

Databricks-Certified-Data-Analyst-Associate Training Questions, Databricks-Certified-Data-Analyst-Associate Latest Exam Vce



In this fast-changing world, the requirements for jobs and talents are higher, and if people want to find a job with high salary they must boost varied skills which not only include the good health but also the working abilities. We provide timely and free update for you to get more Databricks-Certified-Data-Analyst-Associate Questions torrent and follow the latest trend. The Databricks-Certified-Data-Analyst-Associate exam torrent is compiled by the experienced professionals and of great value.

Databricks Databricks-Certified-Data-Analyst-Associate Exam Syllabus Topics:

Topic	Details
Topic 1	• Data Management: The topic describes Delta Lake as a tool for managing data files, Delta Lake manages table metadata, benefits of Delta Lake within the Lakehouse, tables on Databricks, a table owner's responsibilities, and the persistence of data. It also identifies management of a table, usage of Data Explorer by a table owner, and organization-specific considerations of PII data. Lastly, the topic it explains how the LOCATION keyword changes, usage of Data Explorer to secure data.
Topic 2	• Analytics applications: It describes key moments of statistical distributions, data enhancement, and the blending of data between two source applications. Moreover, the topic also explains last-mile ETL, a scenario in which data blending would be beneficial, key statistical measures, descriptive statistics, and discrete and continuous statistics.
Topic 3	• SQL in the Lakehouse: It identifies a query that retrieves data from the database, the output of a SELECT query, a benefit of having ANSI SQL, access, and clean silver-level data. It also compares and contrasts MERGE INTO, INSERT TABLE, and COPY INTO. Lastly, this topic focuses on creating and applying UDFs in common scaling scenarios.

Topic 4	• Data Visualization and Dashboarding: Sub-topics of this topic are about of describing how notifications are sent, how to configure and troubleshoot a basic alert, how to configure a refresh schedule, the pros and cons of sharing dashboards, how query parameters change the output, and how to change the colors of all of the visualizations. It also discusses customized data visualizations, visualization formatting, Query Based Dropdown List, and the method for sharing a dashboard.
Topic 5	• Databricks SQL: This topic discusses key and side audiences, users, Databricks SQL benefits, complementing a basic Databricks SQL query, schema browser, Databricks SQL dashboards, and the purpose of Databricks SQL endpoints • warehouses. Furthermore, the delves into Serverless Databricks SQL endpoint • warehouses, trade-off between cluster size and cost for Databricks SQL endpoints • warehouses, and Partner Connect. Lastly it discusses small-file upload, connecting Databricks SQL to visualization tools, the medallion architecture, the gold layer, and the benefits of working with streaming data.

>> Databricks-Certified-Data-Analyst-Associate Training Questions <<

Databricks-Certified-Data-Analyst-Associate Latest Exam Vce & Pdf Databricks-Certified-Data-Analyst-Associate Torrent

Our Databricks-Certified-Data-Analyst-Associate learn materials include all the qualification tests in recent years, as well as corresponding supporting materials. Such a huge amount of database can greatly satisfy users' learning needs. Not enough valid Databricks-Certified-Data-Analyst-Associate test preparation materials, will bring many inconvenience to the user, such as delay learning progress, these are not conducive to the user pass exam, therefore, in order to solve these problems, our Databricks-Certified-Data-Analyst-Associate Certification material will do a complete summarize and precision of summary analysis to help you pass the Databricks-Certified-Data-Analyst-Associate exam with ease.

Databricks Certified Data Analyst Associate Exam Sample Questions (Q14-Q19):

NEW QUESTION # 14

A data engineering team has created a Structured Streaming pipeline that processes data in micro-batches and populates gold-level tables. The microbatches are triggered every minute.

A data analyst has created a dashboard based on this gold-level data. The project stakeholders want to see the results in the dashboard updated within one minute or less of new data becoming available within the gold-level tables.

Which of the following cautions should the data analyst share prior to setting up the dashboard to complete this task?

- A. The streaming cluster is not fault tolerant
- B. The gold-level tables are not appropriately clean for business reporting
- C. The required compute resources could be costly
- D. The dashboard cannot be refreshed that quickly
- E. The streaming data is not an appropriate data source for a dashboard

Answer: C

Explanation:

A Structured Streaming pipeline that processes data in micro-batches and populates gold-level tables every minute requires a high level of compute resources to handle the frequent data ingestion, processing, and writing. This could result in a significant cost for the organization, especially if the data volume and velocity are large. Therefore, the data analyst should share this caution with the project stakeholders before setting up the dashboard and evaluate the trade-offs between the desired refresh rate and the available budget. The other options are not valid cautions because:

B) The gold-level tables are assumed to be appropriately clean for business reporting, as they are the final output of the data engineering pipeline. If the data quality is not satisfactory, the issue should be addressed at the source or silver level, not at the gold level.

C) The streaming data is an appropriate data source for a dashboard, as it can provide near real-time insights and analytics for the business users. Structured Streaming supports various sources and sinks for streaming data, including Delta Lake, which can enable both batch and streaming queries on the same data.

D) The streaming cluster is fault tolerant, as Structured Streaming provides end-to-end exactly-once fault-tolerance guarantees through checkpointing and write-ahead logs. If a query fails, it can be restarted from the last checkpoint and resume processing.

E) The dashboard can be refreshed within one minute or less of new data becoming available in the gold-level tables, as Structured Streaming can trigger micro-batches as fast as possible (every few seconds) and update the results incrementally. However, this may not be necessary or optimal for the business use case, as it could cause frequent changes in the dashboard and consume more resources. Reference: Streaming on Databricks, Monitoring Structured Streaming queries on Databricks, A look at the new Structured Streaming UI in Apache Spark 3.0, Run your first Structured Streaming workload

NEW QUESTION # 15

The stakeholders.customers table has 15 columns and 3,000 rows of data. The following command is run:

```
CREATE TEMP VIEW stakeholders.eur_customers AS
SELECT * FROM stakeholders.customers
WHERE continent = 'eur';
```

After running `SELECT * FROM stakeholders.eur_customers`, 15 rows are returned. After the command executes completely, the user logs out of Databricks.

After logging back in two days later, what is the status of the `stakeholders.eur_customers` view?

- A. The view is not available in the metastore, but the underlying data can be accessed with `SELECT * FROM delta.`stakeholders.eur_customers``.
- B. The view remains available and `SELECT * FROM stakeholders.eur_customers` will execute correctly.
- C. The view remains available but attempting to `SELECT` from it results in an empty result set because data in views are automatically deleted after logging out.
- D. The view has been converted into a table.
- E. The view has been dropped.

Answer: E

Explanation:

The command you sent creates a TEMP VIEW, which is a type of view that is only visible and accessible to the session that created it. When the session ends or the user logs out, the TEMP VIEW is automatically dropped and cannot be queried anymore.

Therefore, after logging back in two days later, the status of the `stakeholders.eur_customers` view is that it has been dropped and `SELECT * FROM stakeholders.eur_customers` will result in an error. The other options are not correct because:

A) The view does not remain available, as it is a TEMP VIEW that is dropped when the session ends or the user logs out.

C) The view is not available in the metastore, as it is a TEMP VIEW that is not registered in the metastore. The underlying data cannot be accessed with `SELECT * FROM delta.stakeholders.eur_customers`, as this is not a valid syntax for querying a Delta Lake table. The correct syntax would be `SELECT * FROM delta.dbfs/stakeholders/eur_customers`, where the location path is enclosed in backticks. However, this would also result in an error, as the TEMP VIEW does not write any data to the file system and the location path does not exist.

D) The view does not remain available, as it is a TEMP VIEW that is dropped when the session ends or the user logs out. Data in views are not automatically deleted after logging out, as views do not store any data. They are only logical representations of queries on base tables or other views.

E) The view has not been converted into a table, as there is no automatic conversion between views and tables in Databricks. To create a table from a view, you need to use a `CREATE TABLE AS` statement or a similar command. Reference: CREATE VIEW | Databricks on AWS, Solved: How do temp views actually work? - Databricks - 20136, temp tables in Databricks - Databricks - 44012, Temporary View in Databricks - BIG DATA PROGRAMMERS, Solved: What is the difference between a Temporary View and ...

NEW QUESTION # 16

Delta Lake stores table data as a series of data files, but it also stores a lot of other information.

Which of the following is stored alongside data files when using Delta Lake?

- A. Owner account information
- B. None of these
- C. Table metadata, data summary visualizations, and owner account information
- D. Table metadata
- E. Data summary visualizations

Answer: D

Explanation:

Delta Lake is a storage layer that enhances data lakes with features like ACID transactions, schema enforcement, and time travel. While it stores table data as Parquet files, Delta Lake also keeps a transaction log (stored in the `_delta_log` directory) that contains detailed table metadata.

This metadata includes:

Table schema

Partitioning information

Data file paths

Transactional operations like inserts, updates, and deletes

Commit history and version control

This metadata is critical for supporting Delta Lake's advanced capabilities such as time travel and efficient query execution. Delta Lake does not store data summary visualizations or owner account information directly alongside the data files.

NEW QUESTION # 17

What describes Partner Connect in Databricks?

- **A. it allows multi-directional connection between Databricks and Databricks partners easier.**
- B. it allows for free use of Databricks partner tools through a common API.
- C. It exposes connection information to third-party tools via Databricks partners.
- D. It is a feature that runs Databricks partner tools on a Databricks SQL Warehouse (formerly known as a SQL endpoint).

Answer: A

Explanation:

Databricks Partner Connect is designed to simplify and streamline the integration between Databricks and its technology partners. It provides a unified interface within the Databricks platform that facilitates the discovery and connection to a variety of data, analytics, and AI tools. By automating the configuration of necessary resources such as clusters, tokens, and connection files, Partner Connect enables seamless, bi-directional data flow between Databricks and partner solutions. This integration enhances the overall functionality of the Databricks Lakehouse by allowing users to easily incorporate external tools and services into their workflows, thereby expanding the platform's capabilities and fostering a more cohesive data ecosystem.
https://www.databricks.com/blog/2021/11/18/how-generally-available-introducing-databricks-partner-connect-to-discover-and-connect-popular-data-and-ai-tools-to-the-lakehouse?utm_source=chatgpt.com

NEW QUESTION # 18

A data analyst has created a Query in Databricks SQL, and now they want to create two data visualizations from that Query and add both of those data visualizations to the same Databricks SQL Dashboard.

Which of the following steps will they need to take when creating and adding both data visualizations to the Databricks SQL Dashboard?

- A. They will need to copy the Query and create one data visualization per query.
- **B. They will need to add two separate visualizations to the dashboard based on the same Query.**
- C. They will need to create two separate dashboards.
- D. They will need to decide on a single data visualization to add to the dashboard.
- E. They will need to alter the Query to return two separate sets of results.

Answer: B

Explanation:

A data analyst can create multiple visualizations from the same query in Databricks SQL by clicking the + button next to the Results tab and selecting Visualization. Each visualization can have a different type, name, and configuration. To add a visualization to a dashboard, the data analyst can click the vertical ellipsis button beneath the visualization, select + Add to Dashboard, and choose an existing or new dashboard. The data analyst can repeat this process for each visualization they want to add to the same dashboard.

Reference: Visualization in Databricks SQL, Visualize queries and create a dashboard in Databricks SQL

NEW QUESTION # 19

