Databricks

Exam Questions Databricks-Certified-Data-Analyst-Associate

Databricks Certified Data Analyst Associate Exam

NEW QUESTION 1
A data analyst has recently joined a new team that uses Databricks SQL, but the analyst has never used Databricks before. The analyst wants to know where in Databricks SQL they can write and execute SQL queries.
On which of the following pages can the analyst write and execute SQL queries?

A. Data page
B. Dashboards page
C. Queries page
D. Alerts page
E. SQL Editor page

Answer: E

Explanation:
 The SQL Editor page is where the analyst can write and execute SQL queries in Databricks SQL. The SQL Editor page has a query pane where the analyst can type or paste SQL statements, and a results pane where the analyst can view the query results in a table or a chart. The analyst can also browse data objects, edit multiple queries, execute a single query or multiple queries, terminate a query, save a query, download a query result, and more from the SQL Editor page.
References: Create a query in SQL editor

NEW QUESTION 2
A data analyst has a managed table table_name in database database_name. They would now like to remove the table from the database and all of the data files associated with the table. The rest of the tables in the database must continue to exist.
Which of the following commands can the analyst use to complete the task without producing an error?

A. DROP DATABASE database_name;
B. DROP TABLE database_name.table_name;
C. DELETE TABLE database_name.table_name;
D. DELETE TABLE table_name FROM database_name;
E. DROP TABLE table_name FROM database_name;

Answer: B

Explanation:
 The DROP TABLE command removes a table from the metastore and deletes the associated data files. The syntax for this command is DROP TABLE [IF EXISTS] [database_name.]table_name;. The optional IF EXISTS clause prevents an error if the table does not exist. The optional database_name. prefix specifies the database where the table resides. If not specified, the current database is used. Therefore, the correct command to remove the table table_name from the database database_name and all of the data files associated with it is DROP TABLE database_name.table_name;. The other commands are either invalid syntax or would produce undesired
results. References: Databricks - DROP TABLE

NEW QUESTION 3
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 4
Which of the following approaches can be used to connect Databricks to Fivetran for data ingestion?

A. Use Workflows to establish a SQL warehouse (formerly known as a SQL endpoint) for Fivetran to interact with
B. Use Delta Live Tables to establish a cluster for Fivetran to interact with
C. Use Partner Connect's automated workflow to establish a cluster for Fivetran to interact with
D. Use Partner Connect's automated workflow to establish a SQL warehouse (formerly known as a SQL endpoint) for Fivetran to interact with
E. Use Workflows to establish a cluster for Fivetran to interact with

Answer: C

Explanation:
 Partner Connect is a feature that allows you to easily connect your Databricks workspace to Fivetran and other ingestion partners using an automated workflow. You can select a SQL warehouse or a cluster as the destination for your data replication, and the connection details are sent to Fivetran. You can then choose from over 200 data sources that Fivetran supports and start ingesting data into Delta
Lake. References: Connect to Fivetran using Partner Connect, Use Databricks with Fivetran

NEW QUESTION 5

Which of the following benefits of using Databricks SQL is provided by Data Explorer?

A. It can be used to run UPDATE queries to update any tables in a database.
B. It can be used to view metadata and data, as well as view/change permissions.
C. It can be used to produce dashboards that allow data exploration.
D. It can be used to make visualizations that can be shared with stakeholders.
E. It can be used to connect to third party BI cools.

Answer: B

Explanation:
Data Explorer is a user interface that allows you to discover and manage data, schemas, tables, models, and permissions in Databricks SQL. You can use Data Explorer to view schema details, preview sample data, and see table and model details and properties. Administrators can view and change owners, and admins and data object owners can grant and revoke permissions1. References: Discover and manage data using Data Explorer

NEW QUESTION 6
A data analysis team is working with the table_bronze SQL table as a source for one of its most complex projects. A stakeholder of the project notices that some of the downstream data is duplicative. The analysis team identifies table_bronze as the source of the duplication.
Which of the following queries can be used to deduplicate the data from table_bronze and write it to a new table table_silver?
A)
CREATE TABLE table•_silver AS SELECT DISTINCT *
FROM table_bronze;
B)
CREATE TABLE table_silver AS INSERT *
FROM table_bronze;
C)
CREATE TABLE table_silver AS MERGE DEDUPLICATE *
FROM table_bronze;
D)
INSERT INTO TABLE table_silver SELECT * FROM table_bronze;
E)
INSERT OVERWRITE TABLE table_silver SELECT * FROM table_bronze;

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: A

Explanation:
Option A uses the SELECT DISTINCT statement to remove duplicate rows from the table_bronze and create a new table table_silver with the deduplicated data. This is the correct way to deduplicate data using Spark SQL12. Option B simply inserts all the rows from table_bronze into table_silver, without removing any duplicates. Option C is not a valid syntax for Spark SQL, as there is no MERGE DEDUPLICATE statement. Option D appends all the rows from table_bronze into table_silver, without removing any duplicates. Option E overwrites the existing data in table_silver with the data from table_bronze, without removing any duplicates. References: Delete Duplicate using SPARK SQL, Spark SQL - How to Remove Duplicate Rows

NEW QUESTION 7
A data analyst has been asked to use the below tablesales_tableto get the percentage rank of products within region by the sales:

region	product	sales
WEST	A	1880.59
EAST	A	2045.99
EAST	B	4583.23
WEST	B	3391.19

The result of the query should look like this:

region	product	sales
EAST	B	0
EAST	A	1
WEST	B	0
WEST	A	1

Which of the following queries will accomplish this task?

A)
```
SELECT
    region,
    product,
    RANK() OVER (
        PARTITION BY region
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

B)
```
SELECT
    region,
    product,
    PERCENT_RANK () OVER (
        PARTITION BY region
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

C)
```
SELECT
    region,
    product,
    PERCENT_RANK () OVER (
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
```

D)

```
SELECT
    region,
    product,
    PERCENT RANK () OVER (
        PARTITION BY product
        ORDER BY sales DESC
    ) AS rank
FROM sales_table;
GROUP BY region, product;
```

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

Answer: B

Explanation:
The correct query to get the percentage rank of products within region by the sales is option B. This query uses the PERCENT_RANK() window function to calculate the relative rank of each product within each region based on the sales amount. The window function is partitioned by region and ordered by sales in descending order. The result is aliased as rank and displayed along with the region and product columns. The other options are incorrect because:
? A. Option A uses the RANK() window function instead of the PERCENT_RANK() function. The RANK() function returns the rank of each row within the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM().
? C. Option C uses the DENSE_RANK() window function instead of the PERCENT_RANK() function. The DENSE_RANK() function returns the rank of each row within the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM().
? D. Option D uses the ROW_NUMBER() window function instead of the PERCENT_RANK() function. The ROW_NUMBER() function returns the sequential number of each rowwithin the partition, but not the percentage rank. Also, the query does not have a GROUP BY clause, which is required for aggregate functions like SUM(). References:
? 1: PERCENT_RANK (Transact-SQL)
? 2: Window functions in Databricks SQL
? 3: Databricks Certified Data Analyst Associate Exam Guide

NEW QUESTION 8
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 9
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 drogdown 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 10
A data analyst runs the following command: SELECT age, country
FROM my_table
WHERE age >= 75 AND country = 'canada';
Which of the following tables represents the output of the above command?
A)

age	country
80	canada
NULL	canada
90	NULL

B)

age	country
80	NULL
75	NULL
90	NULL

C)

id	age	country
900	80	canada
901	75	canada
902	90	canada

D)

age	country
80	canada
14	canada
90	canada

E)

age	country
80	canada
75	canada
90	canada

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: E

Explanation:
The SQL query provided is designed to filter out records from ??my_table?? where the age is 75 or above and the country is Canada. Since I can??t view the content of the links provided directly, I need to rely on the image attached to this question for context. Based on that, Option E (the image attached) represents a table with columns ??age?? and ??country??, showing records where age is 75 or above and country is Canada. References: The answer can be inferred from understanding SQL queries and their outputs as per Databricks documentation: Databricks SQL

NEW QUESTION 10
In which of the following situations will the mean value and median value of variable be meaningfully different?

A. When the variable contains no outliers
B. When the variable contains no missing values
C. When the variable is of the boolean type
D. When the variable is of the categorical type
E. When the variable contains a lot of extreme outliers

Answer: E

Explanation:
The mean value of a variable is the average of all the values in a data set, calculated by dividing the sum of the values by the number of values. The median value of a variable is the middle value of the ordered data set, or the average of the middle two values if the data set has an even number of values. The mean value is sensitive to outliers, which are values that are verydifferent from the rest of the data. Outliers can skew the mean value and make it less representative of the central tendency of the data. The median value is more robust to outliers, as it only depends on the middle values of the data. Therefore, when the variable contains a lot of extreme outliers, the mean value and the median value will be meaningfully different, as the mean value will be pulled towards the outliers, while the median value will remain close to the majority of the data1. References: Difference Between Mean and Median in Statistics (With Example) - BYJU??S

NEW QUESTION 15
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 19
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 23
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 28
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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