

Databricks

Exam Questions Databricks-Certified-Data-Analyst-Associate

Databricks Certified Data Analyst Associate Exam



NEW QUESTION 1

Which of the following should data analysts consider when working with personally identifiable information (PII) data?

- A. Organization-specific best practices for PII data
- B. Legal requirements for the area in which the data was collected
- C. None of these considerations
- D. Legal requirements for the area in which the analysis is being performed
- E. All of these considerations

Answer: E

Explanation:

Data analysts should consider all of these factors when working with PII data, as they may affect the data security, privacy, compliance, and quality. PII data is any information that can be used to identify a specific individual, such as name, address, phone number, email, social security number, etc. PII data may be subject to different legal and ethical obligations depending on the context and location of the data collection and analysis. For example, some countries or regions may have stricter data protection laws than others, such as the General Data Protection Regulation (GDPR) in the European Union. Data analysts should also follow the organization-specific best practices for PII data, such as encryption, anonymization, masking, access control, auditing, etc. These best practices can help prevent data breaches, unauthorized access, misuse, or loss of PII data. References:

? How to Use Databricks to Encrypt and Protect PII Data

? Automating Sensitive Data (PII/PHI) Detection

? Databricks Certified Data Analyst Associate

NEW QUESTION 2

A data analyst wants to create a dashboard with three main sections: Development, Testing, and Production. They want all three sections on the same dashboard, but they want to clearly designate the sections using text on the dashboard.

Which of the following tools can the data analyst use to designate the Development, Testing, and Production sections using text?

- A. Separate endpoints for each section
- B. Separate queries for each section
- C. Markdown-based text boxes
- D. Direct text written into the dashboard in editing mode
- E. Separate color palettes for each section

Answer: C

Explanation:

Markdown-based text boxes are useful as labels on a dashboard. They allow the data analyst to add text to a dashboard using the %md magic command in a notebook cell and then select the dashboard icon in the cell actions menu. The text can be formatted using markdown syntax and can include headings, lists, links, images, and more. The text boxes can be resized and moved around on the dashboard using the float layout option. References: Dashboards in notebooks, How to add text to a dashboard in Databricks

NEW QUESTION 3

A data analyst runs the following command:

INSERT INTO stakeholders.suppliers TABLE stakeholders.new_suppliers; What is the result of running this command?

- A. The suppliers table now contains both the data it had before the command was run and the data from the new suppliers table, and any duplicate data is deleted.
- B. The command fails because it is written incorrectly.
- C. The suppliers table now contains both the data it had before the command was run and the data from the new suppliers table, including any duplicate data.
- D. The suppliers table now contains the data from the new suppliers table, and the new suppliers table now contains the data from the suppliers table.
- E. The suppliers table now contains only the data from the new suppliers table.

Answer: B

Explanation:

The command INSERT INTO stakeholders.suppliers TABLE stakeholders.new_suppliers is not a valid syntax for inserting data into a table in Databricks SQL.

According to the documentation¹², the correct syntax for inserting data into a table is either:

? INSERT { OVERWRITE | INTO } [TABLE] table_name [PARTITION clause] [(column_name [, ...]) | BY NAME] query

? INSERT INTO [TABLE] table_name REPLACE WHERE predicate query

The command in the question is missing the OVERWRITE or INTO keyword, and the query part that specifies the source of the data to be inserted. The TABLE keyword is optional and can be omitted. The PARTITION clause and the column list are also optional and depend on the table schema and the data source.

Therefore, the command in the question will fail with a syntax error.

References:

? INSERT | Databricks on AWS

? INSERT - Azure Databricks - Databricks SQL | Microsoft Learn

NEW QUESTION 4

A data team has been given a series of projects by a consultant that need to be implemented in the Databricks Lakehouse Platform.

Which of the following projects should be completed in Databricks SQL?

- A. Testing the quality of data as it is imported from a source
- B. Tracking usage of feature variables for machine learning projects
- C. Combining two data sources into a single, comprehensive dataset
- D. Segmenting customers into like groups using a clustering algorithm
- E. Automating complex notebook-based workflows with multiple tasks

Answer: C

Explanation:

Databricks SQL is a service that allows users to query data in the lakehouse using SQL and create visualizations and dashboards¹. One of the common use cases for Databricks SQL is to combine data from different sources and formats into a single, comprehensive dataset that can be used for further analysis or reporting². For example, a data analyst can use Databricks SQL to join data from a CSV file and a Parquet file, or from a Delta table and a JDBC table, and create a new table or view that contains the combined data³. This can help simplify the data management and governance, as well as improve the data quality and consistency. References:

- ? Databricks SQL overview
- ? Databricks SQL use cases
- ? Joining data sources

NEW QUESTION 5

A data analyst has been asked to configure an alert for a query that returns the income in the accounts_receivable table for a date range. The date range is configurable using a Date query parameter.

The Alert does not work.

Which of the following describes why the Alert does not work?

- A. Alerts don't work with queries that access tables.
- B. Queries that return results based on dates cannot be used with Alerts.
- C. The wrong query parameter is being use
- D. Alerts only work with Date and Time query parameters.
- E. Queries that use query parameters cannot be used with Alerts.
- F. The wrong query parameter is being use
- G. Alerts only work with dropdown list query parameters, not dates.

Answer: D

Explanation:

According to the Databricks documentation¹, queries that use query parameters cannot be used with Alerts. This is because Alerts do not support user input or dynamic values. Alerts leverage queries with parameters using the default value specified in the SQL editor for each parameter. Therefore, if the query uses a Date query parameter, the alert will always use the same date range as the default value, regardless of the actual date. This may cause the alert to not work as expected, or to not trigger at all. References:

- ? Databricks SQL alerts: This is the official documentation for Databricks SQL alerts, where you can find information about how to create, configure, and monitor alerts, as well as the limitations and best practices for using alerts.

NEW QUESTION 6

Which of the following statements about a refresh schedule is incorrect?

- A. A query can be refreshed anywhere from 1 minute to 2 weeks
- B. Refresh schedules can be configured in the Query Editor.
- C. A query being refreshed on a schedule does not use a SQL Warehouse (formerly known as SQL Endpoint).
- D. A refresh schedule is not the same as an alert.
- E. You must have workspace administrator privileges to configure a refresh schedule

Answer: C

Explanation:

Refresh schedules are used to rerun queries at specified intervals, and these queries typically require computational resources to execute. In the context of a cloud data service like Databricks, this would typically involve the use of a SQL Warehouse (or a SQL Endpoint, as they were formerly known) to provide the necessary computational resources. Therefore, the statement is incorrect because scheduled query refreshes would indeed use a SQL Warehouse/Endpoint to execute the query.

NEW QUESTION 7

Which of the following describes how Databricks SQL should be used in relation to other business intelligence (BI) tools like Tableau, Power BI, and Looker?

- A. As an exact substitute with the same level of functionality
- B. As a substitute with less functionality
- C. As a complete replacement with additional functionality
- D. As a complementary tool for professional-grade presentations
- E. As a complementary tool for quick in-platform BI work

Answer: E

Explanation:

Databricks SQL is not meant to replace or substitute other BI tools, but rather to complement them by providing a fast and easy way to query, explore, and visualize data on the lakehouse using the built-in SQL editor, visualizations, and dashboards. Databricks SQL also integrates seamlessly with popular BI tools like Tableau, Power BI, and Looker, allowing analysts to use their preferred tools to access data through Databricks clusters and SQL warehouses. Databricks SQL offers low-code and no-code experiences, as well as optimized connectors and serverless compute, to enhance the productivity and performance of BI workloads on the lakehouse. References: Databricks SQL, Connecting Applications and BI Tools to Databricks SQL, Databricks integrations overview, Databricks SQL: Delivering a Production SQL Development Experience on the Lakehouse

NEW QUESTION 8

How can a data analyst determine if query results were pulled from the cache?

- A. Go to the Query History tab and click on the text of the query
- B. The slideout shows if the results came from the cache.
- C. Go to the Alerts tab and check the Cache Status alert.
- D. Go to the Queries tab and click on Cache Status
- E. The status will be green if the results from the last run came from the cache.

- F. Go to the SQL Warehouse (formerly SQL Endpoints) tab and click on Cach
- G. The Cache file will show the contents of the cache.
- H. Go to the Data tab and click Last Quer
- I. The details of the query will show if the results came from the cache.

Answer: A

Explanation:

Databricks SQL uses a query cache to store the results of queries that have been executed previously. This improves the performance and efficiency of repeated queries. To determine if a query result was pulled from the cache, you can go to the Query History tab in the Databricks SQL UI and click on the text of the query. A slideout will appear on the right side of the screen, showing the query details, including the cache status. If the result came from the cache, the cache status will show ??Cached??. If the result did not come from the cache, the cache status will show ??Not cached??. You can also see the cache hit ratio, which is the percentage of queries that were served from the cache. References: The answer can be verified from Databricks SQL documentation which provides information on how to use the query cache and how to check the cache status. Reference link: Databricks SQL - Query Cache

NEW QUESTION 9

A data analyst is working with gold-layer tables to complete an ad-hoc project. A stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use.

Which of the following terms is used to describe this data augmentation?

- A. Data testing
- B. Ad-hoc improvements
- C. Last-mile
- D. Last-mile ETL
- E. Data enhancement

Answer: E

Explanation:

Data enhancement is the process of adding or enriching data with additional information to improve its quality, accuracy, and usefulness. Data enhancement can be used to augment existing data sources with new data sources, such as external datasets, synthetic data, or machine learning models. Data enhancement can help data analysts to gain deeper insights, discover new patterns, and solve complex problems. Data enhancement is one of the applications of generative AI, which can leverage machine learning to generate synthetic data for better models or safer data sharing¹.

In the context of the question, the data analyst is working with gold-layer tables, which are curated business-level tables that are typically organized in consumption-ready project-specific databases²³⁴. The gold-layer tables are the final layer of data transformations and data quality rules in the medallion lakehouse architecture, which is a data design pattern used to logically organize data in a lakehouse². The stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use. This means that the analyst can use the additional dataset to enhance the existing gold-layer tables with more information, such as new features, attributes, or metrics. This data augmentation can help the analyst to complete the ad-hoc project more effectively and efficiently.

References:

- ? What is the medallion lakehouse architecture? - Databricks
- ? Data Warehousing Modeling Techniques and Their Implementation on the Databricks Lakehouse Platform | Databricks Blog
- ? What is the medallion lakehouse architecture? - Azure Databricks
- ? What is a Medallion Architecture? - Databricks
- ? Synthetic Data for Better Machine Learning | Databricks Blog

NEW QUESTION 10

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

Databricks-Certified-Data-Analyst-Associate Practice Exam Features:

- * Databricks-Certified-Data-Analyst-Associate Questions and Answers Updated Frequently
- * Databricks-Certified-Data-Analyst-Associate Practice Questions Verified by Expert Senior Certified Staff
- * Databricks-Certified-Data-Analyst-Associate Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * Databricks-Certified-Data-Analyst-Associate Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The Databricks-Certified-Data-Analyst-Associate Practice Test Here](#)