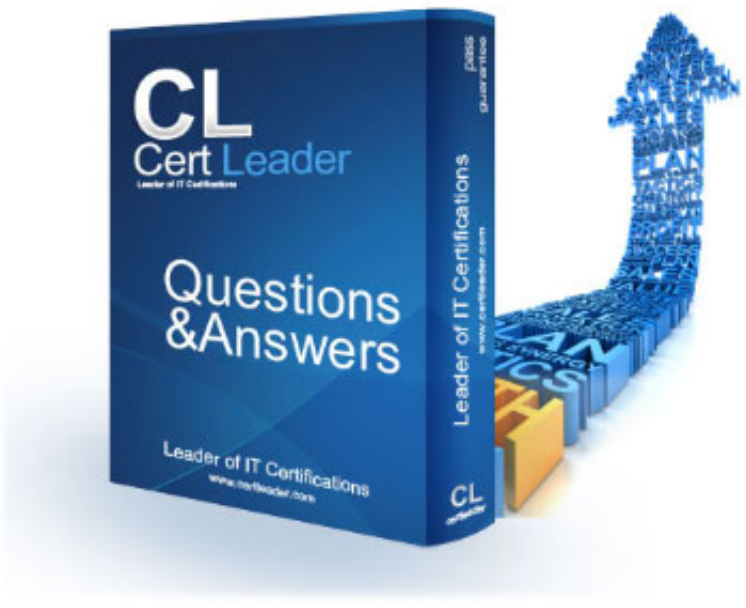
Databricks-Certified-Data-Analyst-Associate Dumps

Databricks Certified Data Analyst Associate Exam

https://www.certleader.com/Databricks-Certified-Data-Analyst-Associate-dumps.html



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, dataenrichment, 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
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 4
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 runningSELECT * 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 thestakeholders.eur_customersview?

A. The view remains available and SELECT * FROM stakeholders.eur_customers will execute correctly.
B. The view has been dropped.
C. The view is not available in the metastore, but the underlying data can be accessed with SELECT * FROM delt
D. `stakeholders.eur_customers`.
E. 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.
F. The view has been converted into a table.

Answer: B

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. References: 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 an ??

NEW QUESTION 5
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 6
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 7
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. Automatingcomplex 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 8
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 9
In which of the following situations should a data analyst use higher-order functions?

A. When custom logic needs to be applied to simple, unnested data
B. When custom logic needs to be converted to Python-native code
C. When custom logic needs to be applied at scale to array data objects
D. When built-in functions are taking too long to perform tasks
E. When built-in functions need to run through the Catalyst Optimizer

Answer: C

Explanation:
Higher-order functions are a simple extension to SQL to manipulate nested data such as arrays. A higher-order function takes an array, implements how the array is processed, and what the result of the computation will be. It delegates to a lambda function how to process each item in the array. This allows you to define functions that manipulate arrays in SQL, without having to unpack and repack them, use UDFs, or rely on limited built-in functions. Higher-order functions provide a performance benefit over user defined functions. References: Higher-order functions | Databricks on AWS, Working with Nested Data Using Higher Order Functions in SQL on Databricks | Databricks Blog, Higher-order functions - Azure Databricks | Microsoft Learn, Optimization recommendations on Databricks | Databricks on AWS

NEW QUESTION 10
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 quer
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 Statu
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 10
Which of the following layers of the medallion architecture is most commonly used by data analysts?

A. None of these layers are used by data analysts
B. Gold
C. All of these layers are used equally by data analysts
D. Silver
E. Bronze

Answer: B

Explanation:
The gold layer of the medallion architecture contains data that is highly refined and aggregated, and powers analytics, machine learning, and production applications. Data analysts typically use the gold layer to access data that has been transformed into knowledge, rather than just information. The gold layer represents the final stage of data quality and optimization in the lakehouse. References: What is the medallion lakehouse architecture?

NEW QUESTION 13
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 18
.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your Databricks-Certified-Data-Analyst-Associate Exam with Our Prep Materials Via below:

https://www.certleader.com/Databricks-Certified-Data-Analyst-Associate-dumps.html