

Databricks

Exam Questions Databricks-Certified-Data-Analyst-Associate

Databricks Certified Data Analyst Associate Exam



NEW QUESTION 1

A data organization has a team of engineers developing data pipelines following the medallion architecture using Delta Live Tables. While the data analysis team working on a project is using gold-layer tables from these pipelines, they need to perform some additional processing of these tables prior to performing their analysis.

Which of the following terms is used to describe this type of work?

- A. Data blending
- B. Last-mile
- C. Data testing
- D. Last-mile ETL
- E. Data enhancement

Answer: D

Explanation:

Last-mile ETL is the term used to describe the additional processing of data that is done by data analysts or data scientists after the data has been ingested, transformed, and stored in the lakehouse by data engineers. Last-mile ETL typically involves tasks such as data cleansing, data enrichment, data aggregation, data filtering, or data sampling that are specific to the analysis or machine learning use case. Last-mile ETL can be done using Databricks SQL, Databricks notebooks, or Databricks Machine

Learning. References: Databricks - Last-mile ETL, Databricks - Data Analysis with Databricks SQL

NEW QUESTION 2

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 alter the Query to return two separate sets of results.
- 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 copy the Query and create one data visualization per query.

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

NEW QUESTION 3

A business analyst has been asked to create a data entity/object called sales_by_employee. It should always stay up-to-date when new data are added to the sales table. The new entity should have the columns sales_person, which will be the name of the employee from the employees table, and sales, which will be all sales for that particular sales person. Both the sales table and the employees table have an employee_id column that is used to identify the sales person.

Which of the following code blocks will accomplish this task?

A)

```
CREATE TEMPORARY TABLE sales_by_employee AS
  SELECT employees.employee_name sales_person,
         sales.sales
  FROM sales
  JOIN employees
  ON employees.employee_id = sales.employee_id;
```

B)

```
CREATE OR REPLACE VIEW sales_by_employee USING
  SELECT employees.employee_name sales_person,
         sales.sales
  FROM sales
  JOIN employees
  ON employees.employee_id = sales.employee_id;
```

C)

```
SELECT employees.employee_name sales_person,
       sales.sales
FROM sales
JOIN employees
ON employees.employee_id = sales.employee_id USING
CREATE OR REPLACE VIEW sales_by_employee;
```

D)

```
CREATE OR REPLACE VIEW sales_by_employee AS
SELECT employees.employee_name sales_person,
       sales.sales FROM sales
JOIN employees
ON employees.employee_id = sales.employee_id;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: D

Explanation:

The SQL code provided in Option D is the correct way to create a view named sales_by_employee that will always stay up-to-date with the sales and employees tables. The code uses the CREATE OR REPLACE VIEW statement to define a new view that joins the sales and employees tables on the employee_id column. It selects the employee_name as sales_person and all sales for each employee, ensuring that the data entity/object is always up-to-date when new data are added to these tables.

References: The answer can be verified from Databricks SQL documentation which provides insights on creating views using SQL queries, joining tables, and selecting specific columns to be included in the view. Reference link: [Databricks SQL](#)

NEW QUESTION 4

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 5

A data analyst needs to use the Databricks Lakehouse Platform to quickly create SQL queries and data visualizations. It is a requirement that the compute resources in the platform can be made serverless, and it is expected that data visualizations can be placed within a dashboard.

Which of the following Databricks Lakehouse Platform services/capabilities meets all of these requirements?

- A. Delta Lake
- B. Databricks Notebooks
- C. Tableau
- D. Databricks Machine Learning
- E. Databricks SQL

Answer: E

Explanation:

Databricks SQL is a serverless data warehouse on the Lakehouse that lets you run all of your SQL and BI applications at scale with your tools of choice, all at a fraction of the cost of traditional cloud data warehouses¹. Databricks SQL allows you to create SQL queries and data visualizations using the SQL Analytics UI or the Databricks

SQL CLI2. You can also place your data visualizations within a dashboard and share it with other users in your organization3. Databricks SQL is powered by Delta Lake, which provides reliability, performance, and governance for your data lake4. References:

- ? Databricks SQL
- ? Query data using SQL Analytics
- ? Visualizations in Databricks notebooks
- ? Delta Lake

NEW QUESTION 6

A data analyst creates a Databricks SQL Query where the result set has the following schema:

region STRING number_of_customer INT

When the analyst clicks on the "Add visualization" button on the SQL Editor page, which of the following types of visualizations will be selected by default?

- A. Violin Chart
- B. Line Chart
- C. IBar Chart
- D. Histogram
- E. There is no default
- F. The user must choose a visualization type.

Answer: C

Explanation:

According to the Databricks SQL documentation, when a data analyst clicks on the ??Add visualization?? button on the SQL Editor page, the default visualization type is Bar Chart. This is because the result set has two columns: one of type STRING and one of type INT. The Bar Chart visualization automatically assigns the STRING column to the X-axis and the INT column to the Y-axis. The Bar Chart visualization is suitable for showing the distribution of a numeric variable across different categories. References: Visualization in Databricks SQL, Visualization types

NEW QUESTION 7

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 8

Which of the following is a benefit of Databricks SQL using ANSI SQL as its standard SQL dialect?

- A. It has increased customization capabilities
- B. It is easy to migrate existing SQL queries to Databricks SQL
- C. It allows for the use of Photon's computation optimizations
- D. It is more performant than other SQL dialects
- E. It is more compatible with Spark's interpreters

Answer: B

Explanation:

Databricks SQL uses ANSI SQL as its standard SQL dialect, which means it follows the SQL specifications defined by the American National Standards Institute (ANSI). This makes it easier to migrate existing SQL queries from other data warehouses or platforms that also use ANSI SQL or a similar dialect, such as PostgreSQL, Oracle, or Teradata. By using ANSI SQL, Databricks SQL avoids surprises in behavior or unfamiliar syntax that may arise from using a non-standard SQL dialect, such as Spark SQL or Hive SQL12. Moreover, Databricks SQL also adds compatibility features to support common SQL constructs that are widely used in other data warehouses, such as `QUALIFY`, `FILTER`, and user-defined functions2. References: ANSI compliance in Databricks Runtime, Evolution of the SQL language at Databricks: ANSI standard by default and easier migrations from data warehouses

NEW QUESTION 9

A data analyst created and is the owner of the managed table `my_table`. They now want to change ownership of the table to a single other user using Data Explorer.

Which of the following approaches can the analyst use to complete the task?

- A. Edit the Owner field in the table page by removing their own account
- B. Edit the Owner field in the table page by selecting All Users
- C. Edit the Owner field in the table page by selecting the new owner's account

- D. Edit the Owner field in the table page by selecting the Admins group
- E. Edit the Owner field in the table page by removing all access

Answer: C

Explanation:

The Owner field in the table page shows the current owner of the table and allows the owner to change it to another user or group. To change the ownership of the table, the owner can click on the Owner field and select the new owner from the drop-down list. This will transfer the ownership of the table to the selected user or group and remove the previous owner from the list of table access control entries¹. The other options are incorrect because:

- ? A. Removing the owner's account from the Owner field will not change the ownership of the table, but will make the table ownerless².
- ? B. Selecting All Users from the Owner field will not change the ownership of the table, but will grant all users access to the table³.
- ? D. Selecting the Admins group from the Owner field will not change the ownership of the table, but will grant the Admins group access to the table³.
- ? E. Removing all access from the Owner field will not change the ownership of the table, but will revoke all access to the table⁴. References:

- ? 1: Change table ownership
- ? 2: Ownerless tables
- ? 3: Table access control
- ? 4: Revoke access to a table

NEW QUESTION 10

Data professionals with varying titles use the Databricks SQL service as the primary touchpoint with the Databricks Lakehouse Platform. However, some users will use other services like Databricks Machine Learning or Databricks Data Science and Engineering.

Which of the following roles uses Databricks SQL as a secondary service while primarily using one of the other services?

- A. Business analyst
- B. SQL analyst
- C. Data engineer
- D. Business intelligence analyst
- E. Data analyst

Answer: C

Explanation:

Data engineers are primarily responsible for building, managing, and optimizing data pipelines and architectures. They use Databricks Data Science and Engineering service to perform tasks such as data ingestion, transformation, quality, and governance. Data engineers may use Databricks SQL as a secondary service to query, analyze, and visualize data from the lakehouse, but this is not their main

focus. References: Databricks SQL overview, Databricks Data Science and Engineering overview, Data engineering with Databricks

NEW QUESTION 10

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 11

A data analyst is attempting to drop a table `my_table`. The analyst wants to delete all table metadata and data.

They run the following command: `DROP TABLE IF EXISTS my_table;`

While the object no longer appears when they run `SHOW TABLES`, the data files still exist.

Which of the following describes why the data files still exist and the metadata files were deleted?

- A. The table's data was larger than 10 GB
- B. The table did not have a location
- C. The table was external
- D. The table's data was smaller than 10 GB
- E. The table was managed

Answer: C

Explanation:

An external table is a table that is defined in the metastore, but its data is stored outside of the Databricks environment, such as in S3, ADLS, or GCS. When an external table is dropped, only the metadata is deleted from the metastore, but the data files are not affected. This is different from a managed table, which is a table whose data is stored in the Databricks environment, and whose data files are deleted when the table is dropped. To delete the data files of an external table, the analyst needs to specify the `PURGE` option in the `DROP TABLE` command, or manually delete the files from the storage system. References: `DROP TABLE`, Drop Delta table features, Best practices for dropping a managed Delta Lake table

NEW QUESTION 13

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 15

Which of the following is an advantage of using a Delta Lake-based data lakehouse over common data lake solutions?

- A. ACID transactions
- B. Flexible schemas
- C. Data deletion
- D. Scalable storage
- E. Open-source formats

Answer: A

Explanation:

A Delta Lake-based data lakehouse is a data platform architecture that combines the scalability and flexibility of a data lake with the reliability and performance of a data warehouse. One of the key advantages of using a Delta Lake-based data lakehouse over common data lake solutions is that it supports ACID transactions, which ensure data integrity and consistency. ACID transactions enable concurrent reads and writes, schema enforcement and evolution, data versioning and rollback, and data quality checks. These features are not available in traditional data lakes, which rely on file-based storage systems that do not support transactions. References:

- ? Delta Lake: Lakehouse, warehouse, advantages | Definition
- ? Synapse – Data Lake vs. Delta Lake vs. Data Lakehouse
- ? Data Lake vs. Delta Lake - A Detailed Comparison
- ? Building a Data Lakehouse with Delta Lake Architecture: A Comprehensive Guide

NEW QUESTION 16

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 17

.....

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](#)