

GAQM.Databricks-Certified-Data-Engineer-Associate.v2026-06-09.q113

Exam Code:	Databricks-Certified-Data-Engineer-Associate
Exam Name:	Databricks Certified Data Engineer Associate Exam
Certification Provider:	GAQM
Free Question Number:	113
Version:	v2026-06-09
# of views:	102
# of Questions views:	1132
https://www.freecram.net/torrent/GAQM.Databricks-Certified-Data-Engineer-Associate.v2026-06-09.q113.html	

NEW QUESTION: 1

A data engineer has been given a new record of data:

id STRING = 'a1'

rank INTEGER = 6

rating FLOAT = 9.4

Which of the following SQL commands can be used to append the new record to an existing Delta table my_table?

- A. UPDATE VALUES ('a1', 6, 9.4) my_table
- B. my_table UNION VALUES ('a1', 6, 9.4)
- C. INSERT VALUES ('a1', 6, 9.4) INTO my_table
- D. UPDATE my_table VALUES ('a1', 6, 9.4)
- E. INSERT INTO my_table VALUES ('a1', 6, 9.4)

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 2

A data engineer is maintaining a data pipeline. Upon data ingestion, the data engineer notices that the source data is starting to have a lower level of quality. The data engineer would like to automate the process of monitoring the quality level.

Which of the following tools can the data engineer use to solve this problem?

- A. Unity Catalog
- B. Auto Loader
- C. Delta Live Tables
- D. Delta Lake
- E. Data Explorer

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 3

What are the transformations typically included in building the Bronze layer?

- A. Include columns Load date/time, process ID

- B. Business rules and transformations
- C. Perform extensive data cleansing
- D. Aggregate data from multiple sources

Answer: ([SHOW ANSWER](#))

The Bronze layer stores raw ingested data with minimal transformations, typically adding metadata columns such as load date/time and process ID for traceability and auditing.

NEW QUESTION: 4

A data engineer is decommissioning a sandbox schema in Unity Catalog. Some tables are ephemeral staging outputs that can be safely removed entirely, but a few tables point at shared cloud storage used by downstream jobs outside Databricks. The engineer must avoid deleting any shared files when cleaning up catalog objects.

How does Unity Catalog behave when dropping Managed vs External tables?

- A. DROP all tables; Databricks will only remove metadata for both managed and external tables.
- B. DROP managed tables that are ephemeral and DROP external tables; files for both remain for 7 days.
- C. DROP managed staging tables to remove data and metadata, and DROP external tables to remove only metadata
- D. DROP external tables first to delete their files, then DROP managed tables to keep their files for recovery.

Answer: ([SHOW ANSWER](#))

In Unity Catalog, dropping a managed table deletes both the table metadata and the underlying data files, while dropping an external table removes only the catalog metadata and leaves the referenced cloud storage data intact.

NEW QUESTION: 5

A data engineer is writing Spark code to group sales data by region and calculate total revenue for each region. Which Spark DataFrame transformation performs grouping operations?

- A. orderBy
- B. groupBy
- C. filter
- D. select

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 6

A data engineer has realized that they made a mistake when making a daily update to a table.

They need to use Delta time travel to restore the table to a version that is 3 days old. However, when the data engineer attempts to time travel to the older version, they are unable to restore the data because the data files have been deleted.

Which of the following explains why the data files are no longer present?

- A. The VACUUM command was run on the table
- B. The TIME TRAVEL command was run on the table
- C. The DELETE HISTORY command was run on the table
- D. The OPTIMIZE command was run on the table
- E. The HISTORY command was run on the table

Answer: ([SHOW ANSWER](#))

The VACUUM command in Delta Lake is used to clean up and remove unnecessary data files that are no longer needed for time travel or query purposes. When you run VACUUM with certain retention settings, it can delete older data files, which might include versions of data that are older than the specified retention period. If the data engineer is unable to restore the table to a version that is 3 days old because the data files have been deleted, it's likely because the VACUUM command was run on the table, removing the older data files as part of data cleanup.

NEW QUESTION: 7

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table.

The code block used by the data engineer is below:

```
(spark.table("sales")
  .withColumn("avg_price", col("sales") / col("units"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("complete")
  .trigger(          )
  .table("new_sales")
)
```

If the data engineer only wants the query to execute a micro-batch to process data every 5 seconds, which of the following lines of code should the data engineer use to fill in the blank?

- A. trigger(continuous="5 seconds")
- B. trigger("5 seconds")
- C. trigger(processingTime="5 seconds")
- D. trigger(once="5 seconds")
- E. trigger()

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 8

In which of the following scenarios should a data engineer select a Task in the Depends On field of a new Databricks Job Task?

- A. When another task needs to fail before the new task begins
- B. When another task needs to be replaced by the new task
- C. When another task needs to use as little compute resources as possible
- D. When another task has the same dependency libraries as the new task
- E. When another task needs to successfully complete before the new task begins

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 9

A data engineer needs to apply custom logic to identify employees with more than 5 years of experience in array column employees in table stores. The custom logic should create a new column exp_employees that is an array of all of the employees with more than 5 years of experience for each row. In order to apply this custom logic at scale, the data engineer wants to use the FILTER higher-order function.

Which of the following code blocks successfully completes this task?

```
SELECT
  store_id,
  employees,
  FILTER (employees, years_exp > 5) AS exp_employees
FROM stores;
```

A.

```
SELECT
  store_id,
  employees,
  CASE WHEN employees.years_exp > 5 THEN employees
  ELSE NULL
  END AS exp_employees
```

B.

```
SELECT
  store_id,
  employees,
  FILTER (employees, i -> i.years_exp > 5) AS exp_employees
FROM stores;
```

C.

```
SELECT
  store_id,
  employees,
  FILTER (exp_employees, i -> i.years_exp > 5) AS exp_employees
FROM stores;
```

D.

```
SELECT
  store_id,
  employees,
  FILTER (exp_employees, years_exp > 5) AS exp_employees
FROM stores;
```

E.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 10

A Databricks workflow fails at the last stage due to an error in a notebook. This workflow runs daily. The data engineer fixes the mistake and wants to rerun the pipeline. This workflow is very costly and time-intensive to run. Which action should the data engineer do in order to minimise downtime and cost?

A. Re-run the entire workflow

- B. Repair run
- C. Restart the cluster
- D. Switch to another cluster

Answer: ([SHOW ANSWER](#))

Repair run in Databricks allows rerunning only the failed tasks in a workflow instead of re-running the entire pipeline, minimizing downtime and reducing costs.

NEW QUESTION: 11

A data engineer needs to create a table in Databricks using data from a CSV file at location /path/to/csv.

They run the following command:

```
CREATE TABLE new_table
_____
OPTIONS (
  header = "true",
  delimiter = "|"
)
LOCATION "path/to/csv"
```

Which of the following lines of code fills in the above blank to successfully complete the task?

- A. USING DELTA
- B. None of these lines of code are needed to successfully complete the task
- C. USING CSV
- D. FROM CSV
- E. FROM "path/to/csv"

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 12

Which of the following describes a scenario in which a data engineer will want to use a single- node cluster?

- A. When they are working with SQL within Databricks SQL
- B. When they are manually running reports with a large amount of data
- C. When they are running automated reports to be refreshed as quickly as possible
- D. When they are concerned about the ability to automatically scale with larger data
- E. When they are working interactively with a small amount of data

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 13

A data engineer is designing a streaming pipeline and wants to limit how long Spark maintains state information for aggregation queries. Which Structured Streaming feature defines how long late data can be processed?

- A. Watermark
- B. Cache
- C. Checkpoint
- D. Partition

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 14

An engineering manager uses a Databricks SQL query to monitor ingestion latency for each data source. The manager checks the results of the query every day, but they are manually rerunning the query each day and waiting for the results.

Which of the following approaches can the manager use to ensure the results of the query are updated each day?

- A. They can schedule the query to refresh every 1 day from the SQL endpoint's page in Databricks SQL.
- B. They can schedule the query to refresh every 1 day from the query's page in Databricks SQL.
- C. They can schedule the query to run every 1 day from the Jobs UI.
- D. They can schedule the query to refresh every 12 hours from the SQL endpoint's page in Databricks SQL.
- E. They can schedule the query to run every 12 hours from the Jobs UI.

Answer: (SHOW ANSWER)

NEW QUESTION: 15

A dataset has been defined using Delta Live Tables and includes an expectations clause:

```
CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON  
VIOLATION FAIL UPDATE
```

What is the expected behavior when a batch of data containing data that violates these constraints is processed?

- A. Records that violate the expectation are added to the target dataset and recorded as invalid in the event log.
- B. Records that violate the expectation are added to the target dataset and flagged as invalid in a field added to the target dataset.
- C. Records that violate the expectation are dropped from the target dataset and recorded as invalid in the event log.
- D. Records that violate the expectation are dropped from the target dataset and loaded into a quarantine table.
- E. Records that violate the expectation cause the job to fail.

Answer: (SHOW ANSWER)

NEW QUESTION: 16

A data engineer is attempting to drop a Spark SQL table `my_table` and runs the following command:

```
DROP TABLE IF EXISTS my_table;
```

After running this command, the engineer notices that the data files and metadata files have been deleted from the file system.

Which of the following describes why all of these files were deleted?

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

Answer: A ([LEAVE A REPLY](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here: <https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF** Special Discount Code: **freecram**)

NEW QUESTION: 17

Which of the following benefits is provided by the array functions from Spark SQL?

- A. An ability to work with an array of tables for procedural automation
- B. An ability to work with data in a variety of types at once
- C. An ability to work with complex, nested data ingested from JSON files
- D. An ability to work with time-related data in specified intervals
- E. An ability to work with data within certain partitions and windows

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 18

Which of the following data workloads will utilize a Gold table as its source?

- A. A job that cleans data by removing malformed records
- B. A job that queries aggregated data designed to feed into a dashboard
- C. A job that aggregates uncleaned data to create standard summary statistics
- D. A job that ingests raw data from a streaming source into the Lakehouse
- E. A job that enriches data by parsing its timestamps into a human-readable format

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 19

Differentiate between all-purpose clusters and jobs clusters.

A data engineering team has created a python notebook to load data from cloud storage, this job has been tested and now needs to be scheduled in production.

Which would be the best cluster to be used in this case?

- A. Any Unity Catalog-enabled cluster
- B. All purpose cluster
- C. Jobs Cluster
- D. Serverless SQL warehouse

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 20

A data engineer needs to develop integration tests for an ETL process and deploy a version- controlled, packaged workflow into production using an external job scheduler. Which tool should the data engineer use for this job?

- A. Databricks Connect
- B. Databricks Asset Bundles
- C. Databricks Command Line Interface
- D. Databricks Software Development Kit

Answer: ([SHOW ANSWER](#))

Databricks Asset Bundles enable packaging of workflows with version control, allowing integration testing and deployment into production while supporting orchestration by external job schedulers.

NEW QUESTION: 21

A data engineer has a Delta table containing old data files that are no longer referenced after several update and delete operations. To remove these unused files and reclaim storage space, which Delta Lake command should be executed?

- A. OPTIMIZE
- B. VACUUM
- C. DESCRIBE HISTORY
- D. CACHE TABLE

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 22

A global retail company sells products across multiple categories (e.g., Electronics, Clothing) and regions (e.g., North, South, East, West). The sales team has provided the data engineer with a PySpark dataframe named sales_df as below and the team wants the data engineer to analyze the sales data to help them make strategic decisions.

sales_df

product_id	category	sales_amount	region
1	Electronics	100	North
2	Clothing	200	South
3	Electronics	150	North
4	Clothing	300	South
5	Electronics	250	East
6	Clothing	400	West

Calculate the total sales amount for each product category and store the results in a new dataframe called category_sales. What will generate the expected result of category_sales?

category	total_sales_amount
Electronics	500
Clothing	900

A. category_sales =

```
sales_df.groupBy("category").agg(sum("sales_amount").alias("total_sales_amount"))
```

B. category_sales =

```
sales_df.sum("sales_amount").groupBy("category").alias("total_sales_amount"))
```

C. category_sales =

```
sales_df.agg(sum("sales_amount").groupBy("category").alias("total_sales_amount"))
```

D. category_sales =

```
sales_df.groupBy("region").agg(sum("sales_amount").alias("total_sales_amount"))
```

Answer: (SHOW ANSWER)

The correct approach is to group by the product category and then aggregate the sum of sales_amount, aliasing it as total_sales_amount. This produces the expected totals: 500 for Electronics and 900 for Clothing.

NEW QUESTION: 23

Which of the following code blocks will remove the rows where the value in column age is greater than 25 from the existing Delta table my_table and save the updated table?

A. UPDATE my_table WHERE age <= 25;

B. UPDATE my_table WHERE age > 25;

C. DELETE FROM my_table WHERE age <= 25;

D. SELECT * FROM my_table WHERE age > 25;

E. DELETE FROM my_table WHERE age > 25;

Answer: (SHOW ANSWER)

NEW QUESTION: 24

A data architect has determined that a table of the following format is necessary:

employeeId	startDate	avgRating
a1	2009-01-06	5.5
a2	2018-11-21	7.1
...

Which of the following code blocks uses SQL DDL commands to create an empty Delta table in the above format regardless of whether a table already exists with this name?

```
CREATE OR REPLACE TABLE table_name (
  employeeId STRING,
  startDate DATE,
  avgRating FLOAT
)
```

A.

```
CREATE OR REPLACE TABLE table_name WITH COLUMNS (  
  employeeId STRING,  
  startDate DATE,  
  avgRating FLOAT  
) USING DELTA
```

B.

```
CREATE TABLE IF NOT EXISTS table_name (  
  employeeId STRING,  
  startDate DATE,  
  avgRating FLOAT  
)
```

C.

```
CREATE OR REPLACE TABLE table_name AS  
SELECT  
  employeeId STRING,  
  startDate DATE,  
  avgRating FLOAT  
USING DELTA
```

D.

```
CREATE TABLE table_name AS  
SELECT  
  employeeId STRING,  
  startDate DATE,  
  avgRating FLOAT
```

E.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 25

A data engineer team has decided to implement a new data platform on Databricks and is currently deciding how to store each kind of data on each data layer. What is the appropriate layer and data pairing for medallion architecture?

- A. Silver Layer - Raw data from deposit account application
- B. Bronze Layer - Summary of cash deposit amount for each country and city
- C. Silver Layer - Cleansed master customer data
- D. Gold Layer - Deduplicated money transfer transaction

Answer: ([SHOW ANSWER](#))

In the Medallion architecture, the Silver layer holds cleansed and conformed data (such as master customer data), which has been cleaned, validated, and standardized after ingestion from the Bronze layer.

NEW QUESTION: 26

A data engineer wants to create a new table containing the names of customers that live in France.

They have written the following command:

```
CREATE TABLE customersInFrance
  AS
  SELECT id,
         firstName,
         lastName,
  FROM customerLocations
  WHERE country = 'FRANCE';
```

A senior data engineer mentions that it is organization policy to include a table property indicating that the new table includes personally identifiable information (PII).

Which of the following lines of code fills in the above blank to successfully complete the task?

- A. "COMMENT PII"
- B. There is no way to indicate whether a table contains PII.
- C. TBLPROPERTIES PII
- D. COMMENT "Contains PII"
- E. PII

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 27

A data engineer must deliver a trustworthy customer 360 dataset in Databricks for data scientists and BI teams. The engineer plans to join deduplicated customer records with cleaned transaction data, enforce schema and data quality checks, and create a conformed "customer_transactions" view. Later, highly aggregated, domain-specific tables (for weekly spend and executive summaries) will be produced for dashboards.

Where should the engineer build the conformed "customer_transactions" dataset, and where should the aggregated, report-ready tables reside?

- A. Build both "customer_transactions" and aggregated, report-ready tables in Silver to keep the model simpler.
- B. Build "customer_transactions" in Bronze and put the aggregated, report-ready tables in Silver.
- C. Build "customer_transactions" in Gold and put the aggregated, report-ready tables in Silver.
- D. Build "customer_transactions" in Silver and put the aggregated, report-ready tables in Gold.

Answer: D ([LEAVE A REPLY](#))

The conformed customer_transactions dataset belongs in Silver because it represents cleaned, deduplicated, and validated data that is suitable for broad analytical use. Aggregated, domain-specific, and report-ready tables are best placed in Gold, as this layer is designed for highly curated datasets optimized for BI, dashboards, and executive reporting.

NEW QUESTION: 28

Which of the following describes when to use the CREATE STREAMING LIVE TABLE (formerly CREATE INCREMENTAL LIVE TABLE) syntax over the CREATE LIVE TABLE syntax when creating Delta Live Tables (DLT) tables using SQL?

- A. CREATE STREAMING LIVE TABLE should be used when the subsequent step in the DLT pipeline is static.
- B. CREATE STREAMING LIVE TABLE is redundant for DLT and it does not need to be used.
- C. CREATE STREAMING LIVE TABLE should be used when the previous step in the DLT pipeline is static.

D. CREATE STREAMING LIVE TABLE should be used when data needs to be processed incrementally.

E. CREATE STREAMING LIVE TABLE should be used when data needs to be processed through complicated aggregations.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 29

A data engineer needs to apply custom logic to string column city in table stores for a specific use case. In order to apply this custom logic at scale, the data engineer wants to create a SQL user- defined function (UDF).

Which of the following code blocks creates this SQL UDF?

```
CREATE UDF combine_nyc(city STRING)
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
```

A. END;

```
CREATE UDF combine_nyc(city STRING)
RETURNS STRING
CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

B.

```
CREATE UDF combine_nyc(city STRING)
RETURNS STRING
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
```

C. END;

```
CREATE FUNCTION combine_nyc(city STRING)
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
```

D. END;

```
CREATE FUNCTION combine_nyc(city STRING)
RETURNS STRING
RETURN CASE
  WHEN city = "brooklyn" THEN "new york"
  ELSE city
END;
```

E.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 30

An organization has data stored across multiple external systems, including MySQL, Amazon Redshift, and Google BigQuery. The data engineer wants to perform analytics without ingesting directly into Databricks, ensuring unified governance and minimizing data duplication. Which feature of Databricks enables querying these external data sources while maintaining centralized governance?

A. Delta Lake

B. Lakehouse Federation

C. MLflow

D. Databricks Connect

Answer: (SHOW ANSWER)

Lakehouse Federation allows Databricks to query external data sources like MySQL, Redshift, and BigQuery without moving or duplicating data, while maintaining centralized governance through Unity Catalog.

NEW QUESTION: 31

A dataset has been defined using Delta Live Tables and includes an expectations clause:

```
CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON
```

```
VIOLATION DROP ROW
```

What is the expected behavior when a batch of data containing data that violates these constraints is processed?

- A. Records that violate the expectation are added to the target dataset and flagged as invalid in a field added to the target dataset.
- B. Records that violate the expectation are dropped from the target dataset and loaded into a quarantine table.
- C. Records that violate the expectation are dropped from the target dataset and recorded as invalid in the event log.
- D. Records that violate the expectation are added to the target dataset and recorded as invalid in the event log.
- E. Records that violate the expectation cause the job to fail.

Answer: (SHOW ANSWER)

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here: <https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special Discount Code: freecram**)

NEW QUESTION: 32

A Data Engineer is designing Bronze layer in Databricks Medallion Architecture. The raw data is collected from multiple sources (Clickstream in JSON, Transactions in CSV).

The task is to design the Bronze layer of the Medallion Architecture to ingest and store this raw data for further processing.

Which operation applies to the Bronze layer?

- A. Ingest raw data without transformations, preserving original schemas, and store in Delta format.
- B. Clean and standardize raw data by removing null values and enforcing schemas.
- C. Apply complex business logic to enrich raw data with customer segmentation labels.
- D. Aggregate and transform source data to calculate daily sales performance metrics.

Answer: A (LEAVE A REPLY)

In the Bronze layer of the Medallion Architecture, raw data from multiple sources is ingested without transformations while preserving its original schema. The data is then stored in Delta format, serving as the single source of truth for downstream Silver and Gold layers where cleaning, standardization, and business logic are applied.

NEW QUESTION: 33

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table.

The code block used by the data engineer is below:

```
(spark.readStream
  .table("sales")
  .withColumn("avg_price", col("sales") / col("units"))
  .writeStream
  .option("checkpointLocation", checkpointPath)
  .outputMode("complete")
  .trigger(_____)
  .table("new_sales")
)
```

The data engineer only wants the query to process all of the available data in as many batches as required.

Which line of code should the data engineer use to fill in the blank?

- A. trigger(continuous="once")
- B. trigger(processingTime="once")
- C. trigger(availableNow=True)
- D. trigger(once=True)

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 34

Which of the following commands will return the number of null values in the member_id column?

- A. SELECT count(member_id) - count_null(member_id) FROM my_table;
- B. SELECT null(member_id) FROM my_table;
- C. SELECT count_if(member_id IS NULL) FROM my_table;
- D. SELECT count(member_id) FROM my_table;
- E. SELECT count_null(member_id) FROM my_table;

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 35

What Databricks feature can be used to check the data sources and tables used in a workspace?

- A. Use the lineage feature to visualize a graph that highlights where the table is used only in reports.
- B. Use the lineage feature to visualize a graph that highlights where the table is used only in notebooks.
- C. Use the lineage feature to visualize a graph that shows all dependencies, including where the table is used in notebooks, other tables, and reports.
- D. Do not use the lineage feature as it only tracks activity from the last 3 months and will not provide full details on dependencies.

Answer: ([SHOW ANSWER](#))

The lineage feature in Unity Catalog provides a complete visualization of data dependencies, showing how a table is used across notebooks, reports, and other tables. This helps track data usage, dependencies, and impact analysis for governance and auditing.

NEW QUESTION: 36

A data engineering team has noticed that their Databricks SQL queries are running too slowly when they are submitted to a non-running SQL endpoint. The data engineering team wants this issue to be resolved.

Which of the following approaches can the team use to reduce the time it takes to return results in this scenario?

- A. They can turn on the Serverless feature for the SQL endpoint and change the Spot Instance Policy to "Reliability Optimized."
- B. They can increase the cluster size of the SQL endpoint.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can turn on the Serverless feature for the SQL endpoint.
- E. They can increase the maximum bound of the SQL endpoint's scaling range.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 37

A data engineer is implementing a data ingestion pipeline that automatically detects schema changes in incoming JSON files and evolves the target Delta table accordingly. Which feature supports this automatic schema evolution?

- A. Broadcast join
- B. Partition pruning
- C. Schema evolution
- D. Schema enforcement

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 38

A data engineer is setting up a new Databricks pipeline that ingests clickstream events from Kafka and daily product catalogs from cloud object storage. To ensure auditability and easy reprocessing, the engineer wants to land all source data. Later stages will handle cleaning, deduplication, and business modeling before the data is used in dashboards.

Which approach aligns to Medallion Architecture principles?

- A. Land both sources in Gold with denormalized star schemas to optimize BI while retaining full source fidelity.
- B. Land streaming events from Kafka in silver, and the product catalog directly in Gold to minimize layers for batch data.
- C. Land both sources in the Bronze layer append-only with minimal validation, then build Silver/Gold downstream for quality and analytics.
- D. Land both sources directly into the Silver layer with schema enforcement and deduplication to reduce downstream complexity.

Answer: ([SHOW ANSWER](#))

The Medallion Architecture recommends landing all raw source data in the Bronze layer in an append-only, minimally transformed form to preserve source fidelity and support auditability and reprocessing, with cleaning, deduplication, and business modeling performed in downstream Silver and Gold layers.

NEW QUESTION: 39

A data engineer wants to run SQL queries directly on data stored in Delta Lake using a serverless compute environment designed for analytics workloads. Which Databricks component provides this capability?

- A. Spark Driver
- B. DBFS
- C. Databricks SQL Warehouse
- D. MLflow

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 40

A data engineer has a Python notebook in Databricks, but they need to use SQL to accomplish a specific task within a cell. They still want all of the other cells to use Python without making any changes to those cells.

Which of the following describes how the data engineer can use SQL within a cell of their Python notebook?

- A. They can simply write SQL syntax in the cell
- B. They can attach the cell to a SQL endpoint rather than a Databricks cluster
- C. It is not possible to use SQL in a Python notebook
- D. They can change the default language of the notebook to SQL
- E. They can add %sql to the first line of the cell

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 41

A data engineer wants to delete records from a Delta table where the status column equals "inactive". The operation should modify only the affected rows without rewriting the entire dataset.

Which SQL command should be used?

- A. DELETE
- B. INSERT
- C. MERGE
- D. UPDATE

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 42

A data engineer wants to schedule multiple tasks in a workflow where each task depends on the successful completion of previous tasks. The workflow should support retries and monitoring.

Which Databricks feature should be used?

- A. DBFS
- B. Databricks Jobs
- C. Delta Lake
- D. Spark UI

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 43

Which of the following Git operations must be performed outside of Databricks Repos?

- A. Merge
- B. Commit
- C. Pull
- D. Clone
- E. Push

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 44

A data engineer needs to create a table in Databricks using data from their organization's existing SQLite database.

They run the following command:

```
CREATE TABLE jdbc_customer360
USING _____
OPTIONS (
  url "jdbc:sqlite:/customers.db",
  dbtable "customer360"
)
```

Which of the following lines of code fills in the above blank to successfully complete the task?

- A. org.apache.spark.sql.jdbc
- B. sqlite
- C. autoloader
- D. org.apache.spark.sql.sqlite
- E. DELTA

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 45

A data engineer wants to create an external table in Databricks that references data stored in an Azure Data Lake Storage (ADLS) location. The goal is to enable Databricks to access and query this external data without moving it into the Databricks-managed storage. Which step should the data engineer take to successfully create the external table?

- A. Use the CREATE MANAGED TABLE statement and specify the LOCATION clause with the path to the external data.
- B. CREATE UNMANAGED TABLE statement without specifying a LOCATION clause.
- C. Use the CREATE TABLE statement and specify the LOCATION clause with the path to the external data.
- D. CREATE EXTERNAL TABLE statement without specifying a LOCATION clause.

Answer: ([SHOW ANSWER](#))

To reference data stored outside of Databricks-managed storage, the engineer should use CREATE TABLE ... LOCATION 'path', which creates an unmanaged (external) table pointing to the ADLS data without moving it into Databricks storage.

NEW QUESTION: 46

A data engineer notices that a Spark job performing a join between a large table and a small lookup table is slow. The lookup table is only a few megabytes. Which Spark optimization technique should be applied to improve the performance of the join operation?

- A. Shuffle join
- B. Broadcast join
- C. Cartesian join
- D. Sort merge join

Answer: ([SHOW ANSWER](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected**

get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special**

Discount Code: **freecram**)

NEW QUESTION: 47

A data engineer wants to create a data entity from a couple of tables. The data entity must be used by other data engineers in other sessions. It also must be saved to a physical location.

Which of the following data entities should the data engineer create?

- A. Table
- B. Database
- C. View
- D. Function
- E. Temporary view

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 48

Which of the following describes the storage organization of a Delta table?

- A. Delta tables are stored in a collection of files that contain data, history, metadata, and other attributes.
- B. Delta tables store their data in a single file and all metadata in a collection of files in a separate location.
- C. Delta tables are stored in a single file that contains data, history, metadata, and other attributes.
- D. Delta tables are stored in a single file that contains only the data stored within the table.
- E. Delta tables are stored in a collection of files that contain only the data stored within the table.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 49

A data engineer needs to ingest streaming data from Kafka into a Delta Lake table using Databricks. The pipeline must guarantee exactly-once processing semantics and maintain checkpoint information to recover from failures. Which Spark Structured Streaming feature ensures reliable processing and fault tolerance?

- A. Checkpointing
- B. Watermarking
- C. Caching
- D. Partition pruning

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 50

A Databricks single-task workflow fails at the last task due to an error in a notebook. The data engineer fixes the mistake in the notebook. What should the data engineer do to rerun the workflow?

- A. Repair the task
- B. Rerun the pipeline
- C. Restart the cluster
- D. Switch the cluster

Answer: ([SHOW ANSWER](#))

For a single-task workflow, the correct action is to repair the task, which reruns only the failed task after fixing the error, avoiding unnecessary re-execution of the whole workflow.

NEW QUESTION: 51

A data engineer has written a function in a Databricks Notebook to calculate the population of bacteria in a given medium.

```
def calculate_population_of_bacteria (intital_population, exponential_factor)
    return future_population = intital_population ** exponential_factor
```

Analysts use this function in the notebook and sometimes provide input arguments of the wrong data type, which can cause errors during execution.

Which Databricks feature will help the data engineer quickly identify if an incorrect data type has been provided as input?

- A. The Spark User interface has a debug tab that contains the variables that are used in this session.
- B. The Databricks debugger enables breakpoints that will raise an error if the wrong data type is submitted.
- C. The Databricks debugger enables the use of a variable explorer to see at a glance the value of the variables.
- D. The Data Engineer should add print statements to find out what the variable is.

Answer: ([SHOW ANSWER](#))

The Databricks debugger supports setting breakpoints that pause execution and allow inspection of variables. If the wrong data type is passed to the function, the debugger raises an error at runtime, helping the engineer quickly identify the issue.

NEW QUESTION: 52

In which of the following scenarios should a data engineer use the MERGE INTO command instead of the INSERT INTO command?

- A. When the target table cannot contain duplicate records
- B. When the source is not a Delta table
- C. When the location of the data needs to be changed
- D. When the source table can be deleted
- E. When the target table is an external table

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 53

A data engineer wants to reduce costs and optimize cloud spending. The data engineer has decided to use Databricks Serverless for lowering cloud costs while maintaining existing SLAs.

What is the first step in migrating to Databricks Serverless?

- A. Legacy Ingestion pipelines that include ingestion from sources API's, files, JDBC/ODBC connections
- B. A frequently running and efficient Python-based data transformation pipeline compatible with the latest Databricks runtime and Unity Catalog
- C. A frequently running and efficient Scala-based data transformation pipeline compatible with the latest Databricks runtime and Unity Catalog
- D. Low frequency BI Dashboarding and Adhoc SQL Analytics

Answer: ([SHOW ANSWER](#))

The first workloads to migrate to Databricks Serverless should be frequently running, efficient pipelines that are already compatible with the latest Databricks runtime and Unity Catalog. This ensures smooth migration while maintaining SLAs, and Python-based pipelines are the most common and fully supported starting point.

NEW QUESTION: 54

In order for Structured Streaming to reliably track the exact progress of the processing so that it can handle any kind of failure by restarting and/or reprocessing, which of the following two approaches is used by Spark to record the offset range of the data being processed in each trigger?

- A. Structured Streaming cannot record the offset range of the data being processed in each trigger.
- B. Replayable Sources and Idempotent Sinks
- C. Checkpointing and Write-ahead Logs
- D. Write-ahead Logs and Idempotent Sinks
- E. Checkpointing and Idempotent Sinks

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 55

A data engineer is working on a personal laptop and needs to perform complex transformations on data stored in a Delta Lake on cloud storage. The engineer decides to use Databricks Connect to interact with Databricks clusters and work in their local IDE. How does Databricks Connect enable the engineer to develop, test, and debug code seamlessly on their local machine while interacting with Databricks clusters?

- A. By providing a local environment that mimics the Databricks runtime, enabling the engineer to develop, test, and debug code using a specific IDE that is required by Databricks
- B. By providing a local environment that mimics the Databricks runtime, enabling the engineer to develop, test, and debug code only through Databricks' own web interface
- C. By allowing direct execution of Spark jobs from the local machine without needing a network connection
- D. By providing a local environment that mimics the Databricks runtime, enabling the engineer to develop, test, and debug code using their preferred IDE

Answer: ([SHOW ANSWER](#))

Databricks Connect lets engineers use their preferred local IDE while connecting to Databricks clusters remotely. It mimics the Databricks runtime locally, enabling seamless development, testing, and debugging while executing workloads on the Databricks cluster.

NEW QUESTION: 56

A data engineer is working in a Python notebook on Databricks to process data, but notices that the output is not as expected. The data engineer wants to investigate the issue by stepping through the code and checking the values of certain variables during execution. Which tool should the data engineer use to inspect the code execution and variables in real-time?

- A. Cluster Logs
- B. Job Execution Dashboard
- C. Python Notebook Interactive Debugger
- D. SQL Analytics

Answer: ([SHOW ANSWER](#))

The Python Notebook Interactive Debugger in Databricks allows stepping through code execution, setting breakpoints, and inspecting variable values in real-time. This makes it the correct tool for debugging Python code inside a Databricks notebook.

NEW QUESTION: 57

A data engineering team has two tables. The first table `march_transactions` is a collection of all retail transactions in the month of March. The second table `april_transactions` is a collection of all retail transactions in the month of April. There are no duplicate records between the tables.

Which of the following commands should be run to create a new table `all_transactions` that contains all records from `march_transactions` and `april_transactions` without duplicate records?

- A.** `CREATE TABLE all_transactions AS
SELECT * FROM march_transactions
UNION SELECT * FROM april_transactions;`
- B.** `CREATE TABLE all_transactions AS
SELECT * FROM march_transactions
INTERSECT SELECT * FROM april_transactions;`
- C.** `CREATE TABLE all_transactions AS
SELECT * FROM march_transactions
INNER JOIN SELECT * FROM april_transactions;`
- D.** `CREATE TABLE all_transactions AS
SELECT * FROM march_transactions
MERGE SELECT * FROM april_transactions;`
- E.** `CREATE TABLE all_transactions AS
SELECT * FROM march_transactions
OUTER JOIN SELECT * FROM april_transactions;`

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 58

A data engineer is working on a Databricks project that utilizes cloud storage. The data engineer wants to load several JSON files from containers on a storage account as soon as the file arrives within the storage account. Which syntax should the data engineer follow to first load the files into a dataframe and check that it is working as expected using Python?

- A.** `df = spark.read.json("input/path")`
- B.** `df = spark.readStream.format("cloud").option("json").load("/input/path")`
- C.** `df = spark.readStream.format("json").load("input/path")`
- D.** `df = spark.readStream.format("cloudFiles").option("cloudFiles.format", "json").load("/input/path")`

Answer: ([SHOW ANSWER](#))

To automatically ingest JSON files as soon as they arrive in cloud storage, the correct syntax is to use Auto Loader with `spark.readStream.format("cloudFiles").option("cloudFiles.format", "json").load("/input/path")`. This enables continuous ingestion of new files into a streaming dataframe.

NEW QUESTION: 59

A data engineer wants to create a relational object by pulling data from two tables. The relational object does not need to be used by other data engineers in other sessions. In order to save on storage costs, the data engineer wants to avoid copying and storing physical data.

Which of the following relational objects should the data engineer create?

- A. Spark SQL Table
- B. Delta Table
- C. View
- D. Database
- E. Temporary view

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 60

A data engineer has inherited a Databricks pipeline from a previous team. The pipeline is missing SLAs and costs more than the allotted budget. On analysis, it is noted that the cluster is not being fully utilized, and the dataset is getting skewed. How should the data engineer resolve this issue?

- A. Use coalesce() on the dataset to merge partitions and reduce skew.
- B. Increase the number of executors for the job.
- C. Repartition the dataset to have it be more optimally spread across all nodes.
- D. Increase the executor memory for the job.

Answer: ([SHOW ANSWER](#))

Repartitioning the dataset ensures data is evenly distributed across all nodes, reducing skew, improving cluster utilization, and helping control costs while meeting SLAs.

NEW QUESTION: 61

A data engineer is configuring a cluster in Databricks and wants to ensure that multiple Spark executors can run tasks concurrently across worker nodes. Which cluster component is responsible for executing tasks?

- A. Driver
- B. Notebook
- C. Executor
- D. Job scheduler

Answer: ([SHOW ANSWER](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special**

Discount Code: **freecram**)

NEW QUESTION: 62

A data engineer has been using a Databricks SQL dashboard to monitor the cleanliness of the input data to an ELT job. The ELT job has its Databricks SQL query that returns the number of input records containing unexpected NULL values. The data engineer wants their entire team to be notified via a messaging webhook whenever this value reaches 100.

Which of the following approaches can the data engineer use to notify their entire team via a messaging webhook whenever the number of NULL values reaches 100?

- A. They can set up an Alert without notifications.
- B. They can set up an Alert with one-time notifications.
- C. They can set up an Alert with a custom template.
- D. They can set up an Alert with a new email alert destination.
- E. They can set up an Alert with a new webhook alert destination.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 63

A data engineer needs to optimize the data layout and query performance for an e-commerce transactions Delta table. The table is partitioned by "purchase_date" a date column which helps with time-based queries but does not optimize searches on user statistics "customer_id", a high- cardinality column.

The table is usually queried with filters on "customer_id" within specific date ranges, but since this data is spread across multiple files in each partition, it results in full partition scans and increased runtime and costs.

How should the data engineer optimize the Data Layout for efficient reads?

- A. Alter table implementing liquid clustering on "customer_id" while keeping the existing partitioning.
- B. Alter the table implementing liquid clustering by "customer_id" and "purchase_date".
- C. Alter the table to partition by "customer_id".
- D. Enable delta caching on the cluster so that frequent reads are cached for performance.

Answer: ([SHOW ANSWER](#))

Since queries filter by customer_id within purchase_date ranges, partitioning only by purchase_date causes inefficient scans across partitions. The correct optimization is to use Liquid Clustering on both customer_id and purchase_date, which dynamically reorganizes data to optimize for both dimensions and avoids the drawbacks of high-cardinality static partitioning.

NEW QUESTION: 64

A data engineer needs to provide access to a group named manufacturing-team. The team needs privileges to create tables in the quality schema. Which set of SQL commands will grant a group named manufacturing-team to create tables in a schema named production with the parent catalog named manufacturing with the least privileges?

- A. GRANT CREATE TABLE ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT USE SCHEMA ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT USE CATALOG ON CATALOG manufacturing TO manufacturing-team;
- B. GRANT USE TABLE ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT USE SCHEMA ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT USE CATALOG ON CATALOG manufacturing TO manufacturing-team;
- C. GRANT CREATE TABLE ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT CREATE SCHEMA ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT CREATE CATALOG ON CATALOG manufacturing TO manufacturing-team;
- D. GRANT CREATE TABLE ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT CREATE SCHEMA ON SCHEMA manufacturing.quality TO manufacturing-team; GRANT USE CATALOG ON CATALOG manufacturing TO manufacturing-team;

Answer: ([SHOW ANSWER](#))

To allow a group to create tables in a schema with least privileges, they need:

- GRANT CREATE TABLE on the schema
- GRANT USE SCHEMA on the schema
- GRANT USE CATALOG on the parent catalog

This ensures proper access without granting unnecessary privileges like creating schemas or catalogs.

NEW QUESTION: 65

A data engineer is implementing a data pipeline using Databricks notebooks. The engineer wants to share parameters such as file paths and processing dates across multiple tasks within a job.

Which Databricks utility enables parameter passing?

- A. display
- B. dbutils.widgets
- C. spark.conf
- D. dbutils.fs

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 66

A data engineer needs to conduct Exploratory Analysis on data residing in a database that is within the company's custom-defined network in the cloud. The data engineer is using SQL for this task.

Which type of SQL Warehouse will enable the data engineer to process large numbers of queries quickly and cost-effectively?

- A. Classic SQL Warehouse
- B. Serverless SQL Warehouse
- C. Pro SQL Warehouse
- D. All-purpose compute cluster

Answer: (SHOW ANSWER)

A serverless SQL Warehouse automatically scales to handle large numbers of queries, optimizes performance without manual configuration, and charges only for actual usage, making it the most cost-effective and efficient option for exploratory SQL analysis.

NEW QUESTION: 67

A new data engineering team has been assigned to an ELT project. The new data engineering team will need full privileges on the database customers to fully manage the project.

Which of the following commands can be used to grant full permissions on the database to the new data engineering team?

- A. GRANT ALL PRIVILEGES ON DATABASE customers TO team;
- B. GRANT ALL PRIVILEGES ON DATABASE team TO customers;
- C. GRANT SELECT CREATE MODIFY USAGE PRIVILEGES ON DATABASE customers TO team;
- D. GRANT SELECT PRIVILEGES ON DATABASE customers TO teams;
- E. GRANT USAGE ON DATABASE customers TO team;

Answer: (SHOW ANSWER)

NEW QUESTION: 68

A data engineer is migrating pipeline tasks to reduce operational toil. The workspace used Unity Catalog and is in a region that supports serverless. The engineer wants Databricks to auto-select instance types, manage scaling, apply Photon, and handle runtime upgrades automatically for job runs.

How should the data engineer meet this requirement while adhering to Databricks constraints?

- A. Run the job on a serverless compute for workflows configuration, ensuring Unity Catalog is enabled and regional support is available.
- B. Use a Pro SQL warehouse and schedule Python notebook tasks to execute as pipeline steps.
- C. Create a job with a single-task job cluster and manually set instance families and min/max workers.
- D. Use an all-purpose cluster with cluster policies to enforce standard sizes and enable autoscaling.

Answer: ([SHOW ANSWER](#))

Serverless compute for workflows allows Databricks to automatically manage instance selection, scaling, Photon enablement, and runtime upgrades for job runs, provided Unity Catalog is enabled and the workspace region supports serverless.

NEW QUESTION: 69

A data engineer only wants to execute the final block of a Python program if the Python variable `day_of_week` is equal to 1 and the Python variable `review_period` is True.

Which of the following control flow statements should the data engineer use to begin this conditionally executed code block?

- A. `if day_of_week = 1 & review_period: = "True":`
- B. `if day_of_week == 1 and review_period == "True":`
- C. `if day_of_week = 1 and review_period:`
- D. `if day_of_week == 1 and review_period:`
- E. `if day_of_week = 1 and review_period = "True":`

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 70

A data engineering team wants to validate a new ingestion pipeline locally while ensuring large aggregations run in serverless compute in their Databricks workspace. They plan to use Databricks Connect and have the option to attach to either a shared cluster or serverless.

Which workspace requirement should be confirmed first to avoid connection failures?

- A. Verify that the workspace has Unity Catalog disabled and the Databricks Connect version is less than the serverless Runtime version.
- B. Verify that the workspace has Unity Catalog enabled and that the Databricks Connect version supports serverless for the target Runtime release.
- C. Verify that only assigned access mode clusters are used because serverless is not supported by Databricks Connect.
- D. Verify that the local Spark version equals the serverless Spark version to satisfy Spark Connect parity.

Answer: ([SHOW ANSWER](#))

Databricks Connect requires Unity Catalog to be enabled when connecting to serverless compute, and the Databricks Connect version must explicitly support serverless for the target Databricks Runtime to ensure compatibility and prevent connection failures.

NEW QUESTION: 71

Which of the following SQL keywords can be used to convert a table from a long format to a wide format?

- A. TRANSFORM
- B. PIVOT

- C. CONVERT
- D. WHERE
- E. SUM

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 72

A new data engineering team has been assigned to work on a project. The team will need access to database customers in order to see what tables already exist. The team has its own group team.

Which of the following commands can be used to grant the necessary permission on the entire database to the new team?

- A. GRANT CREATE ON DATABASE customers TO team;
- B. GRANT USAGE ON DATABASE customers TO team;
- C. GRANT USAGE ON CATALOG team TO customers;
- D. GRANT CREATE ON DATABASE team TO customers;
- E. GRANT VIEW ON CATALOG customers TO team;

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 73

A data analyst has a series of queries in a SQL program. The data analyst wants this program to run every day. They only want the final query in the program to run on Sundays. They ask for help from the data engineering team to complete this task.

Which of the following approaches could be used by the data engineering team to complete this task?

- A. They could submit a feature request with Databricks to add this functionality.
- B. They could automatically restrict access to the source table in the final query so that it is only accessible on Sundays.
- C. They could redesign the data model to separate the data used in the final query into a new table.
- D. They could only run the entire program on Sundays.
- E. They could wrap the queries using PySpark and use Python's control flow system to determine when to run the final query.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 74

A data engineer is configuring Unity Catalog in Databricks and needs to assign a role to a user who should have the ability to grant and revoke privileges on various data objects within a specific schema, but should not have read/write access over the schema or its objects.

Which role should the data engineer assign to this user?

- A. Table Owner
- B. Catalog Owner
- C. Schema Owner
- D. USE catalog/schema privilege on the schema

Answer: ([SHOW ANSWER](#))

The Schema Owner role in Unity Catalog allows a user to grant and revoke privileges on data objects within the schema without requiring read/write access to the schema's data itself, fitting the requirement.

NEW QUESTION: 75

A data engineer has a single-task Job that runs each morning before they begin working. After identifying an upstream data issue, they need to set up another task to run a new notebook prior to the original task.

Which of the following approaches can the data engineer use to set up the new task?

- A. They can clone the existing task in the existing Job and update it to run the new notebook.
- B. They can create a new job from scratch and add both tasks to run concurrently.
- C. They can create a new task in the existing Job and then add it as a dependency of the original task.
- D. They can clone the existing task to a new Job and then edit it to run the new notebook.
- E. They can create a new task in the existing Job and then add the original task as a dependency of the new task.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 76

A data engineer needs to append new records to an existing Delta table while preserving existing data. The ingestion pipeline runs every hour and adds incremental data without replacing previous records. Which write mode should be used?

- A. append
- B. errorIfExists
- C. overwrite
- D. ignore

Answer: ([SHOW ANSWER](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special**

Discount Code: freecram)

NEW QUESTION: 77

What is the functionality of AutoLoader in Databricks?

- A. Auto Loader automatically ingests and processes new files from cloud storage, handling both batch and streaming data with support for schema evolution.
- B. Auto Loader automatically ingests and processes new files from cloud storage, handling batch and streaming data with no support for schema evolution.
- C. Auto Loader automatically ingests and processes new files from cloud storage, handling only streaming data with no support for schema evolution.
- D. Auto Loader automatically ingests and processes new files from cloud storage, handling batch data with support for schema evolution.

Answer: ([SHOW ANSWER](#))

Auto Loader in Databricks automatically ingests new files from cloud storage, supports both batch and streaming ingestion, and handles schema evolution, making it ideal for scalable pipelines.

NEW QUESTION: 78

A data engineer needs to conduct Exploratory Analysis on data residing in a database that is within the company's custom-defined network in the cloud. The data engineer is using SQL for this task. Which type of SQL Warehouse will enable the data engineer to process large numbers of queries quickly and cost-effectively?

- A. Serverless compute for notebooks
- B. Pro SQL Warehouse
- C. Classic SQL Warehouse
- D. Serverless SQL Warehouse

Answer: [\(SHOW ANSWER\)](#)

A Pro SQL Warehouse is designed for high-performance, cost-effective query execution at scale.

It is optimized for running large volumes of queries quickly, making it ideal for exploratory analysis on enterprise datasets.

NEW QUESTION: 79

A data engineer needs to use a Delta table as part of a data pipeline, but they do not know if they have the appropriate permissions.

In which of the following locations can the data engineer review their permissions on the table?

- A. Jobs
- B. Dashboards
- C. Databricks Filesystem
- D. Repos
- E. Data Explorer

Answer: [E \(LEAVE A REPLY\)](#)

NEW QUESTION: 80

What can be used to simplify and unify siloed data architectures that are specialized for specific use cases?

- A. Delta Lake
- B. Data lake
- C. Data warehouse
- D. Data lakehouse

Answer: [\(SHOW ANSWER\)](#)

A data lakehouse is a unified architecture that combines the best features of data lakes and data warehouses. It simplifies and unifies siloed data architectures by allowing both structured and unstructured data to coexist in the same system, supporting a wide variety of use cases, such as analytics, machine learning, and business intelligence.

Data lakes are good for storing large amounts of raw data but lack the structure and performance needed for high-performance analytics.

Data warehouses are designed for structured data and analytics but aren't optimized for large volumes of unstructured or semi-structured data.

Delta Lake is a storage layer that brings reliability to data lakes, but a lakehouse is the broader architecture that unifies these approaches.

NEW QUESTION: 81

A data engineer is processing ingested streaming tables and needs to filter out NULL values in the order_datetime column from the raw streaming table orders_raw and store the results in a new table orders_valid using DLT. Which code snippet should the data engineer use?

```
CREATE OR REFRESH STREAMING TABLE orders_valid(
  CONSTRAINT valid_date
  EXPECT (order_datetime IS NOT NULL)
  ON VIOLATION DROP ROW
)
AS SELECT * FROM STREAM orders_raw;
```

A.

```
CREATE OR REFRESH STREAMING TABLE orders_valid
(
  CONSTRAINT valid_date EXPECT (order_datetime IS NOT NULL)
  ON VIOLATION DROP ROW
)
AS SELECT * FROM orders_raw;
```

B.

```
CREATE OR REFRESH STREAMING TABLE orders_valid
AS
SELECT *
FROM STREAM(orders_raw)
WHERE order_datetime IS NOT NULL;
```

C.

```
CREATE OR REPLACE STREAMING TABLE orders_valid
(
  FILTER (order_datetime IS NOT NULL)
)
AS SELECT * FROM STREAM(orders_raw);
```

D.

Answer: (SHOW ANSWER)

The correct approach is to use a DLT expectation with constraint CONSTRAINT valid_date EXPECT (order_datetime IS NOT NULL) ON VIOLATION DROP ROW This ensures rows with NULL order_datetime are filtered out automatically during streaming ingestion while keeping the pipeline valid.

NEW QUESTION: 82

A data engineer needs access to a table new_table, but they do not have the correct permissions.

They can ask the table owner for permission, but they do not know who the table owner is.

Which of the following approaches can be used to identify the owner of new_table?

- A. Review the Owner field in the table's page in Data Explorer
- B. Review the Owner field in the table's page in the cloud storage solution
- C. Review the Permissions tab in the table's page in Data Explorer
- D. All of these options can be used to identify the owner of the table
- E. There is no way to identify the owner of the table

Answer: (SHOW ANSWER)

NEW QUESTION: 83

A data engineer is standardizing repository layouts for multiple teams adopting Databricks Asset Bundles. The engineer wants to ensure every project has a single authoritative configuration file at the repository root that defines the bundle name, targets, workspace settings, permissions, and resource mappings (for jobs and pipelines).

What strategy should the data engineer use to meet the goal?

- A. Place multiple databricks.yml files under each subfolder (for example, jobs/pipelines/, workspace/) and merge them at deploy time using the include mapping.
- B. Place exactly one databricks.yml at the repository root; it is the main configuration file and may reference additional configuration files via the include mapping.
- C. Place a databricks.yml in a databricks/hidden folder at the repository root; only hidden locations are valid for bundle configs.
- D. Place a databricks.yaml at the repository root and optional databricks.yml in subfolders; the CLI prefers .yaml over .yml when both exist.

Answer: ([SHOW ANSWER](#))

Databricks Asset Bundles require a single authoritative databricks.yml file at the repository root that defines core bundle configuration, while allowing modularization by referencing additional files through the include mapping.

NEW QUESTION: 84

Which of the following statements regarding the relationship between Silver tables and Bronze tables is always true?

- A. Silver tables contain more data than Bronze tables.
- B. Silver tables contain a more refined and cleaner view of data than Bronze tables.
- C. Silver tables contain less data than Bronze tables.
- D. Silver tables contain a less refined, less clean view of data than Bronze data.
- E. Silver tables contain aggregates while Bronze data is unaggregated.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 85

A data engineering project involves processing large batches of data on a daily schedule using ETL. The jobs are resource-intensive and vary in size, requiring a scalable, cost-efficient compute solution that can automatically scale based on the workload. Which compute approach will satisfy the needs described?

- A. Job Cluster
- B. Dedicated Cluster
- C. All-Purpose Cluster
- D. Databricks SQL Serverless

Answer: ([SHOW ANSWER](#))

Job Clusters are ephemeral clusters created for scheduled or automated workloads. They are cost-efficient, scale automatically to match workload demands, and terminate after job completion, making them the ideal choice for daily large-batch ETL processing.

NEW QUESTION: 86

A data engineer has a Python variable `table_name` that they would like to use in a SQL query.

They want to construct a Python code block that will run the query using `table_name`.

They have the following incomplete code block:

```
_____(f"SELECT customer_id, spend FROM {table_name}")
```

Which of the following can be used to fill in the blank to successfully complete the task?

- A. spark.sql
- B. spark.delta.sql
- C. spark.delta.table
- D. dbutils.sql
- E. spark.table

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 87

A data engineer is attempting to write Python and SQL in the same command cell and is running into an error. The engineer thought that it was possible to use a Python variable in a select statement. Why does the command fail?

- A. Databricks supports language interoperability in the same cell but only between Scala and SQL.
- B. Databricks supports multiple languages but only one per notebook.
- C. Databricks supports one language per cell.
- D. Databricks supports language interoperability but only if a special character is used.

Answer: ([SHOW ANSWER](#))

In Databricks, each notebook cell supports only one language. While a notebook can contain multiple languages (Python, SQL, Scala, R), you must define the language per cell, so mixing Python and SQL directly in the same cell causes the error.

NEW QUESTION: 88

A data organization leader is upset about the data analysis team's reports being different from the data engineering team's reports. The leader believes the siloed nature of their organization's data engineering and data analysis architectures is to blame.

Which of the following describes how a data lakehouse could alleviate this issue?

- A. Both teams would be able to collaborate on projects in real-time
- B. Both teams would reorganize to report to the same department
- C. Both teams would use the same source of truth for their work
- D. Both teams would autoscale their work as data size evolves
- E. Both teams would respond more quickly to ad-hoc requests

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 89

A data engineer wants to read only a subset of columns from a Delta table to reduce memory usage and improve performance. Which Spark DataFrame operation should be used?

- A. filter
- B. select
- C. orderBy
- D. groupBy

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 90

A data engineer wants to store intermediate data in a temporary view for use within the same Spark session. The data should not persist after the session ends. Which Spark SQL command should be used?

- A. CREATE TABLE
- B. CREATE TEMP VIEW
- C. CREATE VIEW
- D. CREATE DATABASE

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 91

Which of the following describes a benefit of creating an external table from Parquet rather than CSV when using a CREATE TABLE AS SELECT statement?

- A. Parquet files have the ability to be optimized
- B. Parquet files can be partitioned
- C. Parquet files will become Delta tables
- D. Parquet files have a well-defined schema
- E. CREATE TABLE AS SELECT statements cannot be used on files

Answer: ([SHOW ANSWER](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special**

Discount Code: **freecram**)

NEW QUESTION: 92

A Python file is ready for production and the client wants to use the most efficient yet cost-effective type of cluster possible. The workload is quite small, only processing 10GBs of data with only simple joins and no complex aggregations or wide transformations.

Which cluster meets the requirement?

- A. Job cluster with spot instances enabled
- B. Job cluster with Photon enabled
- C. Job cluster with spot instances disabled
- D. Interactive cluster

Answer: ([SHOW ANSWER](#))

A job cluster is the most cost-effective option for production workloads because it is created only for the duration of the job and then terminated. Enabling spot instances further reduces cost, and the small, simple workload does not require advanced performance optimizations.

NEW QUESTION: 93

A data engineer is configuring a batch processing job in Databricks to read large Parquet datasets from cloud storage. The engineer wants Spark to automatically distribute computation across multiple worker nodes without manually assigning tasks. Which Spark abstraction enables automatic parallel processing?

- A. Executor
- B. Cluster
- C. DataFrame
- D. Driver

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 94

A new data engineering team has been assigned to an ELT project. The new data engineering team will need full privileges on the table sales to fully manage the project.

Which of the following commands can be used to grant full permissions on the database to the new data engineering team?

- A. GRANT ALL PRIVILEGES ON TABLE sales TO team;
- B. GRANT SELECT ON TABLE sales TO team;
- C. GRANT USAGE ON TABLE sales TO team;
- D. GRANT SELECT CREATE MODIFY ON TABLE sales TO team;
- E. GRANT ALL PRIVILEGES ON TABLE team TO sales;

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 95

An organization needs to share a dataset stored in its Databricks Unity Catalog with an external partner who uses a different data platform that is not Databricks. The goal is to maintain data security and ensure the partner can access the data efficiently. Which method should the data engineer use to securely share the dataset with the external partner?

- A. Using Delta Sharing with the open sharing protocol
- B. Exporting data as CSV files and emailing them
- C. Using a third-party API to access the Delta table
- D. Databricks-to-Databricks Sharing

Answer: ([SHOW ANSWER](#))

Delta Sharing with the open sharing protocol allows secure data sharing from Unity Catalog to external partners, even if they are not using Databricks, while maintaining governance and efficient access.

NEW QUESTION: 96

Which of the following describes a scenario in which a data team will want to utilize cluster pools?

- A. An automated report needs to be refreshed as quickly as possible.
- B. An automated report needs to be made reproducible.
- C. An automated report needs to be tested to identify errors.
- D. An automated report needs to be version-controlled across multiple collaborators.
- E. An automated report needs to be runnable by all stakeholders.

Answer: ([SHOW ANSWER](#))

Cluster pools in Databricks are used to ensure that a set of pre-warmed clusters is readily available to run workloads. This means that when a job is submitted, it can be executed more quickly because there is no need to wait for a cluster to spin up. Therefore, if a data team needs to refresh an automated report as quickly as possible, they will want to utilize cluster pools to ensure that the job can be executed as quickly as possible.

NEW QUESTION: 97

A data engineer notices that a Spark job repeatedly scans a large Delta table even though the dataset does not change during the session. Which technique can store the dataset in memory to speed up repeated queries?

- A. VACUUM
- B. MERGE
- C. OPTIMIZE
- D. CACHE TABLE

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 98

A data engineer is working with a very large dataset and wants to improve query performance by colocating related data in the same files. The engineer decides to apply Z-ordering on frequently filtered columns. Which Delta Lake command is used?

- A. ANALYZE TABLE
- B. CACHE TABLE
- C. OPTIMIZE ZORDER BY
- D. VACUUM ZORDER

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 99

A data engineer has been given a new record of data:

id STRING = 'a1'

rank INTEGER = 6

rating FLOAT = 9.4

Which of the following SQL commands can be used to append the new record to an existing Delta table my_table?

- A. INSERT VALUES ('a1' , 6, 9.4) INTO my_table
- B. UPDATE VALUES ('a1', 6, 9.4) my_table
- C. UPDATE my_table VALUES ('a1', 6, 9.4)
- D. my_table UNION VALUES ('a1', 6, 9.4)
- E. INSERT INTO my_table VALUES ('a1', 6, 9.4)

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 100

A data engineer has been using a Databricks SQL dashboard to monitor the cleanliness of the input data to a data analytics dashboard for a retail use case. The job has a Databricks SQL query that returns the number of store-level records where sales is equal to zero. The data engineer wants their entire team to be notified via a messaging webhook whenever this value is greater than 0. Which of the following approaches can the data engineer use to notify their entire team via a messaging webhook whenever the number of stores with \$0 in sales is greater than zero?

- A. They can set up an Alert with a custom template.
- B. They can set up an Alert without notifications.
- C. They can set up an Alert with one-time notifications.
- D. They can set up an Alert with a new email alert destination.
- E. They can set up an Alert with a new webhook alert destination.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 101

A data engineer is developing an ETL process based on Spark SQL. The execution fails. The data engineer checks the Spark UI and can see the ERRORS as follows:

"java.lang.OutOfMemoryError: Java heap space"

Which two corrective actions should the data engineer perform to resolve this issue? (Choose two.)

- A. Narrow the filters in order to collect less data in the query
- B. Upsize the worker nodes and activate autoshuffle partitions
- C. Upsize the driver node and deactivate autoshuffle partitions
- D. Cache the dataset in order to boost the query performance
- E. Fix the shuffle partitions to 50 to ensure the allocation

Answer: ([SHOW ANSWER](#))

Reducing input with narrower filters lowers memory pressure, and increasing executor resources (upsize worker nodes) plus enabling adaptive/auto shuffle partitioning provides more memory and better partitioning to avoid Java heap OOM during Spark SQL execution.

NEW QUESTION: 102

A data engineer is designing an ETL pipeline to process both streaming and batch data from multiple sources. The pipeline must ensure data quality, handle schema evolution, and provide easy maintenance. The team is considering using Delta Live Tables (DLT) in Databricks to achieve these goals. They want to understand the key features and benefits of DLT that make it suitable for this use case. Why is Delta Live Tables (DLT) an appropriate choice?

- A. Automatic data quality checks, built-in support for schema evolution, and declarative pipeline development
- B. Manual schema enforcement, high operational overhead, and limited scalability
- C. Requires custom code for data quality checks, no support for streaming data, and complex pipeline maintenance
- D. Supports only batch processing, no data versioning, and high infrastructure costs

Answer: ([SHOW ANSWER](#))

Delta Live Tables (DLT) is appropriate because it provides automatic data quality checks through expectations, built-in support for schema evolution, and a declarative approach to pipeline development, making it easier to build and maintain both streaming and batch ETL pipelines.

NEW QUESTION: 103

A data engineer is reading a very large dataset that contains skewed key values. Some partitions contain significantly more records than others, causing uneven workload distribution across executors. Which Spark issue does this scenario describe?

- A. Schema evolution
- B. Data skew
- C. Lazy evaluation
- D. Broadcast failure

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 104

A data engineer needs to use a Delta table as part of a data pipeline, but they do not know if they have the appropriate permissions. In which location can the data engineer review their permissions on the table?

- A. Jobs
- B. Dashboards
- C. Catalog Explorer
- D. Repos

Answer: ([SHOW ANSWER](#))

<https://docs.databricks.com/en/delta-live-tables/tutorial-pipelines.html>

NEW QUESTION: 105

A data engineer needs to access the view created by the sales team, using a shared cluster. The data engineer has been provided usage permissions on the catalog and schema. In order to access the view created by sales team. What are the minimum permissions the data engineer would require in addition?

- A. Needs SELECT permission only on the VIEW
- B. Needs SELECT permission on the VIEW and the underlying TABLE.
- C. Needs ALL PRIVILEGES on the VIEW
- D. Needs ALL PRIVILEGES at the SCHEMA level

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 106

A data engineer has a Job that has a complex run schedule, and they want to transfer that schedule to other Jobs.

Rather than manually selecting each value in the scheduling form in Databricks, which of the following tools can the data engineer use to represent and submit the schedule programmatically?

- A. There is no way to represent and submit this information programmatically
- B. datetime
- C. `pyspark.sql.types.TimestampType`
- D. `pyspark.sql.types.DateType`
- E. Cron syntax

Answer: ([SHOW ANSWER](#))

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF Special Discount Code: freecram**)

NEW QUESTION: 107

A data engineer needs to ingest from both streaming and batch sources for a firm that relies on highly accurate data. Occasionally, some of the data picked up by the sensors that provide a streaming input are outside the expected parameters. If this occurs, the data must be dropped, but the stream should not fail. Which feature of Delta Live Tables meets this requirement?

- A. Change Data Capture
- B. Error Handling
- C. Monitoring
- D. Expectations

Answer: (SHOW ANSWER)

Delta Live Tables Expectations allow defining data quality rules. Records that fail the expectations (such as sensor readings outside valid ranges) can be dropped without stopping the streaming pipeline, ensuring continuous and accurate processing.

NEW QUESTION: 108

A data engineer is onboarding a new bronze ingestion pipeline in Databricks with Unity Catalog.

The team wants Databricks to handle storage layout, apply platform optimizations over time, and simplify lifecycle management so that when a table is dropped, its underlying data is also cleaned up according to Databricks-managed retention policies.

Which table type should the data engineer create for these ingestion tables?

- A. Managed tables so that Unity Catalog manages both metadata and underlying data lifecycle.
- B. External tables with a LOCATION pointing to an external volume for full control of file layout.
- C. Foreign tables federated from an external catalog to delegate optimization to the source system.
- D. Temporary views over files to avoid table-level governance and lifecycle coupling.

Answer: (SHOW ANSWER)

Managed tables allow Unity Catalog and Databricks to control both metadata and underlying data storage, enabling automatic optimizations, simplified lifecycle management, and cleanup of data files when tables are dropped according to platform-managed policies.

NEW QUESTION: 109

A data engineer wants to ensure that duplicate records are removed from a dataset based on specific columns before writing to a Delta table.

Which Spark DataFrame method removes duplicate rows?

- A. removeDuplicates
- B. dropDuplicates
- C. uniqueRows
- D. distinctBy

Answer: (SHOW ANSWER)

NEW QUESTION: 110

Which SQL code snippet will correctly demonstrate a Data Definition Language (DDL) operation used to create a table?

- A. CREATE TABLE employees (
id INT,
name STRING
);
- B. DROP TABLE employees;
- C. ALTER TABLE employees ADD COLUMN salary DECIMAL(10,2);
- D. INSERT INTO employees (id, name) VALUES (1 'Alice');

Answer: ([SHOW ANSWER](#))

CREATE TABLE employees (...) is a Data Definition Language (DDL) operation that defines a new table and its schema.

NEW QUESTION: 111

A data engineer needs to reduce the number of small files generated during streaming ingestion into Delta tables. Small files are causing performance issues during downstream queries. Which Delta Lake feature automatically compacts small files during writes?

- A. VACUUM
- B. CACHE
- C. autoCompact
- D. OPTIMIZE

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 112

A data engineering team is using Kafka to capture event data and then ingest it into Databricks.

The team wants to be able to see these historical events. Medallion architecture is already in place. The team wants to be mindful of costs. Where should this historical event data be stored?

- A. Gold
- B. Silver
- C. Bronze
- D. Raw layer

Answer: ([SHOW ANSWER](#))

In the Medallion architecture, the Bronze layer stores raw ingested data, including historical event data from Kafka. This approach preserves the full history at lower cost while allowing further refinement into Silver and Gold layers.

NEW QUESTION: 113

A Databricks single-task workflow fails due to an error in a notebook. The data engineer fixes the mistake in the notebook.

What should the data engineer do to rerun the workflow?

- A. Repair the task
- B. Repair the run
- C. Restart the Cluster

D. Swap the cluster

Answer: ([SHOW ANSWER](#))

Rerunning a failed workflow after fixing the notebook requires repairing the run so the task is re-executed with the updated code, without recreating the entire workflow or changing cluster configuration.

Valid Databricks-Certified-Data-Engineer-Associate Dumps shared by EduDump.com for Helping Passing Databricks-Certified-Data-Engineer-Associate Exam! EduDump.com now offer the **newest Databricks-Certified-Data-Engineer-Associate exam dumps**, the EduDump.com Databricks-Certified-Data-Engineer-Associate exam **questions have been updated** and **answers have been corrected** get the **newest** EduDump.com Databricks-Certified-Data-Engineer-Associate dumps with Test Engine here:

<https://www.edudump.com/exams/GAQM/Databricks-Certified-Data-Engineer-Associate/premium/> (228 Q&As Dumps, **35%OFF** Special

Discount Code: [freecram](#))