

Databricks-Certified-Professional-Data-Engineer Test Score Report, Reliable Databricks-Certified-Professional-Data-Engineer Exam Braindumps



After you pass the test Databricks-Certified-Professional-Data-Engineer certification, your working abilities will be recognized by the society and you will find a good job. If you master our Databricks-Certified-Professional-Data-Engineer quiz torrent and pass the exam. You will be respected by your colleagues, your boss, your relatives, your friends and the society. All in all, buying our Databricks-Certified-Professional-Data-Engineer Test Prep can not only help you pass the exam but also help realize your dream about your career and your future. So don't be hesitated to buy our Databricks-Certified-Professional-Data-Engineer exam materials and take action immediately.

Databricks Certified Professional Data Engineer exam is designed to test a candidate's knowledge and skills in building, designing, and managing data pipelines on the Databricks platform. Databricks-Certified-Professional-Data-Engineer Exam covers a range of topics, including data processing, data storage, data warehousing, data modeling, and data architecture. Candidates are expected to have a deep understanding of these topics and be able to apply them in real-world scenarios.

>> **Databricks-Certified-Professional-Data-Engineer Test Score Report** <<

Free PDF Quiz Databricks-Certified-Professional-Data-Engineer - Databricks Certified Professional Data Engineer Exam –High-quality Test Score Report

When you choose Databricks-Certified-Professional-Data-Engineer valid study pdf, you will get a chance to participate in the simulated exam before you take your actual test. The contents of Databricks-Certified-Professional-Data-Engineer exam torrent are compiled by our experts through several times of verification and confirmation. So the Databricks-Certified-Professional-Data-Engineer questions & answers are valid and reliable to use. You can find all the key points in the Databricks-Certified-Professional-Data-Engineer practice torrent. Besides, the Databricks-Certified-Professional-Data-Engineer test engine training equipped with various self-assessment functions like exam history, result scores and time setting, etc.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q150-Q155):

NEW QUESTION # 150

A data architect has heard about lake's built-in versioning and time travel capabilities. For auditing purposes they have a requirement to maintain a full of all valid street addresses as they appear in the customers table.

The architect is interested in implementing a Type 1 table, overwriting existing records with new values and relying on Delta Lake time travel to support long-term auditing. A data engineer on the project feels that a Type 2 table will provide better performance and scalability.

Which piece of information is critical to this decision?

- A. Data corruption can occur if a query fails in a partially completed state because Type 2 tables requires Setting multiple

fields in a single update.

- B. Delta Lake time travel cannot be used to query previous versions of these tables because Type 1 changes modify data files in place.
- C. Delta Lake time travel does not scale well in cost or latency to provide a long-term versioning solution.
- D. Shallow clones can be combined with Type 1 tables to accelerate historic queries for long-term versioning.

Answer: C

Explanation:

Delta Lake's time travel feature allows users to access previous versions of a table, providing a powerful tool for auditing and versioning. However, using time travel as a long-term versioning solution for auditing purposes can be less optimal in terms of cost and performance, especially as the volume of data and the number of versions grow. For maintaining a full history of valid street addresses as they appear in a customers table, using a Type 2 table (where each update creates a new record with versioning) might provide better scalability and performance by avoiding the overhead associated with accessing older versions of a large table. While Type 1 tables, where existing records are overwritten with new values, seem simpler and can leverage time travel for auditing, the critical piece of information is that time travel might not scale well in cost or latency for long-term versioning needs, making a Type 2 approach more viable for performance and scalability.

:

Databricks Documentation on Delta Lake's Time Travel: Delta Lake Time Travel Databricks Blog on Managing Slowly Changing Dimensions in Delta Lake: Managing SCDs in Delta Lake

NEW QUESTION # 151

A data engineering team has been using a Databricks SQL query to monitor the performance of an ELT job. The ELT job is triggered by a specific number of input records being ready to process. The Databricks SQL query returns the number of minutes since the job's most recent runtime.

Which of the following approaches can enable the data engineering team to be notified if the ELT job has not been run in an hour?

- A. They can set up an Alert for the query to notify them if the returned value is greater than 60
- B. They can set up an Alert for the query to notify when the ELT job fails
- C. This type of alerting is not possible in Databricks
- D. They can set up an Alert for the accompanying dashboard to notify them if the returned value is greater than 60
- E. They can set up an Alert for the accompanying dashboard to notify when it has not re-freshed in 60 minutes

Answer: A

NEW QUESTION # 152

When scheduling Structured Streaming jobs for production, which configuration automatically recovers from query failures and keeps costs low?

- A. Cluster: New Job Cluster;
Retries: None;
Maximum Concurrent Runs: 1
- B. Cluster: New Job Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: Unlimited
- C. Cluster: Existing All-Purpose Cluster;
Retries: None;
Maximum Concurrent Runs: 1
- D. Cluster: Existing All-Purpose Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: 1
- E. Cluster: Existing All-Purpose Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: 1

Answer: A

Explanation:

Explanation

This is the best configuration for scheduling Structured Streaming jobs for production, as it automatically recovers from query failures and keeps costs low. A new job cluster is created for each run of the job and terminated when the job completes, which saves costs and avoids resource contention. Retries are not needed for Structured Streaming jobs, as they can automatically recover from failures using checkpointing and write-ahead logs. Maximum concurrent runs should be set to 1 to avoid duplicate output or data loss. Verified References: Databricks Certified Data Engineer Professional, under "Monitoring & Logging" section; Databricks Documentation, under "Schedule streaming jobs" section.

NEW QUESTION # 153

A data engineer needs to dynamically create a table name string using three Python variables: region, store, and year. An example of a table name is below when region = "nyc", store = "100", and year = "2021":

nyc100_sales_2021

Which of the following commands should the data engineer use to construct the table name in Python?

- A. "{region}+{store}+_sales_"+"{year} "
- B. "{region}+{store}+_sales_+"{year} "
- C. f"{region} {store} _sales_ {year} "
- D. f"{region}+{store}+_sales_+"{year} "
- E. "{region} {store} _sales_ {year} "

Answer: C

NEW QUESTION # 154

Which statement characterizes the general programming model used by Spark Structured Streaming?

- A. Structured Streaming models new data arriving in a data stream as new rows appended to an unbounded table.
- B. Structured Streaming relies on a distributed network of nodes that hold incremental state values for cached stages.
- C. Structured Streaming leverages the parallel processing of GPUs to achieve highly parallel data throughput.
- D. Structured Streaming is implemented as a messaging bus and is derived from Apache Kafka.
- E. Structured Streaming uses specialized hardware and I/O streams to achieve sub-second latency for data transfer.

Answer: D

Explanation:

This is the correct answer because it characterizes the general programming model used by Spark Structured Streaming, which is to treat a live data stream as a table that is being continuously appended. This leads to a new stream processing model that is very similar to a batch processing model, where users can express their streaming computation using the same Dataset/DataFrame API as they would use for static data. The Spark SQL engine will take care of running the streaming query incrementally and continuously and updating the final result as streaming data continues to arrive. Verified References: [Databricks Certified Data Engineer Professional], under "Structured Streaming" section; Databricks Documentation, under "Overview" section.

NEW QUESTION # 155

.....

It is Databricks-Certified-Professional-Data-Engineer exam qualification certification that gives you capital of standing in society and serving your company. Nowadays, using the Internet to study on our Databricks-Certified-Professional-Data-Engineer exam questions has been a new trend of making people access to knowledge and capability-building. Our Databricks-Certified-Professional-Data-Engineer Preparation materials display a brand-new learning model and a comprehensive knowledge structure on our official exam bank, which aims at improving your technical skills and creating your value to your future.

Reliable Databricks-Certified-Professional-Data-Engineer Exam Braindumps: <https://www.actual4cert.com/Databricks-Certified-Professional-Data-Engineer-real-questions.html>

- Reliable Databricks-Certified-Professional-Data-Engineer Exam Test Databricks-Certified-Professional-Data-Engineer Practice Braindumps Test Databricks-Certified-Professional-Data-Engineer Simulator Fee Enter  www.troytecdumps.com  and search for { Databricks-Certified-Professional-Data-Engineer } to download for free Latest Databricks-Certified-Professional-Data-Engineer Exam Labs

