# Download Free Databricks-Certified-Data-Engineer-Associate Exam PDF | PrepBolt

Don't miss out! Download the latest free Databricks Certified Data Engineer Associate Exam PDF questions. Access real Databricks-Certified-Data-Engineer-Associate dumps with verified answers and boost your chances to pass your certification on the first try with PrepBolt Databricks-Certified-Data-Engineer-Associate exam pdf questions and answers.

## Thank you for Downloading Databricks-Certified-Data-Engineer-Associate exam PDF Demo

https://prepbolt.com/Databricks-Certified-Data-Engineer-Associate.html

# Question 1

Question Type: MultipleChoice

Which of the following commands will return the location of database customer360?

## Options:

A- DESCRIBE LOCATION customer360;
B- DROP DATABASE customer360;
C- DESCRIBE DATABASE customer360;
D- ALTER DATABASE customer360 SET DBPROPERTIES ('location' = '/user'};
E- USE DATABASE customer360;

## Answer:

C

## Explanation:

The commandDESCRIBE DATABASE customer360;will return the location of the database customer360, along with its comment and properties. This command is an alias forDESCRIBE SCHEMA customer360;, which can also be used to get the same information.The other commands will either drop the database, alter its properties, or use it as the current database, but will not return its location12.Reference:

DESCRIBE DATABASE | Databricks on AWS

DESCRIBE DATABASE - Azure Databricks - Databricks SQL

# Question 2

Question Type: MultipleChoice

Which of the following is stored in the Databricks customer's cloud account?

## Options:

A- Databricks web application

B- Cluster management metadata
C- Repos
D- Data
E- Notebooks

## Answer:

D

## Explanation:

The only option that is stored in the Databricks customer's cloud account is data. Data is stored in the customer's cloud storage service, such as AWS S3 or Azure Data Lake Storage. The customer has full control and ownership of their data and can access it directly from their cloud account.

Option A is not correct, as the Databricks web application is hosted and managed by Databricks on their own cloud infrastructure. The customer does not need to install or maintain the web application, but only needs to access it through a web browser.

Option B is not correct, as the cluster management metadata is stored and managed by Databricks on their own cloud infrastructure. The cluster management metadata includes information such as cluster configuration, status, logs, and metrics. The customer can view and manage their clusters through the Databricks web application, but does not have direct access to the cluster management metadata.

Option C is not correct, as the repos are stored and managed by Databricks on their own cloud infrastructure. Repos are version-controlled repositories that store code and data files for Databricks projects. The customer can create and manage their repos through the Databricks web application, but does not have direct access to the repos.

Option E is not correct, as the notebooks are stored and managed by Databricks on their own cloud infrastructure. Notebooks are interactive documents that contain code, text, and visualizations for Databricks workflows. The customer can create and manage their notebooks through the Databricks web application, but does not have direct access to the notebooks.

Databricks Architecture

Databricks Data Sources

Databricks Repos

[Databricks Notebooks]

[Databricks Data Engineer Professional Exam Guide]

# Question 3

Question Type: MultipleChoice

A data engineer needs to create a table in Databricks using data from a CSV file at location /path/to/csv.

They run the following command:

Which of the following lines of code fills in the above blank to successfully complete the task?

## Options:

A- None of these lines of code are needed to successfully complete the task
B- USING CSV
C- FROM CSV
D- USING DELTA
E- FROM 'path/to/csv'

## Answer:

B

# Question 4

Question Type: MultipleChoice

A data engineer needs to apply custom logic to string column city in table stores for a specific use case. In order to apply this custom logic at scale, the data engineer wants to create a SQL user-defined function (UDF).

Which of the following code blocks creates this SQL UDF?

A.

B.

C.

D.

E.

## Options:

A- Option A
B- Option B
C- Option C
D- Option D
E- Option E

## Answer:

A

## Explanation:

https://www.databricks.com/blog/2021/10/20/introducing-sql-user-defined-functions.html

# Question 5

Question Type: MultipleChoice

A dataset has been defined using Delta Live Tables and includes an expectations clause:

CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON VIOLATION FAIL UPDATE

What is the expected behavior when a batch of data containing data that violates these constraints is processed?

## Options:

A- Records that violate the expectation cause the job to fail.
B- Records that violate the expectation are added to the target dataset and flagged as invalid in a field added to the target dataset.
C- Records that violate the expectation are dropped from the target dataset and recorded as invalid in the event log.
D- Records that violate the expectation are added to the target dataset and recorded as invalid in the event log.

## Answer:

C

## Explanation:

The expected behavior when a batch of data containing data that violates the expectation is processed is that the job will fail. This is because the expectation clause has theON VIOLATION FAIL UPDATEoption, which means that if any record in the batch does not meet the expectation, the entire batch will be rejected and the job will fail. This option is useful for enforcing strict data quality rules and preventing invalid data from entering the target dataset.

Option A is not correct, as theON VIOLATION FAIL UPDATEoption does not drop the records that violate the expectation, but fails the entire batch. To drop the records that violate the expectation and record them as invalid in the event log, theON VIOLATION DROP RECORDoption should be used.

Option C is not correct, as theON VIOLATION FAIL UPDATEoption does not drop the records that violate the expectation, but fails the entire batch. To drop the records that violate the expectation and load them into a quarantine table, theON VIOLATION QUARANTINE RECORDoption should be used.

Option D is not correct, as theON VIOLATION FAIL UPDATEoption does not add the records that violate the expectation, but fails the entire batch. To add the records that violate the expectation and record them as invalid in the event log, theON VIOLATION LOG RECORDoption should be used.

Option E is not correct, as theON VIOLATION FAIL UPDATEoption does not add the records that violate the expectation, but fails the entire batch. To add the records that violate the expectation and flag them as invalid in a field added to the target dataset, theON VIOLATION FLAG RECORDoption should be used.

Delta Live Tables Expectations

[Databricks Data Engineer Professional Exam Guide]

# Question 6

Question Type: MultipleChoice

A data engineer needs to use a Delta table as part of a data pipeline, but they do not know if they have the appropriate permissions.

In which of the following locations can the data engineer review their permissions on the table?

## Options:

A- Databricks Filesystem
B- Jobs
C- Dashboards
D- Repos
E- Data Explorer

## Answer:

E

## Explanation:

Data Explorer is a graphical interface that allows you to browse, create, and manage data objects such as databases, tables, and views in your workspace. You can also review and modify the permissions on these data objects using Data Explorer. To access Data Explorer, you can click on the Data icon in the sidebar, or use the %sql magic command in a notebook. You can then select a database and a table, and click on the Permissions tab to view and edit the access control lists (ACLs) for the table. You can also use SQL commands such as SHOW GRANT and GRANT to query and modify the permissions on a Delta table.Reference:

Data Explorer

Access control for Delta tables

SHOW GRANT

[GRANT]

# Question 7

Question Type: MultipleChoice

Which of the following approaches should be used to send the Databricks Job owner an email in the case that the Job fails?

## Options:

A- Manually programming in an alert system in each cell of the Notebook
B- Setting up an Alert in the Job page
C- Setting up an Alert in the Notebook
D- There is no way to notify the Job owner in the case of Job failure
E- MLflow Model Registry Webhooks

## Answer:

B

## Explanation:

To send the Databricks Job owner an email in the case that the Job fails, the best approach is to set up an Alert in the Job page. This way, the Job owner can configure the email address and the notification type for the Job failure event. The other options are either not feasible, not reliable, or not relevant for this task. Manually programming an alert system in each cell of the Notebook is tedious and error-prone. Setting up an Alert in the Notebook is not possible, as Alerts are only available for Jobs and Clusters. There is a way to notify the Job owner in the case of Job failure, so option D is incorrect. MLflow Model Registry Webhooks are used for model lifecycle events, not Job events, so option E is not applicable.Reference:

Add email and system notifications for job events

Alerts

MLflow Model Registry Webhooks

# Question 8

Question Type: MultipleChoice

A data engineering team has two tables. The first table march_transactions is a collection of all retail transactions in the month of March. The second table april_transactions is a collection of all retail transactions in the month of April. There are no duplicate records between the tables.

Which of the following commands should be run to create a new table all_transactions that contains all records from march_transactions and april_transactions without duplicate records?

A.

CREATE TABLE all_transactions AS

SELECT * FROM march_transactions

INNER JOIN SELECT * FROM april_transactions;

B.

CREATE TABLE all_transactions AS

SELECT * FROM march_transactions

UNION SELECT * FROM april_transactions;

C.

CREATE TABLE all_transactions AS

SELECT * FROM march_transactions

OUTER JOIN SELECT * FROM april_transactions;

D.

CREATE TABLE all_transactions AS

SELECT * FROM march_transactions

INTERSECT SELECT * from april_transactions;

E.

CREATE TABLE all_transactions AS

SELECT * FROM march_transactions

MERGE SELECT * FROM april_transactions;

## Options:

A- Option A
B- Option B
C- Option C
D- Option D
E- Option E

## Answer:

B

## Explanation:

The correct command to create a new table that contains all records from two tables without duplicate records is to use theUNIONoperator. The UNION operator combines the results of two queries and removes any duplicate rows. The INNER JOIN, OUTER JOIN, and MERGE operators do not remove duplicate rows, and the INTERSECT operator only returns the rows that are common to both tables. Therefore, option B is the only correct answer.Reference:Databricks SQL Reference - UNION,Databricks SQL Reference - JOIN,Databricks SQL Reference - MERGE, [Databricks SQL Reference - INTERSECT]

# Question 9

Question Type: MultipleChoice

A data engineer wants to create a relational object by pulling data from two tables. The relational object does not need to be used by other data engineers in other sessions. In order to save on storage costs, the data engineer wants to avoid copying and storing physical data.

Which of the following relational objects should the data engineer create?

## Options:

A- Spark SQL Table
B- View
C- Database
D- Temporary view
E- Delta Table

## Answer:

D

## Explanation:

A temporary view is a relational object that is defined in the metastore and points to an existing DataFrame. It does not copy or store any physical data, but only saves the query that defines the view. The lifetime of a temporary view is tied to the SparkSession that was used to create it, so it does not persist across different sessions or applications. A temporary view is useful for accessing the same data multiple times within the same notebook or session, without incurring additional storage costs. The other options are either materialized (A, E), persistent (B, C), or not relational objects .Reference:Databricks Documentation - Temporary View,Databricks Community - How do temp views actually work?,Databricks Community - What's the difference between a Global view and a Temp view?,Big Data Programmers - Temporary View in Databricks.

# Thank You for trying Databricks-Certified-Data-Engineer-Associate PDF Demo

## To try our Databricks-Certified-Data-Engineer-Associate practice exam software visit link below

https://prepbolt.com/Databricks-Certified-Data-Engineer-Associate.html

# Start Your Databricks-Certified-Data-Engineer-Associate Preparation

Use Coupon “SAVE50” for extra 50% discount on the purchase of Practice Test Software. Test your Databricks-Certified-Data-Engineer-Associate preparation with actual exam questions.