT
3.SELECT * FROM CUSTOMERS_2020
To compare all the rows between both the tables across all the columns using intersect will help us achieve that, an inner join is only going to check if the same column value exists across both the tables on a single column.
INTERSECT [ALL | DISTINCT]
*Returns the set of rows which are in both subqueries.
If ALL is specified a row that appears multiple times in the subquery1 as well as in subquery will be returned multiple times.
If DISTINCT is specified the result does not contain duplicate rows. This is the default.

NEW QUESTION # 144
.....

- Databricks-Certified-Professional-Data-Engineer Exam Brain Dumps □ Reliable Databricks-Certified-Professional-Data-Engineer Study Guide □ Reliable Databricks-Certified-Professional-Data-Engineer Exam Testking □ Immediately open ⇛ www.vceengine.com ⇚ and search for ➽ Databricks-Certified-Professional-Data-Engineer □ to obtain a free download □Databricks-Certified-Professional-Data-Engineer Training Material
- Databricks-Certified-Professional-Data-Engineer Practice Mock □ Databricks-Certified-Professional-Data-Engineer Practice Mock □ New Databricks-Certified-Professional-Data-Engineer Exam Format ⧖ Search for ➡ Databricks-Certified-Professional-Data-Engineer □□□ on 【 www.pdfvce.com 】 immediately to obtain a free download □ □Databricks-Certified-Professional-Data-Engineer Exam Revision Plan
- Databricks Databricks-Certified-Professional-Data-Engineer Exam Questions Are Out - Download And Prepare [2026] □ Easily obtain ➥ Databricks-Certified-Professional-Data-Engineer □ for free download through ☀ www.prep4sures.top □☀□ □Reliable Databricks-Certified-Professional-Data-Engineer Study Guide
- Databricks-Certified-Professional-Data-Engineer Exam Brain Dumps □ Databricks-Certified-Professional-Data-Engineer Braindumps Torrent □ Real Databricks-Certified-Professional-Data-Engineer Exam Answers □ Easily obtain free download of ➤ Databricks-Certified-Professional-Data-Engineer □ by searching on 「 www.pdfvce.com 」 □ □Databricks-Certified-Professional-Data-Engineer PDF
- Databricks-Certified-Professional-Data-Engineer New Guide Files □ Pass4sure Databricks-Certified-Professional-Data-Engineer Study Materials □ Reliable Databricks-Certified-Professional-Data-Engineer Test Forum □ Easily obtain free download of □ Databricks-Certified-Professional-Data-Engineer □ by searching on { www.dumpsquestion.com } □ □Pass4sure Databricks-Certified-Professional-Data-Engineer Study Materials
- www.stes.tyc.edu.tw, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, estar.jp, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, Disposable vapes