

Free PDF 2026 Databricks Useful Databricks-Certified-Professional-Data-Engineer Pass Test



The companies do not want to lose them and they offer a good package to convince the candidate to become a part of their organization. So, to fit in the game, you must go for the [Itbraindumps Databricks Certified Professional Data Engineer Practice Exam](#) that will show you where you stand and how hard you need to work to get the Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) certification exam.

Databricks Certified Professional Data Engineer exam covers a wide range of topics, including data engineering concepts, Databricks architecture, data ingestion and processing, data storage and management, and data security. Databricks-Certified-Professional-Data-Engineer Exam consists of 60 multiple-choice questions and participants have 90 minutes to complete it. Passing the exam requires a score of 70% or higher, and successful candidates receive a certificate that validates their expertise in building and managing data pipelines on the Databricks platform.

>> Databricks-Certified-Professional-Data-Engineer Pass Test <<

**Databricks-Certified-Professional-Data-Engineer Reliable Exam Online |
Online Databricks-Certified-Professional-Data-Engineer Test**

Databricks-Certified-Professional-Data-Engineer exam dumps are valid and we have helped lots of candidates pass the exam successfully, and they send the thankful letter to us. Databricks-Certified-Professional-Data-Engineer exam materials are edited and verified by professional experts, and they posse the professional knowledge for the exam, therefore you can use them at ease. In addition, we offer you free update for one, so you don't have to spend extra money on update version. We have online and offline chat service, and they possess the professional knowledge for Databricks-Certified-Professional-Data-Engineer Exam Braindumps, if you have any questions, you can consult us, we are glad to help you.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q71-Q76):

NEW QUESTION # 71

Drop the customers database and associated tables and data, all of the tables inside the database are managed tables. Which of the following SQL commands will help you accomplish this?

- A. DROP DELTA DATABASE customers
- B. DROP DATABASE customers FORCE
- C. All the tables must be dropped first before dropping database
- D. DROP DATABASE customers CASCADE
- E. DROP DATABASE customers INCLUDE

Answer: E

Explanation:

Explanation

The answer is `DROP DATABASE customers CASCADE`

Drop database with cascade option drops all the tables, since all of the tables inside the database are managed tables we do not need to perform any additional steps to clean the data in the storage.

NEW QUESTION # 72

An hourly batch job is configured to ingest data files from a cloud object storage container where each batch represent all records produced by the source system in a given hour. The batch job to process these records into the Lakehouse is sufficiently delayed to ensure no late-arriving data is missed. The `user_id` field represents a unique key for the data, which has the following schema: `user_id BIGINT, username STRING, user_utc STRING, user_region STRING, last_login BIGINT, auto_pay BOOLEAN, last_updated BIGINT` New records are all ingested into a table named `account_history` which maintains a full record of all data in the same schema as the source. The next table in the system is named `account_current` and is implemented as a Type 1 table representing the most recent value for each unique `user_id`.

Assuming there are millions of user accounts and tens of thousands of records processed hourly, which implementation can be used to efficiently update the described `account_current` table as part of each hourly batch job?

- A. Filter records in account history using the last updated field and the most recent hour processed, making sure to deduplicate on `username`; write a merge statement to update or insert the most recent value for each `username`.
- B. Overwrite the account current table with each batch using the results of a query against the account history table grouping by `user_id` and filtering for the max value of `last_updated`.
- C. Filter records in account history using the last updated field and the most recent hour processed, as well as the max last login by `user_id` write a merge statement to update or insert the most recent value for each `user_id`.
- D. Use Delta Lake version history to get the difference between the latest version of account history and one version prior, then write these records to account current.
- E. Use Auto Loader to subscribe to new files in the account history directory; configure a Structured Streaming trigger once job to batch update newly detected files into the account current table.

Answer: C

Explanation:

This is the correct answer because it efficiently updates the account current table with only the most recent value for each user id. The code filters records in account history using the last updated field and the most recent hour processed, which means it will only process the latest batch of data. It also filters by the max last login by user id, which means it will only keep the most recent record for each user id within that batch. Then, it writes a merge statement to update or insert the most recent value for each user id into account current, which means it will perform an upsert operation based on the user id column. Verified References:

[Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Upsert into a table using merge" section.

NEW QUESTION # 73

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

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

Maximum Concurrent Runs: 1

Answer: A

Explanation:

Maximum concurrent runs: Set to 1. There must be only one instance of each query concurrently active.

Retries: Set to Unlimited. <https://docs.databricks.com/en/structured-streaming/query-recovery.html>

NEW QUESTION # 74

A data pipeline uses Structured Streaming to ingest data from kafka to Delta Lake. Data is being stored in a bronze table, and includes the Kafka_generated timestamp, key, and value. Three months after the pipeline is deployed the data engineering team has noticed some latency issued during certain times of the day.

A senior data engineer updates the Delta Table's schema and ingestion logic to include the current timestamp (as recoded by Apache Spark) as well the Kafka topic and partition. The team plans to use the additional metadata fields to diagnose the transient processing delays:

Which limitation will the team face while diagnosing this problem?

- A. Updating the table schema will invalidate the Delta transaction log metadata.
- B. Spark cannot capture the topic partition fields from the kafka source.
- C. Updating the table schema requires a default value provided for each file added.
- D. **New fields not be computed for historic records.**

Answer: D

Explanation:

When adding new fields to a Delta table's schema, these fields will not be retrospectively applied to historical records that were ingested before the schema change. Consequently, while the team can use the new metadata fields to investigate transient processing delays moving forward, they will be unable to apply this diagnostic approach to past data that lacks these fields.

References:

* Databricks documentation on Delta Lake schema management: <https://docs.databricks.com/delta/delta-batch.html#schema-management>

NEW QUESTION # 75

A data engineer is using Auto Loader to read incoming JSON data as it arrives. They have configured Auto Loader to quarantine invalid JSON records but notice that over time, some records are being quarantined even though they are well-formed JSON.

The code snippet is:

```
df = (spark.readStream
      .format("cloudFiles")
      .option("cloudFiles.format", "json")
      .option("badRecordsPath", "/tmp/somewhere/badRecordsPath")
      .schema("a int, b int")
      .load("Volumes/catalog/schema/raw_data/"))
```

What is the cause of the missing data?

- A. The badRecordsPath location is accumulating many small files.
- B. The engineer forgot to set the option "cloudFiles.quarantineMode" = "rescue".
- C. **The source data is valid JSON but does not conform to the defined schema in some way.**
- D. At some point, the upstream data provider switched everything to multi-line JSON.

Answer: C

Explanation:

Comprehensive and Detailed Explanation From Exact Extract of Databricks Data Engineer Documents:

Databricks Auto Loader quarantines records that fail schema validation, even if they are syntactically valid JSON. The documentation explains that "records that do not match the defined schema are considered corrupt and written to the bad records path." In this case, the data likely contains valid JSON fields whose structure or data types differ from the declared schema (a int, b int). For example, if a field expected as an integer arrives as a string, Auto Loader will quarantine it. The presence of badRecordsPath confirms this behavior. The cloudFiles.quarantineMode option controls quarantine behavior but is unrelated to valid data that fails schema validation. Thus, the correct cause is C.

NEW QUESTION # 76

Itbraindumps's pledge to customers is that we can help customers 100% pass their IT certification exams. The quality of Itbraindumps's product has been recognized by many IT experts. The most important characteristic of our products is their pertinence. It only takes 20 hours for you to complete the training course and then easily pass your first time to attend Databricks Certification Databricks-Certified-Professional-Data-Engineer Exam. You will not regret to choose Itbraindumps, because choosing it represents the success.

Databricks-Certified-Professional-Data-Engineer Reliable Exam Online: https://www.itbraindumps.com/Databricks-Certified-Professional-Data-Engineer_exam.html

