Databricks Databricks-Certified-Professional-Data-Engineer Study Materials - Databricks-Certified-Professional-Data-Engineer Test Quiz

If you are going to take Databricks Databricks-Certified-Professional-Data-Engineer certification exam, it is essential to use Databricks-Certified-Professional-Data-Engineer training materials. If you are looking for reference materials without a clue, stop！ If you don't know what materials you should use, you can try Lead2PassExam Databricks Databricks-Certified-Professional-Data-Engineer exam dumps. The hit rate of the dumps is very high, which guarantees you can pass your exam with ease at the first attempt. Lead2PassExam Databricks Databricks-Certified-Professional-Data-Engineer Practice Test dumps can determine accurately the scope of the examination compared with other exam materials, which can help you improve efficiency of study and help you well prepare for Databricks-Certified-Professional-Data-Engineer exam.

Databricks-Certified-Professional-Data-Engineer exam is a specialized test that focuses on assessing the technical skillsets of candidates in working on cloud-based big data projects. Candidates will be required to demonstrate their proficiency in a wide range of topics, including data structures and algorithms, distributed systems, database design, Hadoop and Spark, and machine learning. Databricks-Certified-Professional-Data-Engineer Exam contains multiple-choice questions that test the candidates' knowledge of these areas.

>> Databricks Databricks-Certified-Professional-Data-Engineer Study Materials <<

100% Pass 2025 Fantastic Databricks-Certified-Professional-Data-Engineer:

Databricks Certified Professional Data Engineer Exam Study Materials

Do you want to pass Databricks-Certified-Professional-Data-Engineer exam and get the related certification within the minimum time and effort? If you would like to give me a positive answer, you really should keep a close eye on our website since you can find the best study material in here--our Databricks-Certified-Professional-Data-Engineer training materials. We have helped millions of thousands of candidates to prepare for the Databricks-Certified-Professional-Data-Engineer Exam and all of them have got a fruitful outcome, I wish you could be one of the beneficiaries of our training materials in the near future. The advantages of our Databricks-Certified-Professional-Data-Engineer test prep are more than you can imagine.

Databricks Certified Professional Data Engineer exam is a certification program offered by Databricks, a unified data analytics platform that provides a collaborative workspace for data science teams. Databricks Certified Professional Data Engineer Exam certification program is designed for data professionals who want to demonstrate their knowledge and skills in building reliable, scalable, and performant data pipelines using Databricks.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q31-Q36):

NEW QUESTION # 31
A data engineer is tasked with ensuring that a Delta table in Databricks continuously retains deleted files for 15 days (instead of the default 7 days), in order to permanently comply with the organization's data retention policy.
Which code snippet correctly sets this retention period for deleted files?

- A. from delta.tables import *
  deltaTable = DeltaTable.forPath(spark, "/mnt/data/my_table")
  deltaTable.deletedFileRetentionDuration = "interval 15 days"
- B. spark.sql("ALTER TABLE my_table SET TBLPROPERTIES ('delta.deletedFileRetentionDuration' = 'interval 15 days')")
- C. spark.sql("VACUUM my_table RETAIN 15 HOURS")
- D. spark.conf.set("spark.databricks.delta.deletedFileRetentionDuration", "15 days")

Answer: B

Explanation:
Comprehensive and Detailed
In Delta Lake, the property delta.deletedFileRetentionDuration controls how long deleted data files are retained before being permanently removed during a VACUUM operation.
By default, this retention duration is set to 7 days.
To comply with stricter retention requirements, organizations can explicitly update the table property using an ALTER TABLE statement.
Option A uses the correct SQL command:
ALTER TABLE my_table SET TBLPROPERTIES ('delta.deletedFileRetentionDuration' = 'interval 15 days') This updates the Delta table metadata so that all future operations respect the 15-day retention policy for deleted files.
Why not the others?
B: This code incorrectly tries to set the property via the DeltaTable API. Delta's Python API does not expose direct attributes like deletedFileRetentionDuration; instead, properties must be set through ALTER TABLE or DataFrameWriter options.
C: VACUUM ... RETAIN specifies a one-time file cleanup action (e.g., retaining 15 hours of history), not a persistent retention policy. It cannot be used to set a continuous retention duration.
D: Setting spark.conf applies a session-level configuration and does not permanently update the table's retention metadata. Once the session ends, this configuration is lost.
Therefore, Option A is the correct and documented approach for persistently enforcing a 15-day deleted file retention period in Delta Lake.

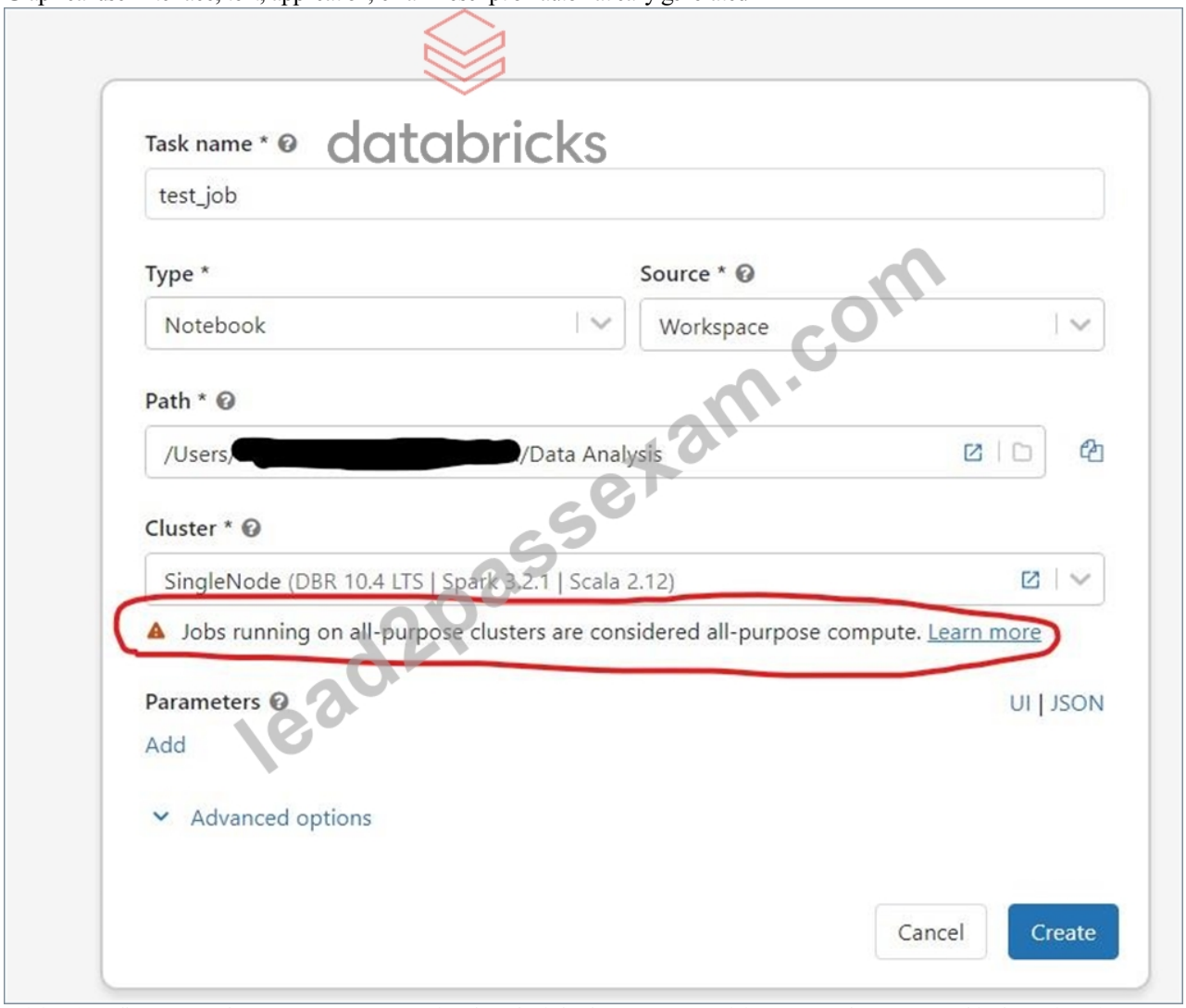
NEW QUESTION # 32
Your colleague was walking you through how a job was setup, but you noticed a warning message that said,
"Jobs running on all-purpose cluster are considered all purpose compute", the colleague was not sure why he was getting the warning message, how do you best explain this warning mes-sage?

- A. All-purpose clusters are less expensive than the job clusters
- B. All-purpose clusters are more expensive than the job clusters
- C. All-purpose clusters take longer to start the cluster vs a job cluster

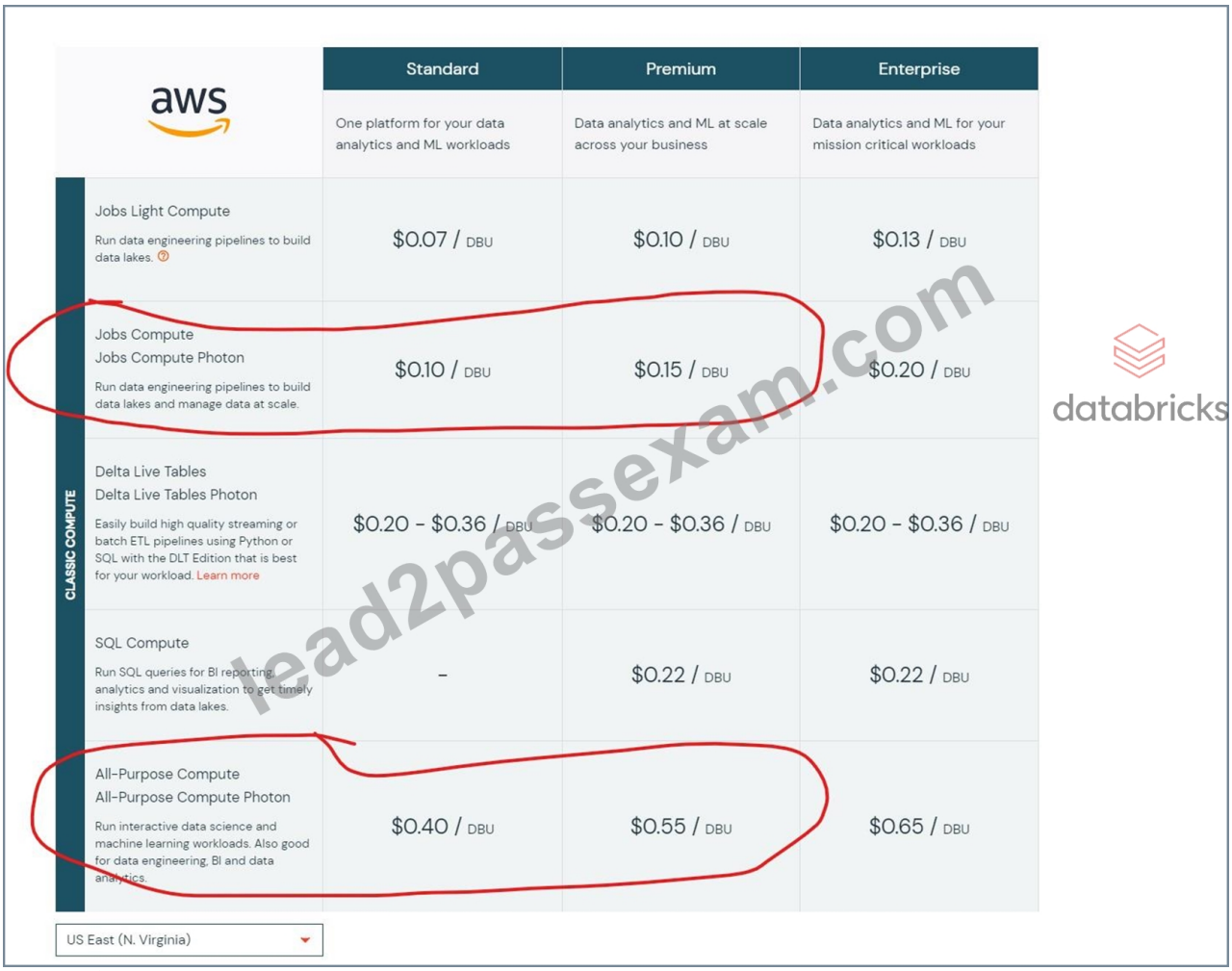
- D. All-purpose cluster provide interactive messages that can not be viewed in a job
- E. All-purpose clusters cannot be used for Job clusters, due to performance issues.

Answer: B

Explanation:
Explanation
Warning message:
Graphical user interface, text, application, email Description automatically generated



Pricing for All-purpose clusters are more expensive than the job clusters AWS pricing(Aug 15th 2022)Graphical user interface Description automatically generated



	Standard	Premium	Enterprise
	One platform for your data analytics and ML workloads	Data analytics and ML at scale across your business	Data analytics and ML for your mission critical workloads
CLASSIC COMPUTE — Jobs Light Compute: Run data engineering pipelines to build data lakes.	$0.07 / DBU	$0.10 / DBU	$0.13 / DBU
Jobs Compute, Jobs Compute Photon: Run data engineering pipelines to build data lakes and manage data at scale.	$0.10 / DBU	$0.15 / DBU	$0.20 / DBU
Delta Live Tables, Delta Live Tables Photon: Easily build high quality streaming or batch ETL pipelines using Python or SQL with the DLT Edition that is best for your workload. Learn more	$0.20 - $0.36 / DBU	$0.20 - $0.36 / DBU	$0.20 - $0.36 / DBU
SQL Compute: Run SQL queries for BI reporting, analytics and visualization to get timely insights from data lakes.	-	$0.22 / DBU	$0.22 / DBU
All-Purpose Compute, All-Purpose Compute Photon: Run interactive data science and machine learning workloads. Also good for data engineering, BI and data analytics.	$0.40 / DBU	$0.55 / DBU	$0.65 / DBU

US East (N. Virginia)

Bottom of Form
Top of Form

NEW QUESTION # 33
The view updates represents an incremental batch of all newly ingested data to be inserted or updated in the customers table.
The following logic is used to process these records.

```
MERGE INTO customers
USING (
  SELECT updates.customer_id as merge_ey, updates.*
  FROM updates

  UNION ALL

  SELECT NULL as merge_key, updates.*
  FROM updates JOIN customers
  ON updates.customer_id = customers.customer_id
  WHERE customers.current = true AND updates.address <> customers.address
) staged_updates
ON customers.customer_id = mergeKey
WHEN MATCHED AND customers.current = true AND customers.address <> staged_updates.address THEN
  UPDATE SET current = false, end_date = staged_updates.effective_date
WHEN NOT MATCHED THEN
  INSERT(customer_id, address, current, effective_date, end_date)
  VALUES(staged_updates.customer_id, staged_updates.address, true, staged_updates.effective_date,
null)
```


Which statement describes this implementation?

- A. The customers table is implemented as a Type 2 table; old values are overwritten and new customers are appended.
- B. The customers table is implemented as a Type 1 table; old values are overwritten by new values and no history is

maintained.

- C. The customers table is implemented as a Type 3 table; old values are maintained as a new column alongside the current value.
- D. The customers table is implemented as a Type 0 table; all writes are append only with no changes to existing values.
- E. The customers table is implemented as a Type 2 table; old values are maintained but marked as no longer current and new values are inserted.

Answer: E

Explanation:
The logic uses the MERGE INTO command to merge new records from the view updates into the table customers. The MERGE INTO command takes two arguments: a target table and a source table or view. The command also specifies a condition to match records between the target and the source, and a set of actions to perform when there is a match or not. In this case, the condition is to match records by customer_id, which is the primary key of the customers table. The actions are to update the existing record in the target with the new values from the source, and set the current_flag to false to indicate that the record is no longer current; and to insert a new record in the target with the new values from the source, and set the current_flag to true to indicate that the record is current. This means that old values are maintained but marked as no longer current and new values are inserted, which is the definition of a Type 2 table. Verified Reference: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Merge Into (Delta Lake on Databricks)" section.

NEW QUESTION # 34
The data engineering team has configured a Databricks SQL query and alert to monitor the values in a Delta Lake table. The recent_sensor_recordings table contains an identifying sensor_id alongside the timestamp and temperature for the most recent 5 minutes of recordings.
The below query is used to create the alert:

```
SELECT MEAN(temperature), MAX(temperature), MIN(temperature)
FROM recent_sensor_recordings
GROUP BY sensor_id
```


The query is set to refresh each minute and always completes in less than 10 seconds. The alert is set to trigger when mean (temperature) > 120. Notifications are triggered to be sent at most every 1 minute.
If this alert raises notifications for 3 consecutive minutes and then stops, which statement must be true?

- A. The source query failed to update properly for three consecutive minutes and then restarted
- B. The maximum temperature recording for at least one sensor exceeded 120 on three consecutive executions of the query
- C. The average temperature recordings for at least one sensor exceeded 120 on three consecutive executions of the query
- D. The recent_sensor_recordingstable was unresponsive for three consecutive runs of the query
- E. The total average temperature across all sensors exceeded 120 on three consecutive executions of the query

Answer: C

Explanation:
This is the correct answer because the query is using a GROUP BY clause on the sensor_id column, which means it will calculate the mean temperature for each sensor separately. The alert will trigger when the mean temperature for any sensor is greater than 120, which means at least one sensor had an average temperature above 120 for three consecutive minutes. The alert will stop when the mean temperature for all sensors drops below 120. Verified Reference: [Databricks Certified Data Engineer Professional], under "SQL Analytics" section; Databricks Documentation, under "Alerts" section.

NEW QUESTION # 35
A Delta Lake table was created with the below query:
Realizing that the original query had a typographical error, the below code was executed:
ALTER TABLE prod.sales_by_stor RENAME TO prod.sales_by_store
Which result will occur after running the second command?

- A. A new Delta transaction log Is created for the renamed table.
- B. All related files and metadata are dropped and recreated in a single ACID transaction.
- C. The table name change is recorded in the Delta transaction log.
- D. The table reference in the metastore is updated and all data files are moved.
- E. The table reference in the metastore is updated and no data is changed.

Answer: E

Explanation:
The query uses the CREATE TABLE USING DELTA syntax to create a Delta Lake table from an existing Parquet file stored in DBFS. The query also uses the LOCATION keyword to specify the path to the Parquet file as /mnt/finance_eda_bucket/tx_sales.parquet. By using the LOCATION keyword, the query creates an external table, which is a table that is stored outside of the default warehouse directory and whose metadata is not managed by Databricks. An external table can be created from an existing directory in a cloud storage system, such as DBFS or S3, that contains data files in a supported format, such as Parquet or CSV.
The result that will occur after running the second command is that the table reference in the metastore is updated and no data is changed. The metastore is a service that stores metadata about tables, such as their schema, location, properties, and partitions. The metastore allows users to access tables using SQL commands or Spark APIs without knowing their physical location or format. When renaming an external table using the ALTER TABLE RENAME TO command, only the table reference in the metastore is updated with the new name; no data files or directories are moved or changed in the storage system. The table will still point to the same location and use the same format as before. However, if renaming a managed table, which is a table whose metadata and data are both managed by Databricks, both the table reference in the metastore and the data files in the default warehouse directory are moved and renamed accordingly. Verified References:
[Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "ALTER TABLE RENAME TO" section; Databricks Documentation, under "Metastore" section; Databricks Documentation, under "Managed and external tables" section.

NEW QUESTION # 36
.....

- New Databricks-Certified-Professional-Data-Engineer Exam Sample □ Test Databricks-Certified-Professional-Data-Engineer Dumps □ Minimum Databricks-Certified-Professional-Data-Engineer Pass Score □ Immediately open ▶ www.pdfvce.com ◀ and search for ➡ Databricks-Certified-Professional-Data-Engineer □□□ to obtain a free download □ □Related Databricks-Certified-Professional-Data-Engineer Exams
- Up-to-Date Online Databricks Databricks-Certified-Professional-Data-Engineer Practice Test Engine □ Search for □ Databricks-Certified-Professional-Data-Engineer □ and obtain a free download on □ www.examcollectionpass.com □ □ □Databricks-Certified-Professional-Data-Engineer New Test Camp
- myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, dogbasicsinfo.us, ac.wizons.com, ncon.edu.sa, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, shortcourses.russellcollege.edu.au, cou.alnoor.edu.iq, shortcourses.russellcollege.edu.au, www.stes.tyc.edu.tw, nativemediastudios.com, Disposable vapes