



**Free Questions for Databricks-Certified-Data-Engineer-Associate by actualtestdumps**

**Shared by Fulton on 29-01-2024**

**For More Free Questions and Preparation Resources**

**Check the Links on Last Page**

## Question 1

---

### Question Type: MultipleChoice

---

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")
  ._____
  .table("new_sales")
)
```

If the data engineer only wants the query to process all of the available data in as many batches as required, which of the following lines of code should the data engineer use to fill in the blank?

### Options:

---

- A- processingTime(1)
- B- trigger(availableNow=True)
- C- trigger(parallelBatch=True)
- D- trigger(processingTime='once')
- E- trigger(continuous='once')

### Answer:

---

B

### Explanation:

---

<https://stackoverflow.com/questions/71061809/trigger-availablenow-for-delta-source-streaming-queries-in-pyspark-databricks>

## Question 2

---

**Question Type:** MultipleChoice

---

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?

### Options:

---

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

### Answer:

---

E

## Question 3

---

**Question Type: MultipleChoice**

---

A data engineer has joined an existing project and they see the following query in the project repository:

```
CREATE STREAMING LIVE TABLE loyal_customers AS
```

```
SELECT customer_id -
```

```
FROM STREAM(LIVE.customers)
```

```
WHERE loyalty_level = 'high';
```

Which of the following describes why the STREAM function is included in the query?

**Options:**

---

- A-** The STREAM function is not needed and will cause an error.
- B-** The table being created is a live table.
- C-** The customers table is a streaming live table.
- D-** The customers table is a reference to a Structured Streaming query on a PySpark DataFrame.
- E-** The data in the customers table has been updated since its last run.

**Answer:**

---

C

## **Explanation:**

---

<https://docs.databricks.com/en/sql/load-data-streaming-table.html>

Load data into a streaming table

To create a streaming table from data in cloud object storage, paste the following into the query editor, and then click Run:

SQL

Copy to clipboardCopy

```
/* Load data from a volume */
```

```
CREATE OR REFRESH STREAMING TABLE <table-name> AS
```

```
SELECT * FROM STREAM read_files('/Volumes/<catalog>/<schema>/<volume>//<folder>')
```

```
/* Load data from an external location */
```

```
CREATE OR REFRESH STREAMING TABLE <table-name> AS
```

```
SELECT * FROM STREAM read_files('s3://<bucket>//<folder>')
```

## Question 4

---

**Question Type:** MultipleChoice

---

A data engineer has realized that the data files associated with a Delta table are incredibly small. They want to compact the small files to form larger files to improve performance.

Which of the following keywords can be used to compact the small files?

### Options:

---

- A- REDUCE
- B- OPTIMIZE
- C- COMPACTION
- D- REPARTITION
- E- VACUUM

### Answer:

---

B

### Explanation:

---

OPTIMIZE can be used to club small files into 1 and improve performance.

## Question 5

---

**Question Type:** MultipleChoice

---

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

### Options:

---

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

### Answer:

---

D



## Question 6

---

### Question Type: MultipleChoice

---

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?

### Options:

---

- A- INSERT INTO my\_table VALUES ('a1', 6, 9.4)
- 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- UPDATE VALUES ('a1', 6, 9.4) my\_table

### Answer:

---

A

**To Get Premium Files for Databricks-Certified-Data-Engineer-Associate Visit**

**<https://www.p2pexams.com/products/databricks-certified-data-engineer-associate>**

**For More Free Questions Visit**

**<https://www.p2pexams.com/databricks/pdf/databricks-certified-data-engineer-associate>**

