

Latest Databricks Databricks-Certified-Professional-Data-Engineer Exam Questions in PDF Format

Databricks Databricks-Certified-Data-Engineer-Associate (Databricks-Certified-Data-Engineer-Associate-Exam)

Mark Question

Question 3 of 10

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. There is no way to indicate whether a table contains PII.

☐ B. "COMMENT PII"

☒ C. TBLPARAMETERS PII

☐ D. COMMENT "Contains PII"

☐ E. PII

Q No. Answer Marked

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Previous Next Show Answer Review Show List Save Session End Exam

Lead2PassExam has many Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) practice questions that reflect the pattern of the real Databricks Databricks-Certified-Professional-Data-Engineer exam. Lead2PassExam allows you to create a Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) exam dumps according to your preparation. It is easy to create the Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) practice questions by following just a few simple steps. Our Databricks-Certified-Professional-Data-Engineer exam dumps are customizable based on the time and type of questions.

Databricks Certified Professional Data Engineer Certification Exam can be attempted by professionals and students who have experience in data engineering, data management, ETL, and data processing. The preparation for the exam can be done via online training courses such as the Databricks Data Engineering Certification Preparation Course, the online Databricks Documentation, and different study materials such as books and videos from verified training providers.

>> Databricks-Certified-Professional-Data-Engineer Training Questions <<

100% Pass Quiz 2026 Databricks Databricks-Certified-Professional-Data-Engineer: Databricks Certified Professional Data Engineer Exam – High Pass-Rate Training Questions

Everyone is not willing to fall behind, but very few people take the initiative to change their situation. Take time to make a change and you will surely do it. Our Databricks-Certified-Professional-Data-Engineer learning materials can give you some help. Our company aims to help ease the pressure on you to prepare for the exam and eventually get a certificate. Obtaining a certificate is equivalent to having a promising future and good professional development. Our Databricks-Certified-Professional-Data-Engineer Learning Materials have a good reputation in the international community and their quality is guaranteed. Why don't you there have a brave attempt? You will certainly benefit from your wise choice.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q73-Q78):

NEW QUESTION # 73

What is the purpose of the bronze layer in a Multi-hop Medallion architecture?

- A. Powers ML applications
- B. Data quality checks, corrupt data quarantined
- C. Contain aggregated data that is to be consumed into Silver
- D. Reduces data storage by compressing the data
- E. Copy of raw data, easy to query and ingest data for downstream processes.

Answer: E

Explanation:

Explanation

The answer is, copy of raw data, easy to query and ingest data for downstream processes, Medallion Architecture - Databricks
Here are the typical role of Bronze Layer in a medallion architecture.

Bronze Layer:

1. Raw copy of ingested data
2. Replaces traditional data lake
3. Provides efficient storage and querying of full, unprocessed history of data
4. No schema is applied at this layer

Exam focus: Please review the below image and understand the role of each layer(bronze, silver, gold) in medallion architecture, you will see varying questions targeting each layer and its purpose.

Sorry I had to add the watermark some people in Udemy are copying my content.

NEW QUESTION # 74

What is the purpose of the bronze layer in a Multi-hop architecture?

- A. Can be used to eliminate duplicate records
- B. Perform data quality checks, corrupt data quarantined
- **C. Provides efficient storage and querying of full unprocessed history of data**
- D. Contains aggregated data that is to be consumed into Silver
- E. Used as a data source for Machine learning applications.

Answer: C

Explanation:

Explanation

The answer is Provides efficient storage and querying of full unprocessed history of data Medallion Architecture - Databricks Bronze Layer:

- 1.Raw copy of ingested data
- 2.Replaces traditional data lake
- 3.Provides efficient storage and querying of full, unprocessed history of data
- 4.No schema is applied at this layer

Exam focus: Please review the below image and understand the role of each layer(bronze, silver, gold) in medallion architecture, you will see varying questions targeting each layer and its purpose.

Sorry I had to add the watermark some people in Udemy are copying my content.

NEW QUESTION # 75

The data science team has created and logged a production model using MLflow. The following code correctly imports and applies the production model to output the predictions as a new DataFrame named preds with the schema "customer_id LONG, predictions DOUBLE, date DATE".

The data science team would like predictions saved to a Delta Lake table with the ability to compare all predictions across time. Churn predictions will be made at most once per day.

Which code block accomplishes this task while minimizing potential compute costs?

- A. `reds.write.format("delta").save("/preds/churn_preds")`
- **B. `preds.write.mode("append").saveAsTable("churn_preds")`**
- C. ☐
- D. ☐
- E. ☐

Answer: B

NEW QUESTION # 76

A Delta Lake table was created with the below query:

☐

Realizing that the original query had a typographical error, the below code was executed:

```
ALTER TABLE prod.sales_by_stor RENAME TO prod.sales_by_store
```

Which result will occur after running the second command?

- A. The table reference in the metastore is updated and all data files are moved.
- **B. The table reference in the metastore is updated and no data is changed.**
- C. All related files and metadata are dropped and recreated in a single ACID transaction.
- D. A new Delta transaction log is created for the renamed table.
- E. The table name change is recorded in the Delta transaction log.

Answer: B

Explanation:

The query uses the CREATE TABLE USING DELTA syntax to create a Delta Lake table from an existing Parquet file stored in DBFS. The query also uses the LOCATION keyword to specify the path to the Parquet file as /mnt/finance_edc_bucket/tx_sales.parquet. By using the LOCATION keyword, the query creates an external table, which is a table that is stored outside of the default warehouse directory and whose metadata is not managed by Databricks. An external table can be created from an existing directory in a cloud storage system, such as DBFS or S3, that contains data files in a supported format, such as Parquet or CSV.

The result that will occur after running the second command is that the table reference in the metastore is updated and no data is changed. The metastore is a service that stores metadata about tables, such as their schema, location, properties, and partitions. The metastore allows users to access tables using SQL commands or Spark APIs without knowing their physical location or format. When renaming an external table using the ALTER TABLE RENAME TO command, only the table reference in the metastore is updated with the new name; no data files or directories are moved or changed in the storage system. The table will still point to the same location and use the same format as before. However, if renaming a managed table, which is a table whose metadata and data are both managed by Databricks, both the table reference in the metastore and the data files in the default warehouse directory are moved and renamed accordingly. Verified References:

[Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "ALTER TABLE RENAME TO" section; Databricks Documentation, under "Metastore" section; Databricks Documentation, under "Managed and external tables" section.

NEW QUESTION # 77

A data team's Structured Streaming job is configured to calculate running aggregates for item sales to update a downstream marketing dashboard. The marketing team has introduced a new field to track the number of times this promotion code is used for each item. A junior data engineer suggests updating the existing query as follows: Note that proposed changes are in bold.

Which step must also be completed to put the proposed query into production?

- A. Remove .option (mergeSchema', true') from the streaming write
- B. Run REFRESH TABLE delta, /item_agg'
- C. Increase the shuffle partitions to account for additional aggregates
- **D. Specify a new checkpoint location**

Answer: D

Explanation:

When introducing a new aggregation or a change in the logic of a Structured Streaming query, it is generally necessary to specify a new checkpoint location. This is because the checkpoint directory contains metadata about the offsets and the state of the aggregations of a streaming query. If the logic of the query changes, such as including a new aggregation field, the state information saved in the current checkpoint would not be compatible with the new logic, potentially leading to incorrect results or failures. Therefore, to accommodate the new field and ensure the streaming job has the correct starting point and state information for aggregations, a new checkpoint location should be specified.

References:

* Databricks documentation on Structured Streaming: <https://docs.databricks.com/spark/latest/structured-streaming/index.html>

* Databricks documentation on streaming checkpoints: <https://docs.databricks.com/spark/latest/structured-streaming/production.html#checkpointing>

NEW QUESTION # 78

.....

[illegible]