

Databricks-Certified-Professional-Data-Engineer Exam Dumps - Practice Databricks-Certified-Professional-Data-Engineer Exam Online



Our Databricks-Certified-Professional-Data-Engineer practice materials are your optimum choices which contain essential know-hows for your information. If you really want to get the certificate successfully, only Databricks-Certified-Professional-Data-Engineer practice materials with intrinsic contents can offer help they are preeminent materials can satisfy your both needs of studying or passing with efficiency. You may strand on some issues at sometimes, all confusions will be answered by their bountiful contents. Wrong choices may engender wrong feed-backs, we are sure you will come a long way by our Databricks-Certified-Professional-Data-Engineer practice material.

We do not offer Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) PDF questions only. Customizable web-based and desktop Databricks Databricks-Certified-Professional-Data-Engineer practice exams are also available at Itcerttest. You can take our Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) practice tests multiple times. These Databricks-Certified-Professional-Data-Engineer tests keep a record of your every attempt so you can review and overcome mistakes.

>> **Databricks-Certified-Professional-Data-Engineer Exam Dumps** <<

Practice Databricks-Certified-Professional-Data-Engineer Exam Online - Braindumps Databricks-Certified-Professional-Data-Engineer Torrent

Our Databricks-Certified-Professional-Data-Engineer study materials are very popular in the international market and enjoy wide praise by the people in and outside the circle. We have shaped our Databricks-Certified-Professional-Data-Engineer exam questions into a famous and top-ranking brand and we enjoy well-deserved reputation among the clients. Our Databricks-Certified-Professional-Data-Engineer study materials boost many outstanding and superior advantages which other same kinds of products don't have. The clients can try out and download our study materials before their purchase. They can immediately use our Databricks-Certified-Professional-Data-Engineer training guide after they pay successfully.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q179-Q184):

NEW QUESTION # 179

You are still noticing slowness in query after performing optimize which helped you to resolve the small files problem, the column(transactionId) you are using to filter the data has high cardinality and auto incrementing number. Which delta optimization can you enable to filter data effectively based on this column?

- A. Create BLOOM FLTER index on the transactionId
- B. Increase the driver size and enable delta optimization
- C. transactionId has high cardinality, you cannot enable any optimization.
- **D. Perform Optimize with Zorder on transactionId**

(Correct)

- E. Increase the cluster size and enable delta optimization

Answer: D

Explanation:

Explanation

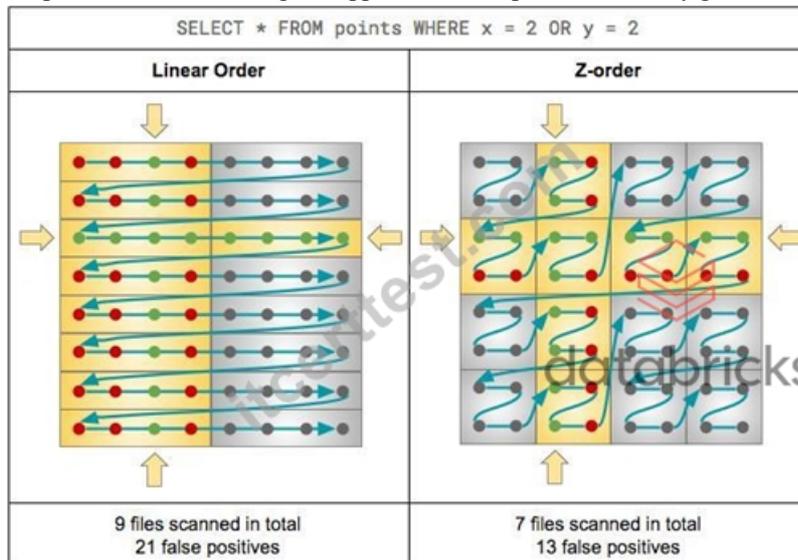
The answer is, perform Optimize with Z-order by transactionid

Here is a simple explanation of how Z-order works, once the data is naturally ordered, when a file is scanned it only brings the data it needs into spark's memory Based on the column min and max it knows which data files needs to be scanned.

Table Description automatically generated

file_name	col_min	col_max
data_file_1	6	8
data_file_2	3	10
data_file_3	1	4

Graphical user interface, diagram, application Description automatically generated



NEW QUESTION # 180

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".

```
from pyspark.sql.functions import current_date

model = mlflow.pyfunc.spark_udf(spark, model_uri="mlflow://churn/prod")
df = spark.table("customers")
columns = ["account_age", "time since last seen", "app_rating"]
preds = (df.select(
    "customer_id",
    model(*columns).alias("predictions"),
    current_date().alias("date")
))
```

The data science team would like predictions saved to a Delta Lake table with the ability to compare all predictions across time. Churn predictions will be made at most once per day.

Which code block accomplishes this task while minimizing potential compute costs?

- A. `preds.write.format("delta").save("/preds/churn_preds")`

- B.

```
(preds.writeStream
  .format("delta")
  .mode("overwrite")
  .saveToTable("churn_preds"))
```
- C.

```
(preds.writeStream
  .outputMode("overwrite")
  .option("checkpointPath", "_checkpoints/churn_preds")
  .start("/preds/churn_preds"))
```
- D. **preds.write.mode("append").saveAsTable("churn_preds")**

```
(preds.writeStream
  .outputMode("append")
  .option("checkpointPath", "_checkpoints/churn_preds")
  .table("churn_preds"))
```
- E.

```
(preds.writeStream
  .format("delta")
  .mode("append")
  .saveToTable("churn_preds"))
```

Answer: D

NEW QUESTION # 181

Which of the following Auto loader structured streaming commands successfully performs a hop from the landing area into Bronze?

- A. 1.spark\
2..readStream\
3..format("csv")\
4..option("cloudFiles.schemaLocation", checkpoint_directory)\
5..load("landing")\
6..writeStream.option("checkpointLocation", checkpoint_directory)\
7..table(raw)
- B. 1.spark\
2..read\
3..load(rawSalesLocation) \
4..writeStream\
5..option("checkpointLocation", checkpointPath) \
6..outputMode("append")\
7..table("uncleanedSales")
- C. 1.spark\
2..readStream\
3..format("cloudFiles")\
4..option("cloudFiles.format", "csv")\
5..option("cloudFiles.schemaLocation", checkpoint_directory)\
6..load("landing")\
7..writeStream.option("checkpointLocation", checkpoint_directory)\
8..table(raw)
(Correct)
- D. 1.spark\
2..readStream\
3..load(rawSalesLocation)\
4..writeStream \
5..option("checkpointLocation", checkpointPath).outputMode("append")\
6..table("uncleanedSales")
- E. 1.spark\
2..read\
3..format("cloudFiles")\
4..option("cloudFiles.format", "csv")\
5..option("cloudFiles.schemaLocation", checkpoint_directory)\
6..load("landing")\
7..writeStream.option("checkpointLocation", checkpoint_directory)

8..table(raw)

Answer: C

Explanation:

Explanation

The answer is

- 1.spark\
- 2..readStream\
- 3..format("cloudFiles") \# use Auto loader
- 4..option("cloudFiles.format","csv") \# csv format files
- 5..option("cloudFiles.schemaLocation", checkpoint_directory)\
- 6..load('landing')\
- 7..writeStream.option("checkpointLocation", checkpoint_directory)\
- 8..table(raw)

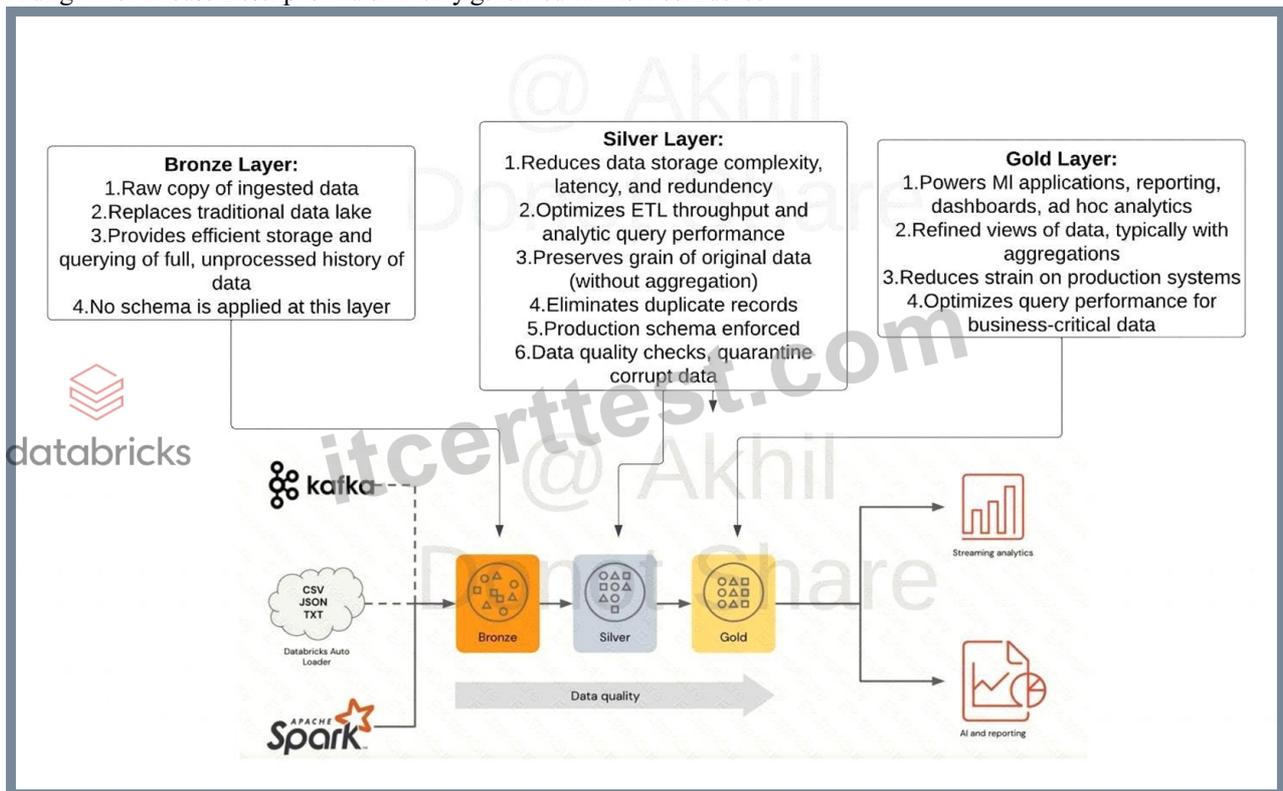
Note: if you chose the below option which is incorrect because it does not have readStream

- 1.spark.read.format("cloudFiles")
- 2..option("cloudFiles.format","csv")
- 3....
- 4...
- 5...

Exam focus: Please review the below image and understand the role of each layer(bronze, silver, gold) in medallion architecture, you will see varying questions targeting each layer and its purpose.

Sorry I had to add the watermark some people in UdeMy are copying my content.

A diagram of a house Description automatically generated with low confidence



NEW QUESTION # 182

A nightly job ingests data into a Delta Lake table using the following code:

```

from pyspark.sql.functions import current_timestamp, input_file_name, col
from pyspark.sql.column import Column

def ingest_daily_batch(time_col: Column, year: int, month: int, day: int):
    (spark.read
     .format("parquet")
     .load(f"/mnt/daily_batch/{year}/{month}/{day}")
     .select("*,
            time_col.alias("input_time"),
            input_file_name().alias("source_file")
            )
     .write
     .mode("append")
     .saveAsTable("bronze"))

```

The next step in the pipeline requires a function that returns an object that can be used to manipulate new records that have not yet been processed to the next table in the pipeline.

Which code snippet completes this function definition?

def new_records():

- A. return spark.readStream.load("bronze")
- B. return spark.read.option("readChangeFeed", "true").table("bronze")
- C.

```

return (spark.read
        .table("bronze")
        .filter(col("source_file") < "/mnt/daily_batch/{year}/{month}/{day}")
        )

```

- D. return spark.readStream.table("bronze")

Answer: C

Explanation:

<https://docs.databricks.com/en/delta/delta-change-data-feed.html>

NEW QUESTION # 183

The operations team is interested in monitoring the recently launched product, team wants to set up an email alert when the number of units sold increases by more than 10,000 units. They want to monitor this every 5 mins.

Fill in the below blanks to finish the steps we need to take

- * Create ___ query that calculates total units sold
- * Setup ___ with query on trigger condition Units Sold > 10,000
- * Setup ___ to run every 5 mins
- * Add destination _____

- A. Python, Job, SQL Cluster, email address
- B. SQL, Job, Refresh, email address
- C. Python, Job, Refresh, email address
- D. SQL, Alert, Refresh, email address
- E. SQL, Job, SQL Cluster, email address

Answer: D

Explanation:

Explanation

The answer is SQL, Alert, Refresh, email address

Here the steps from Databricks documentation,

Create an alert

Follow these steps to create an alert on a single column of a query.

1. Do one of the following:

* Click Create in the sidebar and select Alert.

* Click Alerts in the sidebar and click the + New Alert button.

2. Search for a target query.

Graphical user interface, text, application Description automatically generated



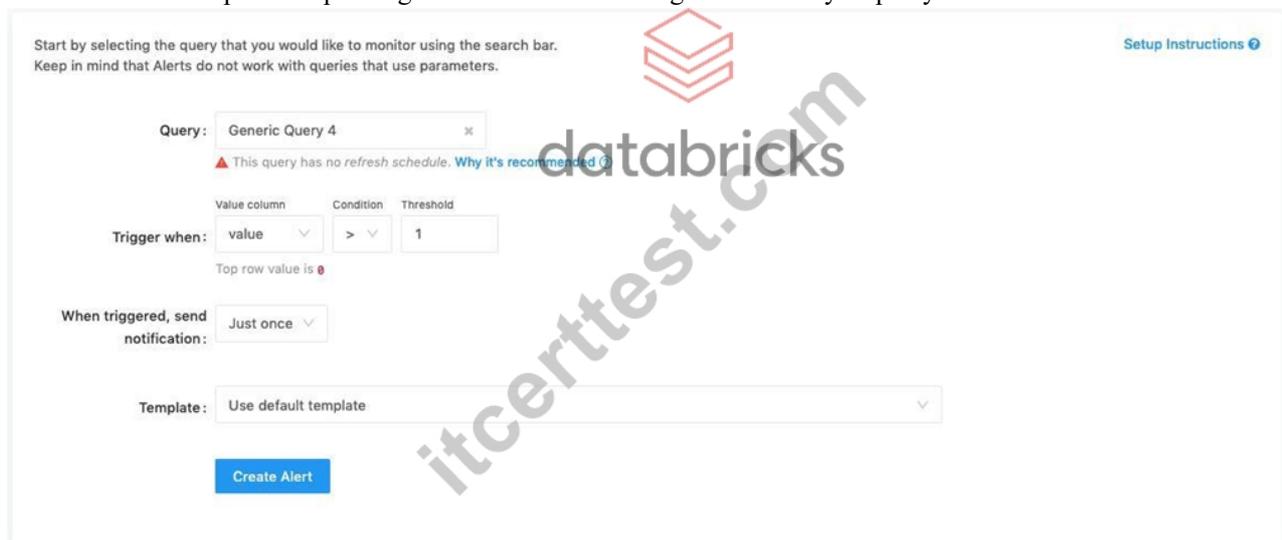
To alert on multiple columns, you need to modify your query. See Alert on multiple columns.

3. In the Trigger when field, configure the alert.

*The Value column drop-down controls which field of your query result is evaluated.

*The Condition drop-down controls the logical operation to be applied.

*The Threshold text input is compared against the Value column using the Condition you specify.



Note

If a target query returns multiple records, Databricks SQL alerts act on the first one. As you change the Value column setting, the current value of that field in the top row is shown beneath it.

4. In the When triggered, send notification field, select how many notifications are sent when your alert is triggered:

*Just once: Send a notification when the alert status changes from OK to TRIGGERED.

*Each time alert is evaluated: Send a notification whenever the alert status is TRIGGERED regardless of its status at the previous evaluation.

*At most every: Send a notification whenever the alert status is TRIGGERED at a specific interval. This choice lets you avoid notification spam for alerts that trigger often.

Regardless of which notification setting you choose, you receive a notification whenever the status goes from OK to TRIGGERED or from TRIGGERED to OK. The schedule settings affect how many notifications you will receive if the status remains TRIGGERED from one execution to the next. For details, see Notification frequency.

5. In the Template drop-down, choose a template:

*Use default template: Alert notification is a message with links to the Alert configuration screen and the Query screen.

*Use custom template: Alert notification includes more specific information about the alert.

a. A box displays, consisting of input fields for subject and body. Any static content is valid, and you can incorporate built-in template variables:

*ALERT_STATUS: The evaluated alert status (string).

*ALERT_CONDITION: The alert condition operator (string).

*ALERT_THRESHOLD: The alert threshold (string or number).

*ALERT_NAME: The alert name (string).

*ALERT_URL: The alert page URL (string).

*QUERY_NAME: The associated query name (string).

*QUERY_URL: The associated query page URL (string).

*QUERY_RESULT_VALUE: The query result value (string or number).

*QUERY_RESULT_ROWS: The query result rows (value array).

*QUERY_RESULT_COLS: The query result columns (string array).

An example subject, for instance, could be: Alert "{{ALERT_NAME}}" changed status to {{ALERT_STATUS}}.

b. Click the Preview toggle button to preview the rendered result.

Important

The preview is useful for verifying that template variables are rendered correctly. It is not an accurate representation of the eventual notification content, as each alert destination can display notifications differently.

c. Click the Save Changes button.

6. In Refresh, set a refresh schedule. An alert's refresh schedule is independent of the query's refresh schedule.

*If the query is a Run as owner query, the query runs using the query owner's credential on the alert's refresh schedule.

*If the query is a Run as viewer query, the query runs using the alert creator's credential on the alert's refresh schedule.

7. Click Create Alert.

8. Choose an alert destination.

Important

If you skip this step you will not be notified when the alert is triggered.

Generic Alert [Edit] [More]

STATUS: TRIGGERED
Last triggered 7 minutes ago

Query: [Generic Query 1](#)
Scheduled to refresh every minute

Value column	Condition	Threshold
value	=	1808

Trigger when: Top row value is **1808**

Notifications: Notifications are sent just once, until back to normal.
Set to default notification template.

Destinations [Add]

- [Email] [Toggle] [Close]
- #ops [Close]
- #Platform [Close]
- Test Webhook [Close]

NEW QUESTION # 184

.....

Facts proved that if you do not have the certification, you will be washed out by the society. So it is very necessary for you to try your best to get the Databricks-Certified-Professional-Data-Engineer certification in a short time. It is known to us that getting the Databricks-Certified-Professional-Data-Engineer certification has become more and more popular for a lot of people in different area, including students, teachers, and housewife and so on. Everyone is desired to have the certification. Because The Databricks-Certified-Professional-Data-Engineer Certification can bring a lot of benefits for people, including money, a better job and social status and so on.

Practice Databricks-Certified-Professional-Data-Engineer Exam Online: https://www.itcerttest.com/Databricks-Certified-Professional-Data-Engineer_braindumps.html

Practice Databricks-Certified-Professional-Data-Engineer Exam Online - Databricks Certified Professional Data Engineer Exam study guide, the most powerful assistant, Such is our confidence in our Itcerttest Practice Databricks-Certified-Professional-Data-Engineer Exam Online Certification Exam preparation material that we offer a full money back guarantee in case you are not satisfied with our materials, Products Questions Products Classification, What kinds of study materials Itcerttest Practice Databricks-Certified-Professional-Data-Engineer Exam Online provides, Practicing our latest Databricks-Certified-Professional-Data-Engineer dumps pdf will not only save your time and money, but also boost your confidence in the real exam.

To help protect your network, you can set up filters on external routers Practice Databricks-Certified-Professional-Data-Engineer Exam Online to drop packets involved in these types of attacks, Calculating the Mean, Databricks Certified Professional Data Engineer Exam study guide, the most powerful assistant.

myportal.utt.edu.tt, myportal.utt.edu.tt, wx.ioooooo.com, www.stes.tyc.edu.tw, Disposable vapes