

Earn The Badge Of Databricks Databricks-Certified-Professional-Data-Engineer Certification Exam On The First Attempt



It is a truism that an internationally recognized Databricks-Certified-Professional-Data-Engineer certification can totally mean you have a good command of the knowledge in certain areas and showcase your capacity to a considerable extend. If you are overwhelmed by workload heavily and cannot take a breath from it, why not choose our Databricks-Certified-Professional-Data-Engineer Preparation torrent? We are specialized in providing our customers with the most reliable and accurate exam materials and help them pass their exams by achieve their satisfied scores. With our Databricks-Certified-Professional-Data-Engineer practice materials, your exam will be a piece of cake.

To become a Databricks Certified Professional Data Engineer, candidates must pass a rigorous exam that covers a wide range of topics related to data engineering. Databricks-Certified-Professional-Data-Engineer Exam is designed to test a candidate's ability to design, build, and maintain data pipelines using Databricks tools and technologies. It also tests their understanding of data modeling, data warehousing, and data integration.

[**>> Sample Databricks-Certified-Professional-Data-Engineer Test Online <<**](#)

Preparation Databricks-Certified-Professional-Data-Engineer Store & Reliable Databricks-Certified-Professional-Data-Engineer Study Materials

With the development of society and the perfection of relative laws and regulations, the Databricks-Certified-Professional-Data-Engineer certificate in our career field becomes a necessity for our country. Passing the Databricks-Certified-Professional-Data-Engineer and obtaining the certificate may be the fastest and most direct way to change your position and achieve your goal. And we are just right here to give you help to pass the Databricks-Certified-Professional-Data-Engineer Exam. Being considered the most authentic brand in this career, our professional experts are making unremitting efforts to provide our customers the latest and valid Databricks-Certified-Professional-Data-Engineer exam simulation.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q183-Q188):

NEW QUESTION # 183

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

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

Answer: D

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 # 184

Which statement describes Delta Lake Auto Compaction?

- A. Before a Jobs cluster terminates, optimize is executed on all tables modified during the most recent job.
- B. An asynchronous job runs after the write completes to detect if files could be further compacted; if yes, an optimize job is executed toward a default of 128 MB.
- C. Optimized writes use logical partitions instead of directory partitions; because partition boundaries are only represented in metadata, fewer small files are written.
- D. An asynchronous job runs after the write completes to detect if files could be further compacted; if yes, an optimize job is executed toward a default of 1 GB.
- E. Data is queued in a messaging bus instead of committing data directly to memory; all data is committed from the messaging bus in one batch once the job is complete.

Answer: B

Explanation:

Explanation

This is the correct answer because it describes the behavior of Delta Lake Auto Compaction, which is a feature that automatically optimizes the layout of Delta Lake tables by coalescing small files into larger ones. Auto Compaction runs as an asynchronous job after a write to a table has succeeded and checks if files within a partition can be further compacted. If yes, it runs an optimize job with a default target file size of 128 MB.

Auto Compaction only compacts files that have not been compacted previously. Verified References:

[Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Auto Compaction for Delta Lake on Databricks" section.

"Auto compaction occurs after a write to a table has succeeded and runs synchronously on the cluster that has performed the write.

Auto compaction only compacts files that haven't been compacted previously."

<https://learn.microsoft.com/en-us/azure/databricks/delta/tune-file-size>

NEW QUESTION # 185

A new data engineer notices that a critical field was omitted from an application that writes its Kafka source to Delta Lake. This happened even though the critical field was in the Kafka source. That field was further missing from data written to dependent, long-term storage. The retention threshold on the Kafka service is seven days. The pipeline has been in production for three months.

Which describes how Delta Lake can help to avoid data loss of this nature in the future?

- A. Data can never be permanently dropped or deleted from Delta Lake, so data loss is not possible under any circumstance.
- B. Ingesting all raw data and metadata from Kafka to a bronze Delta table creates a permanent, replayable history of the data state.
- C. The Delta log and Structured Streaming checkpoints record the full history of the Kafka producer.
- D. Delta Lake schema evolution can retroactively calculate the correct value for newly added fields, as long as the data was in the original source.
- E. Delta Lake automatically checks that all fields present in the source data are included in the ingestion layer.

Answer: B

Explanation:

Explanation

This is the correct answer because it describes how Delta Lake can help to avoid data loss of this nature in the future. By ingesting

all raw data and metadata from Kafka to a bronze Delta table, Delta Lake creates a permanent, replayable history of the data state that can be used for recovery or reprocessing in case of errors or omissions in downstream applications or pipelines. Delta Lake also supports schema evolution, which allows adding new columns to existing tables without affecting existing queries or pipelines. Therefore, if a critical field was omitted from an application that writes its Kafka source to Delta Lake, it can be easily added later and the data can be reprocessed from the bronze table without losing any information. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Delta Lake core features" section.

NEW QUESTION # 186

A data engineer has created a transactions Delta table on Databricks that should be used by the analytics team. The analytics team wants to use the table with another tool that requires Apache Iceberg format.

What should the data engineer do?

- A. Create an Iceberg copy of the transactions Delta table which can be used by the analytics team.
- B. Enable uniform on the transactions table to 'iceberg' so that the table can be read as an Iceberg table.
- C. Convert the transactions Delta table to Iceberg and enable uniform so that the table can be read as a Delta table.
- D. Require the analytics team to use a tool that supports Delta table.

Answer: C

Explanation:

Delta Lake introduced Delta Universal Format (Delta UniForm), which allows seamless interoperability between Delta Lake and Apache Iceberg. This means a Delta table can be converted into an Iceberg table while maintaining Delta capabilities.

Explanation of Each Option:

(A) Require the analytics team to use a tool that supports Delta table

Incorrect: While Delta Lake is widely used, requiring the team to change tools is not a flexible or scalable solution.

(B) Enable uniform on the transactions table to 'iceberg' so that the table can be read as an Iceberg table

Incorrect: The uniform feature must be enabled after conversion.

You cannot directly enable uniform without first converting the table.

(C) Create an Iceberg copy of the transactions Delta table which can be used by the analytics team

Incorrect: Creating a separate Iceberg copy would duplicate storage and increase maintenance complexity.

This is not necessary when Delta UniForm allows direct compatibility with Iceberg.

(D) Convert the transactions Delta table to Iceberg and enable uniform so that the table can be read as a Delta table

Correct: The best approach is to convert the existing Delta table to Iceberg using the Databricks Delta to Iceberg migration tools.

After conversion, enabling uniform ensures the table remains accessible in both Delta and Iceberg formats.

Conclusion:

The best practice for interoperability between Delta and Iceberg is to convert the Delta table to Iceberg and enable uniform, ensuring cross-compatibility without data duplication.

Thus, Option (D) is the correct answer.

Reference:

Delta UniForm for Apache Iceberg - Databricks Documentation

Convert Delta to Iceberg - Databricks

NEW QUESTION # 187

Which statement describes the correct use of `pyspark.sql.functions.broadcast`?

- A. It marks a column as small enough to store in memory on all executors, allowing a broadcast join.
- B. It marks a column as having low enough cardinality to properly map distinct values to available partitions, allowing a broadcast join.
- C. It marks a DataFrame as small enough to store in memory on all executors, allowing a broadcast join.
- D. It caches a copy of the indicated table on attached storage volumes for all active clusters within a Databricks workspace.
- E. It caches a copy of the indicated table on all nodes in the cluster for use in all future queries during the cluster lifetime.

Answer: C

Explanation:

Explanation

<https://spark.apache.org/docs/3.1.3/api/python/reference/api/pyspark.sql.functions.broadcast.html>

NEW QUESTION # 188

If you are overwhelmed with the job at hand, and struggle to figure out how to prioritize your efforts, these would be the basic problem of low efficiency and production. You will never doubt anymore with our Databricks-Certified-Professional-Data-Engineer test prep. With our Databricks-Certified-Professional-Data-Engineer exam questions, you will not only get the Databricks-Certified-Professional-Data-Engineer Certification quickly, but also you can get the best and helpful knowledge. And that when you make a payment for our Databricks-Certified-Professional-Data-Engineer quiz torrent, you will possess this product in 5-10 minutes and enjoy the pleasure and satisfaction of your study time.

Preparation Databricks-Certified-Professional-Data-Engineer Store: <https://www.actualcollection.com/Databricks-Certified-Professional-Data-Engineer-exam-questions.html>

