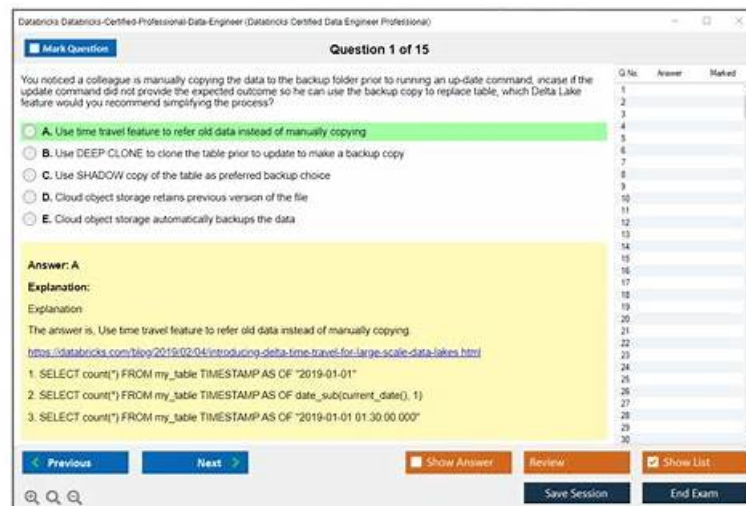


# Databricks-Certified-Professional-Data-Engineer Study Questions - Databricks-Certified-Professional-Data-Engineer Free Demo & Databricks-Certified-Professional-Data-Engineer Valid Torrent



Databricks dumps are designed according to the Databricks Databricks-Certified-Professional-Data-Engineer certification exam standard and have hundreds of questions similar to the actual Databricks-Certified-Professional-Data-Engineer exam. VCETorrent Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) web-based practice exam software also works without installation. It is browser-based; therefore no need to install it, and you can start practicing for the Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) exam by creating the Databricks Databricks-Certified-Professional-Data-Engineer practice test.

Databricks Certified Professional Data Engineer exam covers a wide range of topics, such as data ingestion, transformation, storage, and processing. Databricks-Certified-Professional-Data-Engineer exam tests the candidates' ability to use Databricks tools and technologies to solve real-world problems and challenges. Candidates who pass Databricks-Certified-Professional-Data-Engineer exam demonstrate their proficiency in designing, building, and managing data pipelines with Databricks, which is a leading cloud-based platform for big data processing and analytics.

Databricks Databricks-Certified-Professional-Data-Engineer is a certification that validates the skills and knowledge of professionals who work with big data and data engineering. Databricks-Certified-Professional-Data-Engineer exam is designed for individuals who have experience in designing, building, and maintaining data pipelines on the Databricks platform. Databricks Certified Professional Data Engineer Exam certification confirms that the candidate has proficiency in building robust, scalable, and efficient big data solutions in the cloud.

>> Latest Databricks-Certified-Professional-Data-Engineer Test Testking <<

## Reliable Databricks-Certified-Professional-Data-Engineer Test Preparation - Databricks-Certified-Professional-Data-Engineer Test Vce

Choosing valid Databricks dumps means closer to success. Before you buy our products, you can download the free demo of Databricks-Certified-Professional-Data-Engineer test questions to check the accuracy of our dumps. Besides, there are 24/7 customer assisting to support you in case you may have any questions about Databricks-Certified-Professional-Data-Engineer Dumps PDF or download link.

## Databricks Certified Professional Data Engineer Exam Sample Questions (Q41-Q46):

NEW QUESTION # 41

Newly joined data analyst requested read-only access to tables, assuming you are owner/admin which section of Databricks platform is going to facilitate granting select access to the user

- **A. Data explorer**
- B. Azure Databricks control pane IAM
- C. Admin console
- D. Azure RBAC
- E. User settings

**Answer: A**

Explanation:

Explanation

Answer is Data Explorer

<https://docs.databricks.com/sql/user/data/index.html>

Data explorer lets you easily explore and manage permissions on databases and tables. Users can view schema details, preview sample data, and see table details and properties. Administrators can view and change owners, and admins and data object owners can grant and revoke permissions.

To open data explorer, click Data in the sidebar.

#### NEW QUESTION # 42

The marketing team is launching a new campaign to monitor the performance of the new campaign for the first two weeks, they would like to set up a dashboard with a refresh schedule to run every 5 minutes, which of the below steps can be taken to reduce of the cost of this refresh over time?

- **A. Setup the dashboard refresh schedule to end in two weeks**
- B. Reduce the max size of auto scaling from 10 to 5
- C. Reduce the size of the SQL Cluster size
- D. Change the spot instance policy from reliability optimized to cost optimized
- E. Always use X-small cluster

**Answer: A**

Explanation:

Explanation

The answer is Setup the dashboard refresh schedule to end in two weeks

#### NEW QUESTION # 43

The data governance team has instituted a requirement that all tables containing Personal Identifiable Information (PII) must be clearly annotated. This includes adding column comments, table comments, and setting the custom table property "contains\_pii" = true.

The following SQL DDL statement is executed to create a new table:

Which command allows manual confirmation that these three requirements have been met?

- **A. DESCRIBE EXTENDED dev.pii test**
- B. SHOW TABLES dev
- C. DESCRIBE DETAIL dev.pii test
- D. DESCRIBE HISTORY dev.pii test
- E. SHOW TBLPROPERTIES dev.pii test

**Answer: A**

Explanation:

This is the correct answer because it allows manual confirmation that these three requirements have been met.

The requirements are that all tables containing Personal Identifiable Information (PII) must be clearly annotated, which includes adding column comments, table comments, and setting the custom table property

"contains\_pii" = true. The DESCRIBE EXTENDED command is used to display detailed information about a table, such as its schema, location, properties, and comments. By using this command on the dev.pii\_test table, one can verify that the table has been created with the correct column comments, table comment, and custom table property as specified in the SQL DDL statement.

Verified References: [Databricks Certified Data Engineer Professional], under "Lakehouse" section; Databricks Documentation, under "DESCRIBE EXTENDED" section.

#### NEW QUESTION # 44

A junior data engineer has been asked to develop a streaming data pipeline with a grouped aggregation using DataFrame df. The pipeline needs to calculate the average humidity and average temperature for each non- overlapping five-minute interval. Incremental state information should be maintained for 10 minutes for late- arriving data.

Streaming DataFrame df has the following schema:

"device\_id INT, event\_time TIMESTAMP, temp FLOAT, humidity FLOAT"

Code block:

Choose the response that correctly fills in the blank within the code block to complete this task.

- A. slidingWindow("event\_time", "10 minutes")
- **B. withWatermark("event\_time", "10 minutes")**
- C. awaitArrival("event\_time", "10 minutes")
- D. await("event\_time + '10 minutes'")
- E. delayWrite("event\_time", "10 minutes")

**Answer: B**

Explanation:

The correct answer is A. withWatermark("event\_time", "10 minutes"). This is because the question asks for incremental state information to be maintained for 10 minutes for late-arriving data. The withWatermark method is used to define the watermark for late data. The watermark is a timestamp column and a threshold that tells the system how long to wait for late data. In this case, the watermark is set to 10 minutes. The other options are incorrect because they are not valid methods or syntax for watermarking in Structured Streaming. References:

\* Watermarking: <https://docs.databricks.com/spark/latest/structured-streaming/watermarks.html>

\* Windowed aggregations: <https://docs.databricks.com/spark/latest/structured-streaming/window-operations.html>

#### NEW QUESTION # 45

An analytics team wants to run a short-term experiment in Databricks SQL on the customer transactions Delta table (about 20 billion records) created by the data engineering team. Which strategy should the data engineering team use to ensure minimal downtime and no impact on the ongoing ETL processes?

- **A. Shallow clone the table for the analytics team.**
- B. Give the analytics team direct access to the production table.
- C. Create a new table for the analytics team using a CTAS statement.
- D. Deep clone the table for the analytics team.

**Answer: A**

Explanation:

Comprehensive and Detailed Explanation From Exact Extract:

Exact extract: "A shallow clone creates a copy of the metadata that references the source data files; it is fast and inexpensive." Exact extract: "A deep clone copies the data." Exact extract: "Clones provide a point-in-time snapshot for experimentation without impacting the source." A shallow clone of the production Delta table creates an instantaneous snapshot that references the same data files, so it introduces virtually no downtime or storage overhead and avoids interfering with the ongoing ETL. A deep clone would copy all data (very expensive and slow for 20B rows). CTAS rewrites data and is unnecessary; direct access to prod risks contention and accidental changes.

#### NEW QUESTION # 46

.....

One of the most important functions of our Databricks-Certified-Professional-Data-Engineer preparation questions are that can support almost all electronic equipment, including the computer, mobile phone and so on. If you want to prepare for your exam by the computer, you can buy the Software and APP online versions of our Databricks-Certified-Professional-Data-Engineer training quiz, because these two versions can work well by the computer. Moreover, the APP online version of our Databricks-Certified-

