

Quiz Databricks - Databricks-Certified-Data-Engineer-Professional Authoritative Test Study Guide



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 April 1, 2024. Please check back two weeks before you take your exam to make sure you have the most current version.**

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 scored 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
- Recertification: Recertification is required every two years to maintain your certified status. To recertify, you must take the full exam that is currently live. Please review the "Getting Ready for the Exam" section on the exam webpage to prepare for taking the exam again.

Our company offers valid Databricks Databricks-Certified-Data-Engineer-Professional Exam Cram materials; you can purchase our products any time as we are 7*24 on duty throughout the whole year. We can guarantee you that if you purchase our Databricks-Certified-Data-Engineer-Professional exam cram materials you can pass test at first attempt without large time and energy. If the test questions change, candidates share one year updates materials and service warranty, or if you fail exam we will full refund directly.

Whereas the other two LatestCram Databricks-Certified-Data-Engineer-Professional exam questions formats are concerned, both are customizable practice tests, provide real time environment, track your progress, and help you overcome mistakes. The desktop Databricks Databricks-Certified-Data-Engineer-Professional Practice Test software is compatible with Windows computers. The web based practice exam is supported by all browsers and operating systems.

>> **Databricks-Certified-Data-Engineer-Professional Test Study Guide** <<

Latest Databricks-Certified-Data-Engineer-Professional Exam Papers - Databricks-Certified-Data-Engineer-Professional Reliable Test Prep

During nearly ten years, our Databricks-Certified-Data-Engineer-Professional exam questions have met with warm reception and quick sale in the international market. Our Databricks-Certified-Data-Engineer-Professional study materials are not only as reasonable priced as other makers, but also they are distinctly superior in the many respects. With tens of thousands of our loyal customers supporting us all the way, we believe we will do a better job in this career. More and more candidates will be benefited from our excellent Databricks-Certified-Data-Engineer-Professional training guide!

Databricks Certified Data Engineer Professional Exam Sample Questions (Q106-Q111):

NEW QUESTION # 106

Two data engineers are working on the same Databricks notebook in separate branches. Both have edited the same section of code. When one tries to merge the other's branch into their own using the Databricks Git folders UI, a merge conflict occurs on that notebook file. The UI highlights the conflict and presents options for resolution. How should the data engineers resolve this merge conflict using Databricks Git folders?

- A. Abort the merge, discard all local changes, and try the merge operation again without reviewing the conflicting code.
- B. Use the Git CLI in the cluster's web terminal to force-push the conflicted merge (git push -force), overriding the remote branch with the local version and discarding changes.
- C. Use the Git folders UI to manually edit the notebook file, selecting the desired lines from both versions and removing the conflict markers, then mark the conflict as resolved.
- D. Delete the conflicted notebook file via the Databricks workspace UI, commit the deletion, and recreate the notebook from scratch in a new commit to bypass the conflict entirely.

Answer: C

Explanation:

In the Databricks Git folders integration, when merge conflicts arise in notebooks, the UI provides a visual diff editor that highlights conflicting code segments. Users can manually choose which changes to keep from each branch, edit directly in the notebook UI, and remove conflict markers.

After resolving, the engineer must mark the conflict as resolved, save, and commit the final version.

This process ensures that both contributors' valid code segments are merged correctly and version history is maintained.

Forcing a push (C) or deleting notebooks (B) introduces data loss or versioning issues. Aborting without review (A) violates collaborative best practices. Therefore, D is the only correct and Databricks-approved way to resolve notebook merge conflicts.

NEW QUESTION # 107

A data engineer is using Auto Loader to read incoming JSON data as it arrives. They have configured Auto Loader to quarantine invalid JSON records but notice that over time, some records are being quarantined even though they are well-formed JSON.

The code snippet is:

```
df = (spark.readStream
  .format("cloudFiles")
  .option("cloudFiles.format", "json")
  .option("badRecordsPath", "/tmp/somewhere/badRecordsPath")
  .schema("a int, b int")
  .load("/Volumes/catalog/schema/raw_data/"))
```

What is the cause of the missing data?

- A. The source data is valid JSON but does not conform to the defined schema in some way.
- B. At some point, the upstream data provider switched everything to multi-line JSON.
- C. The badRecordsPath location is accumulating many small files.
- D. The engineer forgot to set the option "cloudFiles.quarantineMode" = "rescue".

Answer: A

Explanation:

Auto Loader quarantines records that cannot be parsed according to the specified schema, even if the JSON itself is well formed. If incoming records contain additional fields, missing fields, or incompatible data types compared to the declared schema, they are treated as invalid and routed to the bad records path.

NEW QUESTION # 108

A data engineering team has a time-consuming data ingestion job with three data sources. Each notebook takes about one hour to load new data. One day, the job fails because a notebook update introduced a new required configuration parameter. The team must quickly fix the issue and load the latest data from the failing source. Which action should the team take?

- A. Share the analysis with the failing notebook owner so that they can fix it quickly.

- B. Repair the run with the new parameter.
- C. Repair the run with the new parameter, and update the task by adding the missing task parameter.
- D. Update the task by adding the missing task parameter, and manually run the job.

Answer: C

Explanation:

The repair run capability in Databricks Jobs allows re-execution of failed tasks without re-running successful ones. When a parameterized job fails due to missing or incorrect task configuration, engineers can perform a repair run to fix inputs or parameters and resume from the failed state.

This approach saves time, reduces cost, and ensures workflow continuity by avoiding unnecessary recomputation. Additionally, updating the task definition with the missing parameter prevents future runs from failing.

Running the job manually (B) loses run context; (C) alone does not prevent recurrence; (D) delays resolution. Thus, A follows the correct operational and recovery practice.

NEW QUESTION # 109

A security analytics pipeline must enrich billions of raw connection logs with geolocation data.

The join hinges on finding which IPv4 range each event's address falls into.

Table 1: network_events (5 billion rows)

```
event_id int
ip_int
42 3232235777
```

Table 2: ip_ranges (2 million rows)

```
start_ip int end_ip int country
3232235520 3232236031 US
```

The query is currently very slow:

```
SELECT n.event_id, n.ip_int, r.country
FROM network_events n
JOIN ip_ranges r
ON n.ip_int BETWEEN r.start_ip_int AND r.end_ip_int;
```

Which change will most dramatically accelerate the query while preserving its logic?

- A. Add a broadcast hint: /*+ BROADCAST(r) */ for ip_ranges.
- B. Increase spark.sql.shuffle.partitions from 200 to 10000.
- C. Force a sort-merge join with /*+ MERGE(r) */.
- D. Add a range-join hint /*+ RANGE_JOIN(r, 65536) */.

Answer: D

Explanation:

The query joins billions of rows (network_events) with millions of rows (ip_ranges) using a range predicate (BETWEEN). Unlike equality joins (=), range joins are not efficiently handled by broadcast or sort-merge joins because:

Broadcast Join (D): Effective for small tables but only for equality joins. Since this query uses a range condition, broadcast will not reduce the complexity of scanning billions of records across non-equality conditions.

Sort-Merge Join (C): Works for ordered joins but is inefficient on range conditions. Sorting billions of records adds excessive overhead and will not resolve the bottleneck.

Increasing Shuffle Partitions (A): Only spreads out shuffle work but does not address the fundamental inefficiency of range-based lookups at scale.

Range Joins in Spark (RANGE_JOIN hint):

Databricks provides range join optimizations specifically for conditions such as BETWEEN. By applying a RANGE_JOIN hint, Spark can build optimized data structures (such as interval indexes or partition pruning strategies) that map billions of input rows to ranges much faster. This avoids brute-force scans and unnecessary shuffle costs.

Thus, Option B is the correct solution because:

It leverages range-join optimization, which is purpose-built for queries joining massive event logs to smaller lookup tables with IP ranges.

This ensures Spark can evaluate billions of rows against millions of ranges with optimized matching logic, drastically improving query performance while preserving correctness.

NEW QUESTION # 110

What statement is true regarding the retention of job run history?

- A. It is retained for 60 days, during which you can export notebook run results to HTML
- B. It is retained for 60 days, after which logs are archived
- C. It is retained for 90 days or until the run-id is re-used through custom run configuration
- D. It is retained for 30 days, during which time you can deliver job run logs to DBFS or S3
- E. It is retained until you export or delete job run logs

Answer: A

Explanation:

<https://docs.databricks.com/en/workflows/jobs/monitor-job-runs.html>

NEW QUESTION # 111

.....

If you are busy with your work and have little time to prepare for the exam. You can just choose our Databricks-Certified-Data-Engineer-Professional learning materials, and you will save your time. You just need to spend about 48 to 72 hours on practicing, and you can pass the exam successfully. Databricks-Certified-Data-Engineer-Professional exam materials are edited by professional experts, therefore they are high-quality. And Databricks-Certified-Data-Engineer-Professional Learning Materials of us also have certain quantity, and they will be enough for you to carry on practice. We offer you free demo for you to try before buying Databricks-Certified-Data-Engineer-Professional exam dumps, so that you can know the format of the complete version.

Latest Databricks-Certified-Data-Engineer-Professional Exam Papers: <https://www.latestcram.com/Databricks-Certified-Data-Engineer-Professional-exam-cram-questions.html>

Databricks Databricks-Certified-Data-Engineer-Professional Test Study Guide If you can own the certification means that you can do the job well in the area so you can get easy and quick promotion, Databricks Databricks-Certified-Data-Engineer-Professional Test Study Guide This is a great way to prepare for a lab exam, Databricks Databricks-Certified-Data-Engineer-Professional Test Study Guide We make it a reality, Just choose Databricks Certification guide question to improve your knowledge to pass Databricks-Certified-Data-Engineer-Professional exam, which is your testimony of competence, This Databricks-Certified-Data-Engineer-Professional exam study material is offered to you at a very low price.

If you are not sure about your exam, choosing our Databricks-Certified-Data-Engineer-Professional training materials will be a good choice for candidates, Many skilled cybersecurity teachers have found the private sector more lucrative than the educational field.

Databricks Databricks-Certified-Data-Engineer-Professional Exam | Databricks-Certified-Data-Engineer-Professional Test Study Guide - Always Available for your Demands

If you can own the certification means that you can do the Databricks-Certified-Data-Engineer-Professional job well in the area so you can get easy and quick promotion, This is a great way to prepare for a lab exam.

We make it a reality, Just choose Databricks Certification guide question to improve your knowledge to pass Databricks-Certified-Data-Engineer-Professional exam, which is your testimony of competence, This Databricks-Certified-Data-Engineer-Professional exam study material is offered to you at a very low price.

- Exam Databricks-Certified-Data-Engineer-Professional Answers New Databricks-Certified-Data-Engineer-Professional Test Voucher Latest Databricks-Certified-Data-Engineer-Professional Exam Discount Easily obtain **▶** Databricks-Certified-Data-Engineer-Professional for free download through **>** www.troytecdumps.com **<** Databricks-Certified-Data-Engineer-Professional Fresh Dumps
- Databricks Certified Data Engineer Professional Exam Exam Guide Have Reasonable Prices but Various Benefits Study Questions Easily obtain free download of **▶** Databricks-Certified-Data-Engineer-Professional by searching on **▶** www.pdfvce.com Reliable Databricks-Certified-Data-Engineer-Professional Braindumps Book
- Latest Databricks-Certified-Data-Engineer-Professional Exam Discount Databricks-Certified-Data-Engineer-Professional Actual Questions Dumps Databricks-Certified-Data-Engineer-Professional Questions Easily obtain free download of **▶** Databricks-Certified-Data-Engineer-Professional by searching on www.torrentvce.com Databricks-Certified-Data-Engineer-Professional Valid Dumps
- Dumps Databricks-Certified-Data-Engineer-Professional Questions New Databricks-Certified-Data-Engineer-Professional Mock Exam New Databricks-Certified-Data-Engineer-Professional Mock Exam Search for [Databricks-Certified-Data-Engineer-Professional] and easily obtain a free download on **⇒** www.pdfvce.com **⇐**

