

Valid Databricks-Certified-Professional-Data-Engineer test answers & Databricks Databricks-Certified-Professional-Data-Engineer pass test & Databricks-Certified-Professional-Data-Engineer lead2pass review

```
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")
)
```

Are you planning to crack the Databricks Databricks Certified Professional Data Engineer Exam Databricks-Certified-Professional-Data-Engineer certification test in a short time and don't know how to prepare for it? PrepAwayTest has updated Databricks-Certified-Professional-Data-Engineer Dumps questions for the applicants who want to prepare for the Databricks Databricks-Certified-Professional-Data-Engineer Certification test successfully within a few days. This study material is available in three different formats that you can trust to crack the Databricks Databricks-Certified-Professional-Data-Engineer certification test with ease.

Every detail of our Databricks-Certified-Professional-Data-Engineer exam guide is going through professional evaluation and test. Other workers are also dedicated to their jobs. Even the proofreading works of the Databricks-Certified-Professional-Data-Engineer study materials are complex and difficult. They still attentively accomplish their tasks. Please have a try and give us an opportunity. Our Databricks-Certified-Professional-Data-Engineer Preparation guide will totally amaze you and bring you good luck. And it deserves you to have a try!

>> **Databricks-Certified-Professional-Data-Engineer Valid Exam Camp Pdf** <<

Databricks Certified Professional Data Engineer Exam Exam Practice Dump Provide Best Databricks-Certified-Professional-Data-Engineer Study Questions

Of course, when you are seeking for exam materials, it is certain that you will find many different materials. However, through investigation or personal experience, you will find PrepAwayTest questions and answers are the best ones for your need. The candidates have not enough time to prepare the exam, while PrepAwayTest certification training materials are to develop to solve the problem. So, it can save much time for us. What's more important, 100% guarantee to pass Databricks Databricks-Certified-Professional-Data-Engineer Exam at the first attempt. In addition, PrepAwayTest exam dumps will be updated at any time. If exam outline and the content change, PrepAwayTest can provide you with the latest information.

Databricks Certified Professional Data Engineer (Databricks-Certified-Professional-Data-Engineer) Certification Exam is a highly respected credential within the data engineering industry. Databricks Certified Professional Data Engineer Exam certification is specifically designed for professionals who have a deep understanding of data engineering principles, practices, and technologies. With this certification, data engineers can demonstrate their expertise in designing and building data pipelines, managing data workflows, and implementing data analytics solutions using Databricks.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q156-Q161):

NEW QUESTION # 156

Which of the following techniques structured streaming uses to ensure recovery of failures during stream processing?

- A. Checkpointing and Watermarking
- B. Delta time travel
- C. Write ahead logging and watermarking
- D. The stream will failover to available nodes in the cluster
- **E. Checkpointing and write-ahead logging**
- F. Checkpointing and Idempotent sinks

Answer: E

Explanation:

Explanation

The answer is Checkpointing and write-ahead logging.

Structured Streaming uses checkpointing and write-ahead logs to record the offset range of data being processed during each trigger interval.

NEW QUESTION # 157

A Databricks SQL dashboard has been configured to monitor the total number of records present in a collection of Delta Lake tables using the following query pattern:

SELECT COUNT (*) FROM table -

Which of the following describes how results are generated each time the dashboard is updated?

- A. The total count of rows is calculated by scanning all data files
- **B. The total count of records is calculated from the Delta transaction logs**
- C. The total count of rows will be returned from cached results unless REFRESH is run
- D. The total count of records is calculated from the Hive metastore
- E. The total count of records is calculated from the parquet file metadata

Answer: B

Explanation:

<https://delta.io/blog/2023-04-19-faster-aggregations-metadata/#:~:text=You%20can%20get%20the%20number,a%20given%20Delta%20table%20version.>

NEW QUESTION # 158

Which of the below SQL commands creates a session scoped temporary view?

- **A. 1.CREATE OR REPLACE TEMPORARY VIEW view_name
2.AS SELECT * FROM table_name
(Correct)**
- B. 1.CREATE OR REPLACE GLOBAL TEMPORARY VIEW view_name
2.AS SELECT * FROM table_name
- C. 1.CREATE OR REPLACE LOCAL TEMPORARY VIEW view_name
2.AS SELECT * FROM table_name
- D. 1.CREATE OR REPLACE LOCAL VIEW view_name
2.AS SELECT * FROM table_name
- E. 1.CREATE OR REPLACE VIEW view_name
2.AS SELECT * FROM table_name

Answer: A

Explanation:

Explanation

The answer is

1.CREATE OR REPLACE TEMPORARY VIEW view_name

2.AS SELECT * FROM table_name

The default temporary view is session scoped, as soon as the session ends or if a notebook is de-tached session scoped temporary

view is dropped.

NEW QUESTION # 159

A junior data engineer on your team has implemented the following code block.

The view `new_events` contains a batch of records with the same schema as the `eventsDelta` table.

The `event_id` field serves as a unique key for this table.

When this query is executed, what will happen with new records that have the same `event_id` as an existing record?

- A. They are inserted.
- **B. They are ignored.**
- C. They are updated.
- D. They are deleted.
- E. They are merged.

Answer: B

Explanation:

This is the correct answer because it describes what will happen with new records that have the same `event_id` as an existing record when the query is executed. The query uses the `INSERT INTO` command to append new records from the view `new_events` to the table `events`. However, the `INSERT INTO` command does not check for duplicate values in the primary key column (`event_id`) and does not perform any update or delete operations on existing records. Therefore, if there are new records that have the same `event_id` as an existing record, they will be ignored and not inserted into the table `events`. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Append data using `INSERT INTO`" section.

"If none of the `WHEN MATCHED` conditions evaluate to true for a source and target row pair that matches the `merge_condition`, then the target row is left unchanged." <https://docs.databricks.com/en/sql/language-manual/delta-merge-into.html#:~:text=If%20none%20>

NEW QUESTION # 160

A junior data engineer has been asked to develop a streaming data pipeline with a grouped aggregation using `DataFrame df`. The pipeline needs to calculate the average humidity and average temperature for each non-overlapping five-minute interval. Incremental state information should be maintained for 10 minutes for late-arriving data.

Streaming `DataFrame df` has the following schema:

`"device_id INT, event_time TIMESTAMP, temp FLOAT, humidity FLOAT"`

Code block:

Choose the response that correctly fills in the blank within the code block to complete this task.

- A. `delayWrite("event_time", "10 minutes")`
- B. `await("event_time + '10 minutes'")`
- **C. `withWatermark("event_time", "10 minutes")`**
- D. `slidingWindow("event_time", "10 minutes")`
- E. `awaitArrival("event_time", "10 minutes")`

Answer: C

Explanation:

The correct answer is A. `withWatermark("event_time", "10 minutes")`. This is because the question asks for incremental state information to be maintained for 10 minutes for late-arriving data. The `withWatermark` method is used to define the watermark for late data. The watermark is a timestamp column and a threshold that tells the system how long to wait for late data. In this case, the watermark is set to 10 minutes. The other options are incorrect because they are not valid methods or syntax for watermarking in Structured Streaming. References:

* Watermarking: <https://docs.databricks.com/spark/latest/structured-streaming/watermarks.html>

* Windowed aggregations: <https://docs.databricks.com/spark/latest/structured-streaming/window-operations.html>

NEW QUESTION # 161

.....

[illegible]