

Databricks-Certified-Data-Engineer-Professional Dumps Download | Answers Databricks-Certified-Data-Engineer-Professional Real Questions



Any ambiguous points may cause trouble to exam candidates. So clarity of our Databricks-Certified-Data-Engineer-Professional training materials make us irreplaceable including all necessary information to convey the message in details to the readers. All necessary elements are included in our Databricks-Certified-Data-Engineer-Professional practice materials. Effective Databricks-Certified-Data-Engineer-Professional exam simulation can help increase your possibility of winning by establishing solid bond with you, help you gain more self-confidence and more success.

We should admit that gaining the Databricks-Certified-Data-Engineer-Professional test certification will bring your some benefits. You may get a good opportunity in the job interview due to your Databricks Databricks-Certified-Data-Engineer-Professional exam certification. You may have a promotion in your present job and get a considerable salary. So, no matter how difficult it is, many IT candidates still choose to take the Databricks-Certified-Data-Engineer-Professional exam test. SureTorrent Databricks latest practice exam test may contribute to your Databricks-Certified-Data-Engineer-Professional Exam Preparation. We have three different versions for you to choose, the PDF, PC Test Engine, Online Test Engine. You can choose the proper version according to your actual condition. Databricks Databricks-Certified-Data-Engineer-Professional exam torrents are valid and useful which can ensure you 100% pass in the actual test.

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

Answers Databricks-Certified-Data-Engineer-Professional Real Questions - Databricks-Certified-Data-Engineer-Professional Exam Collection

Once you enter into our interface, nothing will disturb your learning the Databricks-Certified-Data-Engineer-Professional training engine except the questions and answers. So all your attention will be concentrated on study. At the same time, each process is easy for you to understand. There will have small buttons on the Databricks-Certified-Data-Engineer-Professional Exam simulation to help you switch between the different pages. It does not matter whether you can operate the computers well. Our Databricks-Certified-Data-Engineer-Professional training engine will never make you confused.

Databricks Certified Data Engineer Professional Exam Sample Questions (Q96-Q101):

NEW QUESTION # 96

A data engineer is running a `groupBy` aggregation on a massive user activity log grouped by `user_id`. A few users have millions of records, causing task skew and long runtimes. Which technique will fix the skew in this aggregation?

- A. Increase the Spark driver memory and retry.
- **B. Use salting by adding a random prefix to skewed keys before aggregation, then aggregate again after removing the prefix.**
- C. Use `reduceByKey` instead of `groupBy` to avoid shuffles.
- D. Filter out the skewed users before the aggregation.

Answer: B

Explanation:

Salting distributes records for heavily skewed keys across multiple partitions by adding a random prefix, which balances task execution during the aggregation. A second aggregation after removing the prefix correctly recombines the partial results, eliminating skew-related bottlenecks without losing accuracy.

NEW QUESTION # 97

A data engineer needs to capture pipeline settings from an existing in the workspace, and use them to create and version a JSON file to create a new pipeline. Which command should the data engineer enter in a web terminal configured with the Databricks CLI?

- A. Use the get command to capture the settings for the existing pipeline; remove the pipeline_id and rename the pipeline; use this in a create command
- B. Use list pipelines to get the specs for all pipelines; get the pipeline spec from the return results parse and use this to create a pipeline
- C. Stop the existing pipeline; use the returned settings in a reset command
- D. Use the alone command to create a copy of an existing pipeline; use the get JSON command to get the pipeline definition; save this to git

Answer: A

Explanation:

The Databricks CLI provides a way to automate interactions with Databricks services. When dealing with pipelines, you can use the databricks pipelines get --pipeline-id command to capture the settings of an existing pipeline in JSON format. This JSON can then be modified by removing the pipeline_id to prevent conflicts and renaming the pipeline to create a new pipeline. The modified JSON file can then be used with the databricks pipelines create command to create a new pipeline with those settings.

NEW QUESTION # 98

A data engineer is building a customer data pipeline in Lakeflow Spark Declarative Pipelines. The source is a cloud-based event stream with limited retention containing inserts, updates, and deletes for customer records. These changes are being applied using the AUTO CDC INTO syntax to maintain an SCD Type 1 table as the target table, customer_dim. How should the data engineer build a downstream job that streams from the customer_dim table to only act on updates and delete events, processing data incrementally?

- A. When stored as SCD 1, the target of AUTO CDC INTO includes updates and deletes. Streaming from customer_dim can fail due to these operations. Instead, build another stream from the original source.
- B. Streaming from customer_dim table would only be possible in the case of SCD 2 retention.
- C. Use ignoreChanges flag while streaming from customer_dim to avoid breaking the pipeline during updates and deletes.
- D. Read change data feed from customer_dim table and apply filters to incrementally act on the change events.

Answer: D

Explanation:

Reading the change data feed from the customer_dim table enables downstream processing to react specifically to update and delete events while operating incrementally. Change data feed exposes row-level change types and versions, making it the correct mechanism for streaming only the relevant changes from an SCD Type 1 table maintained with AUTO CDC INTO.

NEW QUESTION # 99

A data architect has heard about lake's built-in versioning and time travel capabilities. For auditing purposes they have a requirement to maintain a full of all valid street addresses as they appear in the customers table.

The architect is interested in implementing a Type 1 table, overwriting existing records with new values and relying on Delta Lake time travel to support long-term auditing. A data engineer on the project feels that a Type 2 table will provide better performance and scalability. Which piece of information is critical to this decision?

- A. Delta Lake time travel does not scale well in cost or latency to provide a long-term versioning solution.
- B. Shallow clones can be combined with Type 1 tables to accelerate historic queries for long-term versioning.
- C. Data corruption can occur if a query fails in a partially completed state because Type 2 tables requires setting multiple fields in a single update.

- D. Delta Lake only supports Type 0 tables; once records are inserted to a Delta Lake table, they cannot be modified.
- E. Delta Lake time travel cannot be used to query previous versions of these tables because Type 1 changes modify data files in place.

Answer: A

Explanation:

Delta Lake's time travel feature allows users to access previous versions of a table, providing a powerful tool for auditing and versioning. However, using time travel as a long-term versioning solution for auditing purposes can be less optimal in terms of cost and performance, especially as the volume of data and the number of versions grow. For maintaining a full history of valid street addresses as they appear in a customers table, using a Type 2 table (where each update creates a new record with versioning) might provide better scalability and performance by avoiding the overhead associated with accessing older versions of a large table. While Type 1 tables, where existing records are overwritten with new values, seem simpler and can leverage time travel for auditing, the critical piece of information is that time travel might not scale well in cost or latency for long-term versioning needs, making a Type 2 approach more viable for performance and scalability.

NEW QUESTION # 100

A data engineer deploys a multi-task Databricks job that orchestrates three notebooks. One task intermittently fails with Exit Code 1 but succeeds on retry. The engineer needs to collect detailed logs for the failing attempts, including stdout/stderr and cluster lifecycle context, and share them with the platform team. What steps the data engineer needs to follow using built-in tools?

- A. Use the notebook interactive debugger to re-run the entire multi-task job, and capture step- through traces for the failing task.
- B. Export the notebook run results to HTML; this bundle includes complete stdout, stderr, and cluster event history across all tasks.
- C. Download worker logs directly from the Spark UI and ignore driver logs, as worker logs contain stdout/stderr for all tasks and cluster events.
- **D. From the job run details page, export the job's logs or configure log delivery; then retrieve the compute driver logs and event logs from the compute details page to correlate stdout/stderr with cluster events.**

Answer: D

Explanation:

The recommended way to troubleshoot and collect detailed job logs is through the Job Run Details page in Databricks. From there, engineers can export run logs or configure automatic log delivery to a storage destination. The driver and event logs available under compute details provide stdout, stderr, and cluster lifecycle context required for root-cause analysis.

NEW QUESTION # 101

.....

One can start using product of SureTorrent instantly after buying. The 24/7 support system is available for the customers so that they don't stick to any problems. If they do so, they can contact the support system, which will assist them in the right way and solve their issues. A lot of Databricks Certified Data Engineer Professional Exam (Databricks-Certified-Data-Engineer-Professional) exam applicants have used the Databricks Certified Data Engineer Professional Exam (Databricks-Certified-Data-Engineer-Professional) practice material. They are satisfied with it because it is updated.

Answers Databricks-Certified-Data-Engineer-Professional Real Questions: <https://www.suretorrent.com/Databricks-Certified-Data-Engineer-Professional-exam-guide-torrent.html>

Databricks Databricks-Certified-Data-Engineer-Professional Dumps Download If the product activation key has not been entered, the customer has thirty (30) days from the date of purchase to return the product for refund, Databricks Databricks-Certified-Data-Engineer-Professional Dumps Download Generally speaking, customers who study hard enough to make preparations for IT exams can, after 20 or 30 hours of practice, take part in the exam and pass exams for certificates, Databricks Databricks-Certified-Data-Engineer-Professional Dumps Download The answer is that you get the certificate.

Select the Direct Selection tool in the Tools panel, What's a String, If Answers Databricks-Certified-Data-Engineer-Professional Real Questions the product activation key has not been entered, the customer has thirty (30) days from the date of purchase to return the product for refund.

