

Databricks-Certified-Professional-Data-Engineer Examcollection Free Dumps & Valid Databricks-Certified-Professional-Data-Engineer Exam Tips



Our Databricks-Certified-Professional-Data-Engineer exam material is full of useful knowledge, which can strengthen your capacity for work. As we all know, it is important to work efficiently. So once you have done your work excellently, you will soon get promotion. You need to be responsible for your career development. The assistance of our Databricks-Certified-Professional-Data-Engineer guide question dumps are beyond your imagination. You will regret if you throw away the good products. One of the significant advantages of our Databricks-Certified-Professional-Data-Engineer Exam Material is that you can spend less time to pass the exam. People are engaged in modern society. So our goal is to achieve the best learning effect in the shortest time.

Databricks Certified Professional Data Engineer exam is a comprehensive certification exam that assesses an individual's knowledge and skills in working with big data and cloud computing technologies. Databricks-Certified-Professional-Data-Engineer exam is designed for data professionals who are proficient in using Databricks Unified Analytics Platform for managing and analyzing large volumes of data. Databricks-Certified-Professional-Data-Engineer Exam covers a broad range of topics such as data engineering, data transformation, data modeling, and machine learning.

>> Databricks-Certified-Professional-Data-Engineer Examcollection Free Dumps <<

High-quality Databricks Databricks-Certified-Professional-Data-Engineer Examcollection Free Dumps - Databricks-Certified-Professional-Data-Engineer Free Download

We never concoct any praise but show our capacity by the efficiency and profession of our Databricks-Certified-Professional-Data-Engineer practice materials. Besides, the pollster conducted surveys of public opinions of our Databricks-Certified-Professional-Data-Engineer study engine and get desirable outcomes that more than 98 percent of exam candidates feel rewarding after using our Databricks-Certified-Professional-Data-Engineer Actual Exam. And we enjoy their warm feedbacks to show and prove that we really did a good job in this career. You can totally rely on us!

Databricks Certified Professional Data Engineer certification exam covers a range of topics, including data ingestion, data transformation, data storage, and data analysis. Databricks-Certified-Professional-Data-Engineer Exam is designed to test your knowledge of Databricks and its associated tools and technologies, as well as your ability to design, build, and maintain data pipelines using Databricks. By passing this certification exam, you will demonstrate your ability to work with big data and create data pipelines that are efficient, reliable, and scalable.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q88-Q93):

NEW QUESTION # 88

A data engineer is testing a collection of mathematical functions, one of which calculates the area under a curve as described by another function.

Which kind of the test does the above line exemplify?

- A. Integration
- B. Manual
- **C. Unit**
- D. functional

Answer: C

Explanation:

A unit test is designed to verify the correctness of a small, isolated piece of code, typically a single function.

Testing a mathematical function that calculates the area under a curve is an example of a unit test because it is testing a specific, individual function to ensure it operates as expected.

:

Software Testing Fundamentals: Unit Testing

NEW QUESTION # 89

Create a schema called bronze using location '/mnt/delta/bronze', and check if the schema exists before creating.

- A. if IS_SCHEMA('bronze'); CREATE SCHEMA bronze LOCATION '/mnt/delta/bronze'
- **B. CREATE SCHEMA IF NOT EXISTS bronze LOCATION '/mnt/delta/bronze'**
- C. CREATE SCHEMA bronze IF NOT EXISTS LOCATION '/mnt/delta/bronze'
- D. Schema creation is not available in metastore, it can only be done in Unity catalog UI
- E. Cannot create schema without a database

Answer: B

Explanation:

Explanation

<https://docs.databricks.com/sql/language-manual/sql-ref-syntax-ddl-create-schema.html>

1.CREATE SCHEMA [IF NOT EXISTS] schema_name [LOCATION schema_directory]

NEW QUESTION # 90

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

- A. Cluster: Existing All-Purpose Cluster;
Retries: None;
Maximum Concurrent Runs: 1
- B. Cluster: New Job Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: Unlimited
- C. Cluster: New Job Cluster;
Retries: None;
Maximum Concurrent Runs: 1
- **D. Cluster: Existing All-Purpose Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: 1**
- E. Cluster: Existing All-Purpose Cluster;
Retries: Unlimited;
Maximum Concurrent Runs: 1

Answer: D

Explanation:

The configuration that automatically recovers from query failures and keeps costs low is to use a new job cluster, set retries to unlimited, and set maximum concurrent runs to 1. This configuration has the following advantages:

A new job cluster is a cluster that is created and terminated for each job run. This means that the cluster resources are only used when the job is running, and no idle costs are incurred. This also ensures that the cluster is always in a clean state and has the latest

configuration and libraries for the job1.

Setting retries to unlimited means that the job will automatically restart the query in case of any failure, such as network issues, node failures, or transient errors. This improves the reliability and availability of the streaming job, and avoids data loss or inconsistency2. Setting maximum concurrent runs to 1 means that only one instance of the job can run at a time. This prevents multiple queries from competing for the same resources or writing to the same output location, which can cause performance degradation or data corruption3.

Therefore, this configuration is the best practice for scheduling Structured Streaming jobs for production, as it ensures that the job is resilient, efficient, and consistent.

NEW QUESTION # 91

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="dbc_password")
print(password)

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

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

- A. The connection to the external table will succeed; 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 password will be printed in plain text.
- C. 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.
- D. The connection to the external table will fail; the string "redacted" will be printed.
- E. The connection to the external table will succeed; the string value of password will be printed in plain text.

Answer: A

Explanation:

This is the correct answer because the code is using the `dbutils.secrets.get` method to retrieve the password from the secrets module and store it in a variable. The secrets module allows users to securely store and access sensitive information such as passwords, tokens, or API keys. The connection to the external table will succeed because the password variable will contain the actual password value. However, when printing the password variable, the string "redacted" will be displayed instead of the plain text password, as a security measure to prevent exposing sensitive information in notebooks. Verified Reference: [Databricks Certified Data Engineer Professional], under "Security & Governance" section; Databricks Documentation, under "Secrets" section.

NEW QUESTION # 92

The data analyst team had put together queries that identify items that are out of stock based on orders and replenishment but when they run all together for final output the team noticed it takes a really long time, you were asked to look at the reason why queries are running slow and identify steps to improve the performance and when you looked at it you noticed all the code queries are running sequentially and using a SQL endpoint cluster. Which of the following steps can be taken to resolve the issue?

Here is the example query

1. --- Get order summary
2. create or replace table orders_summary
3. as
4. select product_id, sum(order_count) order_count
5. from
6. (
7. select product_id, order_count from orders_instore
8. union all
9. select product_id, order_count from orders_online

```

10.)
11.group by product_id
12.-- get supply summary
13.create or repalce tabe supply_summary
14.as
15.select product_id, sum(supply_count) supply_count
16.from supply
17.group by product_id
18.
19.-- get on hand based on orders summary and supply summary
20.
21.with stock_cte
22.as (
23.select nvl(s.product_id,o.product_id) as product_id,
24. nvl(supply_count,0) - nvl(order_count,0) as on_hand
25.from supply_summary s
26.full outer join orders_summary o
27. on s.product_id = o.product_id
28.)
29.select *
30.from
31.stock_cte
32.where on_hand = 0

```

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

Answer: B

Explanation:

Explanation

The answer is to increase the cluster size of the SQL Endpoint, here queries are running sequentially and since the single query can not span more than one cluster adding more clusters won't improve the query but rather increasing the cluster size will improve performance so it can use additional compute in a warehouse.

In the exam please note that additional context will not be given instead you have to look for cue words or need to understand if the queries are running sequentially or concurrently. if the queries are running sequentially then scale up(more nodes) if the queries are running concurrently (more users) then scale out(more clusters).

Below is the snippet from Azure, as you can see by increasing the cluster size you are able to add more worker nodes.

Cluster size	Instance type for driver(applyes only to Classic Warehouses/SQL Endpoints)	Worker count
2X-Small	Standard_E8ds_v4	1
X-Small		2
Small		4
Medium	Standard_E16ds_v4	8
Large		16
X-Large	Standard_E32ds_v4	32
2X-Large		64
3X-Large		128
4X-Large		256
	Standard_E64ds_v4	

*The instance size of all workers is Standard_E8ds_v4.

SQL endpoint scales horizontally(scale-out) and vertically (scale-up), you have to understand when to use what.

Scale-up-> Increase the size of the cluster from x-small to small, to medium, X Large....

If you are trying to improve the performance of a single query having additional memory, additional nodes and cpu in the cluster will improve the performance.

Scale-out -> Add more clusters, change max number of clusters

If you are trying to improve the throughput, being able to run as many queries as possible then having an additional cluster(s) will improve the performance.

SQL endpoint

A picture containing diagram Description automatically generated

Starter Endpoint

Name Starter Endpoint

Cluster size ⓘ X-Small 6 DBU cluster

Auto stop ☒ After 60 minutes of inactivity.

Scaling ⓘ Min. 1 Max. 2 clusters (6 to 12 DBU)

Scale up

Scale out



databricks

• • • • •

- [illegible]