

Prepare for the Databricks Databricks-Certified-Data-Engineer-Associate Exam with PDF4Test Verified Pdf Questions



P.S. Free 2025 Databricks Databricks-Certified-Data-Engineer-Associate dumps are available on Google Drive shared by PDF4Test: <https://drive.google.com/open?id=1iZ3athW4Byuy5OLZJZH7DZq8JVB4Di7Z>

We have created a number of reports and learning functions for evaluating your proficiency for the Databricks Certified Data Engineer Associate Exam (Databricks-Certified-Data-Engineer-Associate) exam dumps. In preparation, you can optimize Databricks Certified Data Engineer Associate Exam (Databricks-Certified-Data-Engineer-Associate) practice exam time and question type by utilizing our Databricks Databricks-Certified-Data-Engineer-Associate Practice Test software. PDF4Test makes it easy to download Databricks Certified Data Engineer Associate Exam (Databricks-Certified-Data-Engineer-Associate) exam questions immediately after purchase.

Databricks Certified Data Engineer Associate certification exam covers topics such as data engineering concepts, data ingestion, data processing, data storage, and data transformation using Apache Spark and Delta Lake. Candidates who Pass Databricks-Certified-Data-Engineer-Associate Exam will have a deep understanding of the Databricks platform and will be able to design, build, and maintain data pipelines that are scalable, reliable, and efficient. Databricks Certified Data Engineer Associate Exam certification is ideal for data engineers, data analysts, and data scientists who work with big data and want to enhance their skills and advance their careers.

>> **Databricks-Certified-Data-Engineer-Associate Latest Exam Papers** <<

Databricks-Certified-Data-Engineer-Associate Real Exam Questions, Detailed Databricks-Certified-Data-Engineer-Associate Study Dumps

To find the perfect Databricks Certified Data Engineer Associate Exam Databricks-Certified-Data-Engineer-Associate practice materials for the exam, you search and re-search without reaching the final decision and compare advantages and disadvantages with materials in the market. With systemic and methodological content within our Databricks-Certified-Data-Engineer-Associate practice materials, they have helped more than 98 percent of exam candidates who chose our Databricks-Certified-Data-Engineer-Associate guide exam before getting the final certificates successfully.

Databricks Certified Data Engineer Associate Exam Sample Questions (Q97-Q102):

NEW QUESTION # 97

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table.

The code block used by the data engineer is below:

```
(spark.table("sales")
  .withColumn("avg_price", col("sales") / col("units"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("complete")
  .trigger(          )
  .table("new_sales")
)
```

If the data engineer only wants the query to execute a micro-batch to process data every 5 seconds, which of the following lines of code should the data engineer use to fill in the blank?

- A. trigger(once="5 seconds")
- B. trigger(processingTime="5 seconds")
- C. trigger("5 seconds")
- D. trigger()
- E. trigger(continuous="5 seconds")

Answer: B

Explanation:

The processingTime option specifies a time-based trigger interval for fixed interval micro-batches. This means that the query will execute a micro-batch to process data every 5 seconds, regardless of how much data is available. This option is suitable for near-real time processing workloads that require low latency and consistent processing frequency. The other options are either invalid syntax (A, C), default behavior (B), or experimental feature (E). Reference: Databricks Documentation - Configure Structured Streaming trigger intervals, Databricks Documentation - Trigger.

NEW QUESTION # 98

A data engineer wants to create a new table containing the names of customers that live in France. They have written the following command:

```
CREATE TABLE customersInFrance
  AS
  SELECT id,
         firstName,
         lastName,
  FROM customerLocations
  WHERE country = 'FRANCE';
```

A senior data engineer mentions that it is organization policy to include a table property indicating that the new table includes personally identifiable information (PII).

Which of the following lines of code fills in the above blank to successfully complete the task?

- A. PII
- B. TBLPROPERTIES PII
- C. COMMENT "Contains PII"
- D. "COMMENT PII"
- E. There is no way to indicate whether a table contains PII.

Answer: C

Explanation:

In Databricks, when creating a table, you can add a comment to columns or the entire table to provide more information about the data it contains. In this case, since it's organization policy to indicate that the new table includes personally identifiable information (PII), option D is correct. The line of code would be added after defining the table structure and before closing with a semicolon. References: Data Engineer Associate Exam Guide, CREATE TABLE USING (Databricks SQL)

NEW QUESTION # 99

A Delta Live Table pipeline includes two datasets defined using streaming live table. Three datasets are defined against Delta Lake table sources using live table.

The table is configured to run in Production mode using the Continuous Pipeline Mode.

What is the expected outcome after clicking Start to update the pipeline assuming previously unprocessed data exists and all definitions are valid?

- A. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will persist to allow for additional testing.
- B. All datasets will be updated once and the pipeline will shut down. The compute resources will persist to allow for additional testing.
- C. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will be deployed for the update and terminated when the pipeline is stopped.
- D. All datasets will be updated once and the pipeline will shut down. The compute resources will be terminated.

Answer: C

Explanation:

In Delta Live Tables (DLT), when configured to run in Continuous Pipeline Mode, particularly in a production environment, the system is designed to continuously process and update data as it becomes available. This mode keeps the compute resources active to handle ongoing data processing and automatically updates all datasets defined in the pipeline at predefined intervals. Once the pipeline is manually stopped, the compute resources are terminated to conserve resources and reduce costs. This mode is suitable for production environments where datasets need to be kept up-to-date with the latest data.

Reference:

Databricks documentation on Delta Live Tables: Delta Live Tables Guide

NEW QUESTION # 100

A data engineer needs to apply custom logic to identify employees with more than 5 years of experience in array column employees in table stores. The custom logic should create a new column exp_employees that is an array of all of the employees with more than 5 years of experience for each row. In order to apply this custom logic at scale, the data engineer wants to use the FILTER higher-order function.

Which of the following code blocks successfully completes this task?

```

SELECT
    store_id,
A.    employees,
    FILTER (employees, i -> i.years_exp > 5) AS exp_employees
FROM stores;

SELECT
    store_id,
B.    employees,
    FILTER (exp_employees, years_exp > 5) AS exp_employees
FROM stores;

SELECT
    store_id,
C.    employees,
    FILTER (employees, years_exp > 5) AS exp_employees
FROM stores;

SELECT
    store_id,
    employees,
D.    CASE WHEN employees.years_exp > 5 THEN employees
        ELSE NULL
    END AS exp_employees
FROM stores;

SELECT
    store_id,
E.    employees,
    FILTER (exp_employees, i -> i.years_exp > 5) AS exp_employees
FROM stores;

```

- A. Option B
- B. Option D
- C. Option E
- D. Option C
- E. Option A

Answer: E

NEW QUESTION # 101

In order for Structured Streaming to reliably track the exact progress of the processing so that it can handle any kind of failure by restarting and/or reprocessing, which of the following two approaches is used by Spark to record the offset range of the data being processed in each trigger?

- A. Write-ahead Logs and Idempotent Sinks
- B. Replayable Sources and Idempotent Sinks
- C. Structured Streaming cannot record the offset range of the data being processed in each trigger.
- D. Checkpointing and Write-ahead Logs
- E. Checkpointing and Idempotent Sinks

Answer: D

Explanation:

Explanation

The engine uses checkpointing and write-ahead logs to record the offset range of the data being processed in each trigger. -- in the link search for "The engine uses " you'll find the answer. <https://spark.apache.org/docs/latest/structured-streaming-programming->

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, pct.edu.pk, www.stes.tyc.edu.tw, 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, www.stes.tyc.edu.tw, Disposable vapes

P.S. Free & New Databricks-Certified-Data-Engineer-Associate dumps are available on Google Drive shared by PDF4Test:
<https://drive.google.com/open?id=1iZ3athW4Byuy5OLZJZH7DZq8JVB4Di7Z>