

Databricks

Exam Questions Databricks-Certified-Data-Analyst-Associate

Databricks Certified Data Analyst Associate Exam



NEW QUESTION 1

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 2

A data engineer is working with a nested array column products in table transactions. They want to expand the table so each unique item in products for each row has its own row where the transaction_id column is duplicated as necessary.

They are using the following incomplete command:

```
SELECT
    transaction_id,
    _____ AS product
FROM transactions;
```

Which of the following lines of code can they use to fill in the blank in the above code block so that it successfully completes the task?

- A. array distinct(produces)
- B. explode(produces)
- C. reduce(produces)
- D. array(produces)
- E. flatten(produces)

Answer: B

Explanation:

The explode function is used to transform a DataFrame column of arrays or maps into multiple rows, duplicating the other column's values. In this context, it will be used to expand the nested array column products in the transactions table so that each unique item in products for each row has its own row and the transaction_id column is duplicated as necessary. References: Databricks Documentation

I also noticed that you sent me an image along with your message. The image shows a snippet of SQL code that is incomplete. It begins with ??SELECT??

indicating a query to retrieve data. ??transaction_id,?? suggests that transaction_id is one of the columns being selected. There are blanks indicated by underscores where certain parts of the SQL command should be, including what appears to be an alias for a column and part of the FROM clause. The query ends with ??FROM transactions;?? indicating data is being selected from a ??transactions?? table.

If you are interested in learning more about Databricks Data Analyst Associate certification, you can check out the following resources:

? Databricks Certified Data Analyst Associate: This is the official page for the certification exam, where you can find the exam guide, registration details, and preparation tips.

? Data Analysis With Databricks SQL: This is a self-paced course that covers the topics and skills required for the certification exam. You can access it for free on Databricks Academy.

? Tips for the Databricks Certified Data Analyst Associate Certification: This is a blog post that provides some useful advice and study tips for passing the certification exam.

? Databricks Certified Data Analyst Associate Certification: This is another blog post that gives an overview of the certification exam and its benefits.

NEW QUESTION 3

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 4

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 5

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 6

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 7

A data analyst has created a user-defined function using the following line of code: CREATE FUNCTION price(spend DOUBLE, units DOUBLE) RETURNS DOUBLE

RETURN spend / units;

Which of the following code blocks can be used to apply this function to the customer_spend and customer_units columns of the table customer_summary to create column customer_price?

- A. SELECT PRICE customer_spend, customer_units AS customer_price FROM customer_summary
- B. SELECT price FROM customer_summary
- C. SELECT function(price(customer_spend, customer_units)) AS customer_price FROM customer_summary
- D. SELECT double(price(customer_spend, customer_units)) AS customer_price FROM customer_summary
- E. SELECT price(customer_spend, customer_units) AS customer_price FROM customer_summary

Answer: E

Explanation:

A user-defined function (UDF) is a function defined by a user, allowing custom logic to be reused in the user environment¹. To apply a UDF to a table, the syntax is SELECT udf_name(column_name) AS alias FROM table_name². Therefore, option E is

the correct way to use the UDF price to create a new column customer_price based on the existing columns customer_spend and customer_units from the table customer_summary. References:

? What are user-defined functions (UDFs)?

? User-defined scalar functions - SQL V

NEW QUESTION 8

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 9

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 10

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 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](#)