DBCLOUD
EXAMS

QUESTIONS & ANSWERS

verified



Technology Provider: Databricks

Exam: Certified Data Engineer Professional

Number of questions in the database: 268

Exam Version: NOV.2026

Question #1Topic 1

An upstream system has been configured to pass the date for a given batch of data to the Databricks Jobs API as a parameter. The notebook to be scheduled will use this parameter to load data with the following code: df = spark.read.format("parquet").load(f"/mnt/source/(date)")

Which code block should be used to create the date Python variable used in the above code block?

A. date = spark.conf.get("date")

B. input_dict = input()

date= input_dict["date"]

C. import sys

date = sys.argv[1]

D. date = dbutils.notebooks.getParam("date")

E. dbutils.widgets.text("date", "null")

date = dbutils.widgets.get("date")

Correct Answer: *E*

Question #2Topic 1

The Databricks workspace administrator has configured interactive clusters for each of the data engineering groups. To control costs, clusters are set to terminate after 30 minutes of inactivity. Each user should be able to execute workloads against their assigned clusters at any time of the day.

Assuming users have been added to a workspace but not granted any permissions, which of the following describes the minimal permissions a user would need to start and attach to an already configured cluster.

A. "Can Manage" privileges on the required cluster

https://dbcloud.academy/

B. Workspace Admin privileges, cluster creation allowed, "Can Attach To" privileges on the required cluster

C. Cluster creation allowed, "Can Attach To" privileges on the required cluster

D. "Can Restart" privileges on the required cluster

E. Cluster creation allowed, "Can Restart" privileges on the required cluster

Correct Answer: *D*

Question #3Topic 1

When scheduling Structured Streaming jobs for production, which configuration automatically recovers from query failures and keeps costs low?

A. Cluster: New Job Cluster;

Retries: Unlimited;

Maximum Concurrent Runs: Unlimited

B. Cluster: New Job Cluster;

Retries: None;

Maximum Concurrent Runs: 1

C. Cluster: Existing All-Purpose Cluster;

Retries: Unlimited;

Maximum Concurrent Runs: 1

D. Cluster: New Job Cluster;

Retries: Unlimited;

Maximum Concurrent Runs: 1

E. Cluster: Existing All-Purpose Cluster;

Retries: None;

Maximum Concurrent Runs: 1

Correct Answer: *D*

Question #4*Topic 1*

https://dbcloud.academy/

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 total average temperature across all sensors exceeded 120 on three consecutive executions of the query

B. The recent_sensor_recordings table was unresponsive for three consecutive runs of the query

C. The source query failed to update properly for three consecutive minutes and then restarted

D. The maximum temperature recording for at least one sensor exceeded 120 on three consecutive executions of the query

E. The average temperature recordings for at least one sensor exceeded 120 on three consecutive executions of the query

Correct Answer: *E*

Question #5Topic 1

A junior developer complains that the code in their notebook isn't producing the correct results in the development environment. A shared screenshot reveals that while they're using a notebook versioned with Databricks Repos, they're using a personal branch that contains old logic. The desired branch named dev-2.3.9 is not available from the branch selection dropdown.

Which approach will allow this developer to review the current logic for this notebook?

A. Use Repos to make a pull request use the Databricks REST API to update the current branch to dev-2.3.9

B. Use Repos to pull changes from the remote Git repository and select the dev-2.3.9 branch.

C. Use Repos to checkout the dev-2.3.9 branch and auto-resolve conflicts with the current branch

D. Merge all changes back to the main branch in the remote Git repository and clone the repo again

https://dbcloud.academy/

E. Use Repos to merge the current branch and the dev-2.3.9 branch, then make a pull request to sync with the remote repository

Correct Answer: *B*

Question #6Topic 1

The security team is exploring whether or not the Databricks secrets module can be leveraged for connecting to an external database.
After testing the code with all Python variables being defined with strings, they upload the password to the secrets module and configure the correct permissions for the currently active user. They then modify their code to the following (leaving all other variables unchanged).

```
password = dbutils.secrets.get(scope="db_creds", key="jdbc_password")

print(password)

df = (spark
  .read
  .format("jdbc")
  .option("url", connection)
  .option("dbtable", tablename)
  .option("user", username)
  .option("password", password)
)
```

Which statement describes what will happen when the above code is executed?

A. The connection to the external table will fail; the string "REDACTED" will be printed.

B. An interactive input box will appear in the notebook; if the right password is provided, the connection will succeed and the encoded password will be saved to DBFS.

C. An interactive input box will appear in the notebook; if the right password is provided, the connection will succeed and the password will be printed in plain text.

D. The connection to the external table will succeed; the string value of password will be printed in plain text.

E. The connection to the external table will succeed; the string "REDACTED" will be printed.

Correct Answer: *E*

Question #7Topic 1

The data science team has created and logged a production model using MLflow. The following code correctly imports and applies the production model to output the predictions as a new DataFrame named preds with the schema "customer_id LONG, predictions DOUBLE, date DATE".

https://dbcloud.academy/