

Databricks-Certified-Professional-Data-Engineer Test Certification Cost | Vce Databricks-Certified-Professional-Data-Engineer Test Simulator



Moreover, you do not need an active internet connection to utilize PrepAwayTest Databricks Databricks-Certified-Professional-Data-Engineer practice exam software. It works without the internet after software installation on Windows computers. The PrepAwayTest web-based Databricks Databricks-Certified-Professional-Data-Engineer Practice Test requires an active internet and it is compatible with all operating systems. You can conveniently test your performance by checking your score each time you use our Databricks Databricks-Certified-Professional-Data-Engineer practice exam software.

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.

>> **Databricks-Certified-Professional-Data-Engineer Test Certification Cost** <<

Quiz 2026 Pass-Sure Databricks-Certified-Professional-Data-Engineer: Databricks Certified Professional Data Engineer Exam Test Certification Cost

We would like to benefit our customers from different countries who decide to choose our Databricks-Certified-Professional-Data-Engineer study guide in the long run, so we cooperation with the leading experts in the field to renew and update our Databricks-Certified-Professional-Data-Engineer learning materials. Our leading experts aim to provide you the newest information in this field in order to help you to keep pace with the times and fill your knowledge gap. As long as you bought our Databricks-Certified-Professional-Data-Engineer Practice Engine, you are bound to pass the Databricks-Certified-Professional-Data-Engineer exam for sure.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q187-Q192):

NEW QUESTION # 187

A data engineer is designing a data pipeline. The source system generates files in a shared directory that is also used by other processes. As a result, the files should be kept as is and will accumulate in the directory. The data engineer needs to identify which files are new since the previous run in the pipeline, and set up the pipeline to only ingest those new files with each run.

Which of the following tools can the data engineer use to solve this problem?

- A. Unity Catalog
- B. Auto Loader

- C. Data Explorer
- D. Delta Lake
- E. Databricks SQL

Answer: B

NEW QUESTION # 188

A data architect has designed a system in which two Structured Streaming jobs will concurrently write to a single bronze Delta table. Each job is subscribing to a different topic from an Apache Kafka source, but they will write data with the same schema. To keep the directory structure simple, a data engineer has decided to nest a checkpoint directory to be shared by both streams.

The proposed directory structure is displayed below:

Which statement describes whether this checkpoint directory structure is valid for the given scenario and why?

- **A. No; each of the streams needs to have its own checkpoint directory.**
- B. No; only one stream can write to a Delta Lake table.
- C. Yes; both of the streams can share a single checkpoint directory.
- D. Yes; Delta Lake supports infinite concurrent writers.
- E. No; Delta Lake manages streaming checkpoints in the transaction log.

Answer: A

Explanation:

This is the correct answer because checkpointing is a critical feature of Structured Streaming that provides fault tolerance and recovery in case of failures. Checkpointing stores the current state and progress of a streaming query in a reliable storage system, such as DBFS or S3. Each streaming query must have its own checkpoint directory that is unique and exclusive to that query. If two streaming queries share the same checkpoint directory, they will interfere with each other and cause unexpected errors or data loss. Verified Reference: [Databricks Certified Data Engineer Professional], under "Structured Streaming" section; Databricks Documentation, under "Checkpointing" section.

NEW QUESTION # 189

The business intelligence team has a dashboard configured to track various summary metrics for retail stores.

This includes total sales for the previous day alongside totals and averages for a variety of time periods. The fields required to populate this dashboard have the following schema:

For Demand forecasting, the Lakehouse contains a validated table of all itemized sales updated incrementally in near real-time. This table named `products_per_order`, includes the following fields:

Because reporting on long-term sales trends is less volatile, analysts using the new dashboard only require data to be refreshed once daily. Because the dashboard will be queried interactively by many users throughout a normal business day, it should return results quickly and reduce total compute associated with each materialization.

Which solution meets the expectations of the end users while controlling and limiting possible costs?

- A. Populate the dashboard by configuring a nightly batch job to save the required to quickly update the dashboard with each query.
- B. Use the Delta Cache to persist the `products_per_order` table in memory to quickly the dashboard with each query.
- **C. Define a view against the `products_per_order` table and define the dashboard against this view.**
- D. Use Structure Streaming to configure a live dashboard against the `products_per_order` table within a Databricks notebook.

Answer: C

Explanation:

Given the requirement for daily refresh of data and the need to ensure quick response times for interactive queries while controlling costs, a nightly batch job to pre-compute and save the required summary metrics is the most suitable approach.

* By pre-aggregating data during off-peak hours, the dashboard can serve queries quickly without requiring on-the-fly computation, which can be resource-intensive and slow, especially with many users.

* This approach also limits the cost by avoiding continuous computation throughout the day and instead leverages a batch process that efficiently computes and stores the necessary data.

* The other options (A, C, D) either do not address the cost and performance requirements effectively or

* are not suitable for the use case of less frequent data refresh and high interactivity.

References:

* Databricks Documentation on Batch Processing: Databricks Batch Processing

* Data Lakehouse Patterns: Data Lakehouse Best Practices

NEW QUESTION # 190

Your colleague was walking you through how a job was setup, but you noticed a warning message that said, "Jobs running on all-purpose cluster are considered all purpose compute", the colleague was not sure why he was getting the warning message, how do you best explain this warning message?

- A. All-purpose clusters take longer to start the cluster vs a job cluster
- B. All-purpose clusters are less expensive than the job clusters
- **C. All-purpose clusters are more expensive than the job clusters**
- D. All-purpose cluster provide interactive messages that can not be viewed in a job
- E. All-purpose clusters cannot be used for Job clusters, due to performance issues.

Answer: C

Explanation:

Explanation

Warning message:

Graphical user interface, text, application, email Description automatically generated

□ Pricing for All-purpose clusters are more expensive than the job clusters AWS pricing(Aug 15th 2022)Graphical user interface

Description automatically generated

□ Bottom of Form

Top of Form

NEW QUESTION # 191

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

Answer: D

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 # 192

.....

You can download the Databricks Certified Professional Data Engineer Exam Databricks-Certified-Professional-Data-Engineer product right after purchasing and start your journey toward your big career. The Databricks Databricks-Certified-Professional-Data-Engineer exam questions are very similar to actual Databricks Databricks-Certified-Professional-Data-Engineer Exam

