Reliable Certified-Data-Engineer-Professional Dumps Sheet - Exam Certified-Data-Engineer-Professional Duration


databricks

Databricks Exam Guide

Databricks Certified Data Engineer Professional

Provide Exam Guide Feedback

Purpose of this Exam Guide

The purpose of this exam guide is to give you an overview of the exam and what is covered on the exam to help you determine your exam readiness. This document will get updated anytime there are any changes to an exam (and when those changes will take effect on an exam) so that you can be prepared. This version covers the currently live version as of August 1, 2023

Audience Description

The Databricks Certified Data Engineer Professional certification exam assesses an individual's ability to use Databricks to perform advanced data engineering tasks. This includes an understanding of the Databricks platform and developer tools like Apache Spark, Delta Lake, MLflow, and the Databricks CLI and REST API. It also assesses the ability to build optimized and cleaned ETL pipelines. Additionally, modeling data into a Lakehouse using knowledge of general data modeling concepts will also be assessed. Finally, ensuring that data pipelines are secure, reliable, monitored, and tested before deployment will also be included in this exam. Individuals who pass this certification exam can be expected to complete advanced data engineering tasks using Databricks and its associated tools.

About the Exam

- Number of items: 60 multiple-choice questions
- Time limit: 120 minutes
- Registration fee: USD 200, plus applicable taxes as required per local law
- Delivery method: Online Proctored
- Test aides: none allowed.
- Prerequisite: None required; course attendance and 1 year of hands-on experience in Databricks is highly recommended
- Validity: 2 years
- Unscored Content: Exams may include unscored items to gather statistical information for future use. These items are not identified on the form and do not impact your score, and additional time is factored into account for this content.

With the best quality and high accuracy, our Certified-Data-Engineer-Professional vce braindumps are the best study materials for the certification exam among the dumps vendors. Our experts constantly keep the pace of the current exam requirement for Certified-Data-Engineer-Professional Actual Test to ensure the accuracy of our questions. The pass rate of our Certified-Data-Engineer-Professional exam dumps almost reach to 98% because our questions and answers always updated according to the latest exam information.

Databricks Certified-Data-Engineer-Professional Exam Syllabus Topics:

Section	Weight	Objectives
Topic 1: Data Modeling	~10%	- Apply dimensional modeling techniques - Design scalable Delta Lake schemas and clustering
Topic 2: Data Sharing and Federation	~8%	- Configure Delta Sharing and Lakehouse Federation
Topic 3: Security and Governance	~10%	- Manage Unity Catalog permissions and ACLs - Implement row-level security, column masking, and compliance
Topic 4: Streaming Workloads and Change Data Capture	~11%	- Apply AUTO CDC APIs and exactly-once semantics - Implement reliable streaming pipelines

Topic 5: Developing Code for Data Processing using Python and SQL	~22%	- Build pipelines with Lakeflow Spark Declarative Pipelines and Auto Loader - Implement scalable Python/SQL code and project structures - Manage dependencies, libraries, and UDFs
Topic 6: CI/CD, Testing, and Deployment	~6%	- Deploy with Declarative Automation Bundles, CLI, and REST API - Implement testing and deployment pipelines
Topic 7: Data Transformation, Cleansing, and Quality	~12%	- Apply advanced Spark transformations - Enforce data quality and quarantine bad data
Topic 8: Cost and Performance Optimization	~13%	- Optimize queries, clusters, and storage - Leverage system tables and observability tools
Topic 9: Monitoring, Logging, and Troubleshooting	~8%	- Use Spark UI, Query Profiler, and system tables - Diagnose common pipeline and job failures

Exam Certified-Data-Engineer-Professional Duration | Certified-Data-Engineer-Professional Hottest Certification

It is known to us that the 21st century is an information era of rapid development. Now the people who have the opportunity to gain the newest information, who can top win profit maximization. In a similar way, people who want to pass Certified-Data-Engineer-Professional exam also need to have a good command of the newest information about the coming exam. However, it is not easy for a lot of people to learn more about the information about the study materials. Luckily, the Certified-Data-Engineer-Professional exam dumps from our company will help all people to have a good command of the newest information. Because our company have employed a lot of experts and professors to renew and update the Certified-Data-Engineer-Professional test training guide for all customer in order to provide all customers with the newest information. If you also choose the Certified-Data-Engineer-Professional study questions from our company, we can promise that you will have the chance to enjoy the newest information provided by our company.

Databricks Certified Data Engineer Professional Sample Questions (Q79-Q84):

NEW QUESTION # 79
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 average temperature recordings for at least one sensor exceeded 120 on three consecutive executions of the query
- B. The source query failed to update properly for three consecutive minutes and then restarted
- C. The recent_sensor_recordingstable was unresponsive for three consecutive runs of the query
- D. The total average temperature across all sensors exceeded 120 on three consecutive executions of the query
- E. The maximum temperature recording for at least one sensor exceeded 120 on three consecutive executions of the query

Answer: A

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.

NEW QUESTION # 80
A data engineer is designing a system to process batch patient encounter data stored in an S3 bucket, creating a Delta table (patient_encounters) with columns encounter_id, patient_id, encounter_date, diagnosis_code, and treatment_cost. The table is queried frequently by patient_id and encounter_date, requiring fast performance. Fine-grained access controls must be enforced. The engineer wants to minimize maintenance and boost performance. How should the data engineer create the patient_encounters table?

- A. Create a managed table in Unity Catalog. Configure Unity Catalog permissions for access controls, schedule jobs to run OPTIMIZE and VACUUM commands daily to achieve best performance.
- B. Create a managed table in Unity Catalog. Configure Unity Catalog permissions for access controls, and rely on predictive optimization to enhance query performance and simplify maintenance.
- C. Create a managed table in Hive Metastore. Configure Hive Metastore permissions for access controls, and rely on predictive optimization to enhance query performance and simplify maintenance.
- D. Create an external table in Unity Catalog, specifying an S3 location for the data files. Enable predictive optimization through table properties, and configure Unity Catalog permissions for access controls.

Answer: B

Explanation:
Databricks documentation specifies that Unity Catalog managed tables are the preferred choice for secure, low-maintenance Delta Lake architectures. Managed tables provide full lifecycle management, including metadata, file storage, and access control integration with Unity Catalog.
Fine-grained permissions can be enforced at the column and row level through built-in Unity Catalog governance.
Additionally, Predictive Optimization (Auto Optimize + Auto Compaction) automatically manages file sizes, metadata pruning, and layout optimization, eliminating the need for manual maintenance such as scheduling OPTIMIZE or VACUUM.
External tables (A) require manual path management, and Hive Metastore tables (D) do not support Unity Catalog access policies.
Therefore, creating a managed Unity Catalog table with predictive optimization provides both the security and performance benefits needed, making B the correct solution.

NEW QUESTION # 81
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 succeed; the string "redacted" will be printed.
- B. The connection to the external table will fail; the string "redacted" will be printed.
- C. The connection to the external table will succeed; the string value of password will be printed in plain text.
- D. 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.
- E. 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.

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.

NEW QUESTION # 82
A data engineer is optimizing a managed Delta table that suffers from data skew and frequently changing query filter columns. The engineer wants to avoid costly data rewrites when query patterns evolve. The table size is under 1 TB. How should the data engineer meet this requirement?

- A. Use Hive-style partitioning, as it provides efficient data skipping and is easy to change partition columns at any time.
- B. Combine partitioning and Z-ordering to maximize flexibility and minimize maintenance as query patterns change.
- C. Apply Z-ordering, since it allows flexible reorganization of data layout without rewriting existing files and adapts easily to new filter columns.
- D. Enable liquid clustering, as it efficiently handles data skew, allows clustering keys to be changed without rewriting existing data, and adapts to evolving query patterns.

Answer: D

Explanation:
Liquid clustering is designed for managed tables under 1TB with evolving query patterns. It efficiently addresses data skew, continuously optimizes data layout, and allows clustering keys to be changed without requiring full data rewrites, making it well suited for frequently changing filter columns while minimizing maintenance overhead.

NEW QUESTION # 83
A data engineer is attempting to execute the following PySpark code:
df = spark.read.table("sales")
result = df.groupBy("region").agg(sum("revenue"))
However, upon inspecting the execution plan and profiling the Spark job, they observe excessive data shuffling during the aggregation phase.
Which technique should be applied to reduce shuffling during the groupBy aggregation operation?

- A. Use coalesce() after the aggregation.
- B. Use broadcast join.
- C. Repartition by region before aggregation.
- D. Caching the DataFrame df.

Answer: C

Explanation:
Repartitioning the DataFrame by the grouping key ensures that records with the same region are colocated in the same partitions before the aggregation runs. This significantly reduces the amount of data shuffled during the groupBy operation, leading to more efficient execution.

NEW QUESTION # 84
.....

Our company is professional brand. There are a lot of experts and professors in the field in our company. All the experts in our company are devoting all of their time to design the best Certified-Data-Engineer-Professionaltest question for all people. In order to ensure quality of the products, a lot of experts keep themselves working day and night. We can make sure that you cannot find the more suitable Certified-Data-Engineer-Professionalcertification guide than our study materials, so hurry to choose the study materials from our company as your study tool, it will be very useful for you to prepare for the Certified-Data-Engineer-Professional exam.

Exam Certified-Data-Engineer-Professional Duration: https://www.lead2passexam.com/Databricks/valid-Certified-Data-

Engineer-Professional-exam-dumps.html

- High Pass-Rate Reliable Certified-Data-Engineer-Professional Dumps Sheet to Obtain Databricks Certification □ Search for ▸ Certified-Data-Engineer-Professional ◂ and obtain a free download on （ www.pdfdumps.com ） □New Certified-Data-Engineer-Professional Real Exam
- Databricks Certified-Data-Engineer-Professional Exam | Reliable Certified-Data-Engineer-Professional Dumps Sheet - Sample Download Free of Exam Certified-Data-Engineer-Professional Duration □ Enter ▹ www.pdfvce.com ◃ and search for □ Certified-Data-Engineer-Professional □ to download for free □New Certified-Data-Engineer-Professional Real Exam
- Latest Study Certified-Data-Engineer-Professional Questions □ Actual Certified-Data-Engineer-Professional Test Pdf □ Certified-Data-Engineer-Professional PDF Download □ Open ▹ www.vce4dumps.com ◃ enter ✔ Certified-Data-Engineer-Professional □✔□ and obtain a free download □Certified-Data-Engineer-Professional Mock Test
- Useful Reliable Certified-Data-Engineer-Professional Dumps Sheet Provide Prefect Assistance in Certified-Data-Engineer-Professional Preparation □ Search for □ Certified-Data-Engineer-Professional □ and easily obtain a free download on " www.pdfvce.com " □Premium Certified-Data-Engineer-Professional Files
- Certified-Data-Engineer-Professional Prepaway Dumps □ Latest Study Certified-Data-Engineer-Professional Questions □ □ Valid Braindumps Certified-Data-Engineer-Professional Free ⧗ Search for 【 Certified-Data-Engineer-Professional 】 and download it for free immediately on ☀ www.troytecdumps.com □☀□ □Premium Certified-Data-Engineer-Professional Files
- Useful Reliable Certified-Data-Engineer-Professional Dumps Sheet Provide Prefect Assistance in Certified-Data-Engineer-Professional Preparation □ Easily obtain ✔ Certified-Data-Engineer-Professional □✔□ for free download through ➥ www.pdfvce.com □ □Premium Certified-Data-Engineer-Professional Files
- Reliable Certified-Data-Engineer-Professional Dumps Sheet Imparts You the Best Knowledge of Certified-Data-Engineer-Professional Exam □ Enter 【 www.pdfdumps.com 】 and search for [Certified-Data-Engineer-Professional] to download for free □Latest Study Certified-Data-Engineer-Professional Questions
- Certified-Data-Engineer-Professional Practice Exam Fee □ Latest Study Certified-Data-Engineer-Professional Questions □ Certified-Data-Engineer-Professional Certification Exam Cost □ Search for 【 Certified-Data-Engineer-Professional 】 and easily obtain a free download on { www.pdfvce.com } □Certified-Data-Engineer-Professional Certification Exam Cost
- Get Success In Databricks Certified-Data-Engineer-Professional Exam With www.practicevce.com Quickly □ Search for ➤ Certified-Data-Engineer-Professional □ and easily obtain a free download on ➥ www.practicevce.com □ □Exams Certified-Data-Engineer-Professional Torrent
- Latest Study Certified-Data-Engineer-Professional Questions □ Certified-Data-Engineer-Professional Prepaway Dumps □ □ Certified-Data-Engineer-Professional Prepaway Dumps □ Search on ➡ www.pdfvce.com □□□ for ➽ Certified-Data-Engineer-Professional □ to obtain exam materials for free download □Premium Certified-Data-Engineer-Professional Files
- First-Grade Databricks Certified-Data-Engineer-Professional: Reliable Databricks Certified Data Engineer Professional Dumps Sheet - Pass-Sure www.prepawayexam.com Exam Certified-Data-Engineer-Professional Duration □ 《 www.prepawayexam.com 》 is best website to obtain ➤ Certified-Data-Engineer-Professional □ for free download □ □Certification Certified-Data-Engineer-Professional Test Questions
- www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, writeablog.net, myportal.utt.edu.tt, www.stes.tyc.edu.tw, 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, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, Disposable vapes