

Exam Topics Questions

@thatonecodes

Exam Certified Data Engineer Associate topic 1 question 3 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 3 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following is hosted completely in the control plane of the classic Databricks architecture? Suggested Answer: C

- A. Worker node
- B. JDBC data source
- C. Databricks web application
- D. Databricks Filesystem
- E. Driver node

Answer: C

Timestamp: March 27, 2023, 10:09 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 4 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 4 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following benefits of using the Databricks Lakehouse Platform is provided by Delta Lake? Suggested Answer: D

- A. The ability to manipulate the same data using a variety of languages
- B. The ability to collaborate in real time on a single notebook
- C. The ability to set up alerts for query failures
- D. The ability to support batch and streaming workloads
- E. The ability to distribute complex data operations

Answer: D

Timestamp: April 1, 2023, 3:32 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 5 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 5 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes the storage organization of a Delta table? Suggested Answer: C

- A. Delta tables are stored in a single file that contains data, history, metadata, and other attributes.
- B. Delta tables store their data in a single file and all metadata in a collection of files in a separate location.
- C. Delta tables are stored in a collection of files that contain data, history, metadata, and other attributes.
- D. Delta tables are stored in a collection of files that contain only the data stored within the table.
- E. Delta tables are stored in a single file that contains only the data stored within the table.

Answer: C

Timestamp: April 1, 2023, 3:37 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 6 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 6 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following code blocks will remove the rows where the value in column age is greater than 25 from the existing Delta table my_table and save the updated table? Suggested Answer: C

- A. `SELECT * FROM my_table WHERE age > 25;`
- B. `UPDATE my_table WHERE age > 25;`
- C. `DELETE FROM my_table WHERE age > 25;`
- D. `UPDATE my_table WHERE age <= 25;`
- E. `DELETE FROM my_table WHERE age <= 25;`

Answer: C

Timestamp: April 1, 2023, 3:38 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 9 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 9 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following data lakehouse features results in improved data quality over a traditional data lake? Suggested Answer: B

- A. A data lakehouse provides storage solutions for structured and unstructured data.
- B. A data lakehouse supports ACID-compliant transactions.
- C. A data lakehouse allows the use of SQL queries to examine data.
- D. A data lakehouse stores data in open formats.
- E. A data lakehouse enables machine learning and artificial Intelligence workloads.

Answer: B

Timestamp: April 1, 2023, 3:53 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 11 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 11 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has left the organization. The data team needs to transfer ownership of the data engineer's Delta tables to a new data engineer. The new data engineer is the lead engineer on the data team. Assuming the original data engineer no longer has access, which of the following individuals must be the one to transfer ownership of the Delta tables in Data Explorer? Suggested Answer: C

A. Databricks account representative

B. This transfer is not possible

C. Workspace administrator

D. New lead data engineer

E. Original data engineer

Answer: C

Timestamp: April 1, 2023, 3:59 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 12 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 12 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data analyst has created a Delta table sales that is used by the entire data analysis team. They want help from the data engineering team to implement a series of tests to ensure the data is clean. However, the data engineering team uses Python for its tests rather than SQL. Which of the following commands could the data engineering team use to access sales in PySpark? Suggested Answer: E

A. `SELECT * FROM sales`

B. There is no way to share data between PySpark and SQL.

C. `spark.sql("sales")` D. `spark.delta.table("sales")`

E. `spark.table("sales")`

Answer: E

Timestamp: April 1, 2023, 4:03 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 13 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 13 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following commands will return the location of database customer360? Suggested Answer: C

- 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

Timestamp: April 1, 2023, 4:04 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 14 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 14 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to create a new table containing the names of customers that live in France. They have written the following command: A senior data engineer mentions that it is organization policy to include a table property indicating that the new table includes personally identifiable information (PII). Which of the following lines of code fills in the above blank to successfully complete the task? Suggested Answer: D

A. There is no way to indicate whether a table contains PII.

B. "COMMENT PII"

C. TBLPROPERTIES PII

D. COMMENT "Contains PII"

E. PII

Answer: D

Timestamp: April 2, 2023, 3:35 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 16 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 16 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following commands can be used to write data into a Delta table while avoiding the writing of duplicate records? Suggested Answer: C

- A. DROP
- B. IGNORE
- C. MERGE
- D. APPEND
- E. INSERT

Answer: C

Timestamp: April 2, 2023, 3:29 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 18 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 18 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data analyst has a series of queries in a SQL program. The data analyst wants this program to run every day. They only want the final query in the program to run on Sundays. They ask for help from the data engineering team to complete this task. Which of the following approaches could be used by the data engineering team to complete this task? Suggested Answer: B

- A. They could submit a feature request with Databricks to add this functionality.
- B. They could wrap the queries using PySpark and use Python's control flow system to determine when to run the final query.
- C. They could only run the entire program on Sundays.
- D. They could automatically restrict access to the source table in the final query so that it is only accessible on Sundays.
- E. They could redesign the data model to separate the data used in the final query into a new table.

Answer: B

Timestamp: April 4, 2023, 1:49 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 19 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 19 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer runs a statement every day to copy the previous day's sales into the table transactions. Each day's sales are in their own file in the location "/transactions/raw". Today, the data engineer runs the following command to complete this task: After running the command today, the data engineer notices that the number of records in table transactions has not changed. Which of the following describes why the statement might not have copied any new records into the table?

Suggested Answer: C

- A. The format of the files to be copied were not included with the FORMAT_OPTIONS keyword.
- B. The names of the files to be copied were not included with the FILES keyword.
- C. The previous day's file has already been copied into the table.
- D. The PARQUET file format does not support COPY INTO.
- E. The COPY INTO statement requires the table to be refreshed to view the copied rows.

Answer: C

Timestamp: April 4, 2023, 2:03 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 23 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 23 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is attempting to drop a Spark SQL table my_table. The data engineer wants to delete all table metadata and data. They run the following command: `DROP TABLE IF EXISTS my_table` - While the object no longer appears when they run `SHOW TABLES`, the data files still exist. Which of the following describes why the data files still exist and the metadata files were deleted? Suggested Answer: C

- A. The table's data was larger than 10 GB
- B. The table's data was smaller than 10 GB
- C. The table was external
- D. The table did not have a location
- E. The table was managed

Answer: C

Timestamp: April 1, 2023, 5:56 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 24 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 24 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to create a data entity from a couple of tables. The data entity must be used by other data engineers in other sessions. It also must be saved to a physical location. Which of the following data entities should the data engineer create? Suggested Answer: E

- A. Database
- B. Function
- C. View
- D. Temporary view
- E. Table

Answer: E

Timestamp: April 1, 2023, 6 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 25 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 25 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is maintaining a data pipeline. Upon data ingestion, the data engineer notices that the source data is starting to have a lower level of quality. The data engineer would like to automate the process of monitoring the quality level. Which of the following tools can the data engineer use to solve this problem? Suggested Answer: D

- A. Unity Catalog
- B. Data Explorer
- C. Delta Lake
- D. Delta Live Tables
- E. Auto Loader

Answer: D

Timestamp: April 1, 2023, 6:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 31 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 31 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table. The code block used by the data engineer is below: If the data engineer only wants the query to execute a micro-batch to process data every 5 seconds, which of the following lines of code should the data engineer use to fill in the blank?

Suggested Answer: D

- A. `trigger("5 seconds")`
- B. `trigger()`
- C. `trigger(once="5 seconds")`
- D. `trigger(processingTime="5 seconds")`
- E. `trigger(continuous="5 seconds")`

Answer: D

Timestamp: April 2, 2023, 2:14 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 32 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 32 Topic #: 1

[All Certified Data Engineer Associate Questions]

A dataset has been defined using Delta Live Tables and includes an expectations clause: `CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON VIOLATION DROP ROW` What is the expected behavior when a batch of data containing data that violates these constraints is processed? Suggested Answer: C

- A. Records that violate the expectation are dropped from the target dataset and loaded into a quarantine table.
- 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.
- E. Records that violate the expectation cause the job to fail.

Answer: C

Timestamp: April 2, 2023, 2:18 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 33 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 33 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes when to use the CREATE STREAMING LIVE TABLE (formerly CREATE INCREMENTAL LIVE TABLE) syntax over the CREATE LIVE TABLE syntax when creating Delta Live Tables (DLT) tables using SQL? Suggested Answer: B

- A. CREATE STREAMING LIVE TABLE should be used when the subsequent step in the DLT pipeline is static.
- B. CREATE STREAMING LIVE TABLE should be used when data needs to be processed incrementally.
- C. CREATE STREAMING LIVE TABLE is redundant for DLT and it does not need to be used.
- D. CREATE STREAMING LIVE TABLE should be used when data needs to be processed through complicated aggregations.
- E. CREATE STREAMING LIVE TABLE should be used when the previous step in the DLT pipeline is static.

Answer: B

Timestamp: April 2, 2023, 2:19 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 36 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 36 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has three tables in a Delta Live Tables (DLT) pipeline. They have configured the pipeline to drop invalid records at each table. They notice that some data is being dropped due to quality concerns at some point in the DLT pipeline. They would like to determine at which table in their pipeline the data is being dropped. Which of the following approaches can the data engineer take to identify the table that is dropping the records? Suggested Answer: D

- A. They can set up separate expectations for each table when developing their DLT pipeline.
- B. They cannot determine which table is dropping the records.
- C. They can set up DLT to notify them via email when records are dropped.
- D. They can navigate to the DLT pipeline page, click on each table, and view the data quality statistics.
- E. They can navigate to the DLT pipeline page, click on the "Error" button, and review the present errors.

Answer: D

Timestamp: April 5, 2023, 8:57 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 37 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 37 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a single-task Job that runs each morning before they begin working. After identifying an upstream data issue, they need to set up another task to run a new notebook prior to the original task. Which of the following approaches can the data engineer use to set up the new task? Suggested Answer: B

- A. They can clone the existing task in the existing Job and update it to run the new notebook.
- B. They can create a new task in the existing Job and then add it as a dependency of the original task.
- C. They can create a new task in the existing Job and then add the original task as a dependency of the new task.
- D. They can create a new job from scratch and add both tasks to run concurrently.
- E. They can clone the existing task to a new Job and then edit it to run the new notebook.

Answer: B

Timestamp: April 5, 2023, 1:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 38 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 38 Topic #: 1

[All Certified Data Engineer Associate Questions]

An engineering manager wants to monitor the performance of a recent project using a Databricks SQL query. For the first week following the project's release, the manager wants the query results to be updated every minute. However, the manager is concerned that the compute resources used for the query will be left running and cost the organization a lot of money beyond the first week of the project's release. Which of the following approaches can the engineering team use to ensure the query does not cost the organization any money beyond the first week of the project's release? Suggested Answer: E

- A. They can set a limit to the number of DBUs that are consumed by the SQL Endpoint.
- B. They can set the query's refresh schedule to end after a certain number of refreshes.
- C. They cannot ensure the query does not cost the organization money beyond the first week of the project's release.
- D. They can set a limit to the number of individuals that are able to manage the query's refresh schedule.
- E. They can set the query's refresh schedule to end on a certain date in the query scheduler.

Answer: E

Timestamp: April 5, 2023, 1:09 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 40 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 40 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to schedule their Databricks SQL dashboard to refresh once per day, but they only want the associated SQL endpoint to be running when it is necessary. Which of the following approaches can the data engineer use to minimize the total running time of the SQL endpoint used in the refresh schedule of their dashboard? Suggested Answer: C

- A. They can ensure the dashboard's SQL endpoint matches each of the queries' SQL endpoints.
- B. They can set up the dashboard's SQL endpoint to be serverless.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can reduce the cluster size of the SQL endpoint.
- E. They can ensure the dashboard's SQL endpoint is not one of the included queries' SQL endpoint.

Answer: C

Timestamp: April 2, 2023, 7:25 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 43 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 43 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a Job with multiple tasks that runs nightly. Each of the tasks runs slowly because the clusters take a long time to start. Which of the following actions can the data engineer perform to improve the start up time for the clusters used for the Job? Suggested Answer: D

•

- A. They can use endpoints available in Databricks SQL
- B. They can use jobs clusters instead of all-purpose clusters
- C. They can configure the clusters to be single-node
- D. They can use clusters that are from a cluster pool
- E. They can configure the clusters to autoscale for larger data sizes

Answer: D

Timestamp: April 2, 2023, 7:32 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 44 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 44 Topic #: 1

[All Certified Data Engineer Associate Questions]

A new data engineering team has been assigned to an ELT project. The new data engineering team will need full privileges on the database customers to fully manage the project. Which of the following commands can be used to grant full permissions on the database to the new data engineering team? Suggested Answer: E

- A. GRANT USAGE ON DATABASE customers TO team;
- B. GRANT ALL PRIVILEGES ON DATABASE team TO customers;
- C. GRANT SELECT PRIVILEGES ON DATABASE customers TO teams;
- D. GRANT SELECT CREATE MODIFY USAGE PRIVILEGES ON DATABASE customers TO team;
- E. GRANT ALL PRIVILEGES ON DATABASE customers TO team;

Answer: E

Timestamp: April 2, 2023, 7:33 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 45 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 45 Topic #: 1

[All Certified Data Engineer Associate Questions]

A new data engineering team has been assigned to work on a project. The team will need access to database customers in order to see what tables already exist. The team has its own group team. Which of the following commands can be used to grant the necessary permission on the entire database to the new team? Suggested Answer: E

- A. GRANT VIEW ON CATALOG customers TO team;
- B. GRANT CREATE ON DATABASE customers TO team;
- C. GRANT USAGE ON CATALOG team TO customers;
- D. GRANT CREATE ON DATABASE team TO customers;
- E. GRANT USAGE ON DATABASE customers TO team;

Answer: E

Timestamp: April 2, 2023, 4:21 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 46 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 46 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is running code in a Databricks Repo that is cloned from a central Git repository. A colleague of the data engineer informs them that changes have been made and synced to the central Git repository. The data engineer now needs to sync their Databricks Repo to get the changes from the central Git repository. Which of the following Git operations does the data engineer need to run to accomplish this task? Suggested Answer: C

- A. Merge
- B. Push
- C. Pull
- D. Commit
- E. Clone

Answer: C

Timestamp: Oct. 20, 2023, 3:48 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 47 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 47 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following is a benefit of the Databricks Lakehouse Platform embracing open source technologies? Suggested Answer: E

- A. Cloud-specific integrations
- B. Simplified governance
- C. Ability to scale storage
- D. Ability to scale workloads
- E. Avoiding vendor lock-in

Answer: E

Timestamp: Oct. 19, 2023, 8:36 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 49 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 49 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes a scenario in which a data engineer will want to use a single-node cluster? Suggested Answer: A

- A. When they are working interactively with a small amount of data
- B. When they are running automated reports to be refreshed as quickly as possible
- C. When they are working with SQL within Databricks SQL
- D. When they are concerned about the ability to automatically scale with larger data
- E. When they are manually running reports with a large amount of data

Answer: A

Timestamp: Oct. 20, 2023, 3:55 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 51 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 51 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has realized that the data files associated with a Delta table are incredibly small. They want to compact the small files to form larger files to improve performance. Which of the following keywords can be used to compact the small files? Suggested Answer: B

- A. REDUCE
- B. OPTIMIZE
- C. COMPACTION
- D. REPARTITION
- E. VACUUM

Answer: B

Timestamp: Oct. 20, 2023, 3:57 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 54 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 54 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following can be used to simplify and unify siloed data architectures that are specialized for specific use cases? Suggested Answer: E

- A. None of these
- B. Data lake
- C. Data warehouse
- D. All of these
- E. Data lakehouse

Answer: E

Timestamp: Oct. 20, 2023, 4:01 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 56 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 56 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a Python notebook in Databricks, but they need to use SQL to accomplish a specific task within a cell. They still want all of the other cells to use Python without making any changes to those cells. Which of the following describes how the data engineer can use SQL within a cell of their Python notebook? Suggested Answer: D

- A. It is not possible to use SQL in a Python notebook
- B. They can attach the cell to a SQL endpoint rather than a Databricks cluster
- C. They can simply write SQL syntax in the cell
- D. They can add %sql to the first line of the cell
- E. They can change the default language of the notebook to SQL

Answer: D

Timestamp: Oct. 22, 2023, 4:36 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 57 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 57 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following SQL keywords can be used to convert a table from a long format to a wide format? Suggested Answer: B

A. TRANSFORM

B. PIVOT

C. SUM

D. CONVERT

E. WHERE

Answer: B

Timestamp: Oct. 22, 2023, 4:35 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 58 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 58 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes a benefit of creating an external table from Parquet rather than CSV when using a CREATE TABLE AS SELECT statement? Suggested Answer: C

- A. Parquet files can be partitioned
- B. CREATE TABLE AS SELECT statements cannot be used on files
- C. Parquet files have a well-defined schema
- D. Parquet files have the ability to be optimized
- E. Parquet files will become Delta tables

Answer: C

Timestamp: Oct. 21, 2023, 4:31 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 62 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 62 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to apply custom logic to identify employees with more than 5 years of experience in array column employees in table stores. The custom logic should create a new column exp_employees that is an array of all of the employees with more than 5 years of experience for each row. In order to apply this custom logic at scale, the data engineer wants to use the FILTER higher-order function. Which of the following code blocks successfully completes this task?

Suggested Answer: A

A.

B.

C.

D.

E.

Answer: A

Timestamp: Oct. 22, 2023, 4:56 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 66 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 66 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer that is new to using Python needs to create a Python function to add two integers together and return the sum? Which of the following code blocks can the data engineer use to complete this task? Suggested Answer: D

A.

B.

C.

D.

E.

Answer: D

Timestamp: Oct. 22, 2023, 5:06 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 68 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 68 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is working with two tables. Each of these tables is displayed below in its entirety. The data engineer runs the following query to join these tables together: Which of the following will be returned by the above query? Suggested Answer: C

A.

B.

C.

D.

E.

Answer: C

Timestamp: Oct. 20, 2023, 7:59 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 71 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 71 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has developed a data pipeline to ingest data from a JSON source using Auto Loader, but the engineer has not provided any type inference or schema hints in their pipeline. Upon reviewing the data, the data engineer has noticed that all of the columns in the target table are of the string type despite some of the fields only including float or boolean values. Which of the following describes why Auto Loader inferred all of the columns to be of the string type? Suggested Answer: B

- A. There was a type mismatch between the specific schema and the inferred schema
- B. JSON data is a text-based format
- C. Auto Loader only works with string data
- D. All of the fields had at least one null value
- E. Auto Loader cannot infer the schema of ingested data

Answer: B

Timestamp: Oct. 22, 2023, 5:19 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 72 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 72 Topic #: 1

[All Certified Data Engineer Associate Questions]

A Delta Live Table pipeline includes two datasets defined using STREAMING LIVE TABLE. Three datasets are defined against Delta Lake table sources using LIVE TABLE. The table is configured to run in Development mode using the Continuous Pipeline Mode. Assuming previously unprocessed data exists and all definitions are valid, what is the expected outcome after clicking Start to update the pipeline? Suggested Answer: E

- A. All datasets will be updated once and the pipeline will shut down. The compute resources will be terminated.
- B. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will persist until the pipeline is shut down.
- C. All datasets will be updated once and the pipeline will persist without any processing. The compute resources will persist but go unused.
- D. All datasets will be updated once and the pipeline will shut down. The compute resources will persist to allow for additional testing.
- E. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will persist to allow for additional testing.

Answer: E

Timestamp: Oct. 21, 2023, 12:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 73 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 73 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following data workloads will utilize a Gold table as its source? Suggested Answer: D

- A. A job that enriches data by parsing its timestamps into a human-readable format
- B. A job that aggregates uncleaned data to create standard summary statistics
- C. A job that cleans data by removing malformed records
- D. A job that queries aggregated data designed to feed into a dashboard
- E. A job that ingests raw data from a streaming source into the Lakehouse

Answer: D

Timestamp: Oct. 22, 2023, 5:25 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 76 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 76 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes the type of workloads that are always compatible with Auto Loader? Suggested Answer: A

- A. Streaming workloads
- B. Machine learning workloads
- C. Serverless workloads
- D. Batch workloads
- E. Dashboard workloads

Answer: A

Timestamp: Oct. 22, 2023, 5:36 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 78 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 78 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is using the following code block as part of a batch ingestion pipeline to read from a composable table: Which of the following changes needs to be made so this code block will work when the transactions table is a stream source? Suggested Answer: E

- A. Replace predict with a stream-friendly prediction function
- B. Replace schema(schema) with option ("maxFilesPerTrigger", 1)
- C. Replace "transactions" with the path to the location of the Delta table
- D. Replace format("delta") with format("stream")
- E. Replace spark.read with spark.readStream

Answer: E

Timestamp: Oct. 22, 2023, 5:43 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 79 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 79 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following queries is performing a streaming hop from raw data to a Bronze table?

Suggested Answer: E

A.

B.

C.

D.

E.

Answer: E

Timestamp: Nov. 4, 2023, 6:25 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 80 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 80 Topic #: 1

[All Certified Data Engineer Associate Questions]

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? Suggested Answer: B

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

Answer: B

Timestamp: Oct. 22, 2023, 5:46 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 81 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 81 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following statements regarding the relationship between Silver tables and Bronze tables is always true? Suggested Answer: D

- A. Silver tables contain a less refined, less clean view of data than Bronze data.
- B. Silver tables contain aggregates while Bronze data is unaggregated.
- C. Silver tables contain more data than Bronze tables.
- D. Silver tables contain a more refined and cleaner view of data than Bronze tables.
- E. Silver tables contain less data than Bronze tables.

Answer: D

Timestamp: Oct. 22, 2023, 5:47 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 82 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 82 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineering team has noticed that their Databricks SQL queries are running too slowly when they are submitted to a non-running SQL endpoint. The data engineering team wants this issue to be resolved. Which of the following approaches can the team use to reduce the time it takes to return results in this scenario? Suggested Answer: D

- A. They can turn on the Serverless feature for the SQL endpoint and change the Spot Instance Policy to "Reliability Optimized."
- B. They can turn on the Auto Stop feature for the SQL endpoint.
- C. They can increase the cluster size of the SQL endpoint.
- D. They can turn on the Serverless feature for the SQL endpoint.
- E. They can increase the maximum bound of the SQL endpoint's scaling range.

Answer: D

Timestamp: Oct. 22, 2023, 5:52 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 84 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 84 Topic #: 1

[All Certified Data Engineer Associate Questions]

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

- 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

Timestamp: Oct. 22, 2023, 5:55 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 87 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 87 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has been using a Databricks SQL dashboard to monitor the cleanliness of the input data to a data analytics dashboard for a retail use case. The job has a Databricks SQL query that returns the number of store-level records where sales is equal to zero. The data engineer wants their entire team to be notified via a messaging webhook whenever this value is greater than 0. Which of the following approaches can the data engineer use to notify their entire team via a messaging webhook whenever the number of stores with \$0 in sales is greater than zero?

Suggested Answer: D

- A. They can set up an Alert with a custom template.
- B. They can set up an Alert with a new email alert destination.
- C. They can set up an Alert with one-time notifications.
- D. They can set up an Alert with a new webhook alert destination.
- E. They can set up an Alert without notifications.

Answer: D

Timestamp: Oct. 22, 2023, 5:15 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 88 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 88 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to schedule their Databricks SQL dashboard to refresh every hour, but they only want the associated SQL endpoint to be running when it is necessary. The dashboard has multiple queries on multiple datasets associated with it. The data that feeds the dashboard is automatically processed using a Databricks Job. Which of the following approaches can the data engineer use to minimize the total running time of the SQL endpoint used in the refresh schedule of their dashboard? Suggested Answer: A

- A. They can turn on the Auto Stop feature for the SQL endpoint.
- B. They can ensure the dashboard's SQL endpoint is not one of the included query's SQL endpoint.
- C. They can reduce the cluster size of the SQL endpoint.
- D. They can ensure the dashboard's SQL endpoint matches each of the queries' SQL endpoints.
- E. They can set up the dashboard's SQL endpoint to be serverless.

Answer: A

Timestamp: Oct. 21, 2023, 12:21 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 91 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 91 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which data lakehouse feature results in improved data quality over a traditional data lake?

Suggested Answer: D

- A. A data lakehouse stores data in open formats.
- B. A data lakehouse allows the use of SQL queries to examine data.
- C. A data lakehouse provides storage solutions for structured and unstructured data.
- D. A data lakehouse supports ACID-compliant transactions.

Answer: D

Timestamp: June 14, 2024, 1:42 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 92 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 92 Topic #: 1

[All Certified Data Engineer Associate Questions]

In which scenario will a data team want to utilize cluster pools? Suggested Answer: C

- A. An automated report needs to be version-controlled across multiple collaborators.
- B. An automated report needs to be runnable by all stakeholders.
- C. An automated report needs to be refreshed as quickly as possible.
- D. An automated report needs to be made reproducible.

Answer: C

Timestamp: May 25, 2024, 8:11 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 93 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 93 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is hosted completely in the control plane of the classic Databricks architecture? Suggested Answer: B

- A. Worker node
- B. Databricks web application
- C. Driver node
- D. Databricks Filesystem

Answer: B

Timestamp: Aug. 11, 2024, 3 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 94 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 94 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to determine whether to use the built-in Databricks Notebooks versioning or version their project using Databricks Repos. What is an advantage of using Databricks Repos over the Databricks Notebooks versioning? Suggested Answer: D

- A. Databricks Repos allows users to revert to previous versions of a notebook
- B. Databricks Repos is wholly housed within the Databricks Data Intelligence Platform
- C. Databricks Repos provides the ability to comment on specific changes
- D. Databricks Repos supports the use of multiple branches

Answer: D

Timestamp: Sept. 20, 2024, 3:37 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 95 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 95 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is a benefit of the Databricks Lakehouse Architecture embracing open source technologies?

Suggested Answer: A

- A. Avoiding vendor lock-in
- B. Simplified governance
- C. Ability to scale workloads
- D. Cloud-specific integrations

Answer: A

Timestamp: Aug. 11, 2024, 3:04 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 96 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 96 Topic #: 1

[All Certified Data Engineer Associate Questions]

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 location can the data engineer review their permissions on the table? Suggested Answer: C

- A. Jobs
- B. Dashboards
- C. Catalog Explorer
- D. Repos

Answer: C

Timestamp: Aug. 11, 2024, 3:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 97 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 97 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is running code in a Databricks Repo that is cloned from a central Git repository. A colleague of the data engineer informs them that changes have been made and synced to the central Git repository. The data engineer now needs to sync their Databricks Repo to get the changes from the central Git repository. Which Git operation does the data engineer need to run to accomplish this task? Suggested Answer: B

A. Clone

B. Pull

C. Merge

D. Push

Answer: B

Timestamp: May 25, 2024, 8:12 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 98 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 98 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which file format is used for storing Delta Lake Table? Suggested Answer: B

A. CSV

B. Parquet

C. JSON

D. Delta

Answer: B

Timestamp: May 20, 2024, 10:37 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 99 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 99 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data architect has determined that a table of the following format is necessary: Which code block is used by SQL DDL command to create an empty Delta table in the above format regardless of whether a table already exists with this name? Suggested Answer: A

A. CREATE OR REPLACE TABLE table_name (employeeId STRING, startDate DATE, avgRating FLOAT)

B. CREATE OR REPLACE TABLE table_name WITH COLUMNS (employeeId STRING, startDate DATE, avgRating FLOAT) USING DELTA

C. CREATE TABLE IF NOT EXISTS table_name (employeeId STRING, startDate DATE, avgRating FLOAT)

D. CREATE TABLE table_name AS SELECT employeeId STRING, startDate DATE, avgRating FLOAT

Answer: A

Timestamp: May 25, 2024, 8:12 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 100 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 100 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has been given a new record of data: id STRING = 'a1' rank INTEGER = 6 rating FLOAT = 9.4 Which SQL commands can be used to append the new record to an existing Delta table my_table? Suggested Answer: A

- A. INSERT INTO my_table VALUES ('a1', 6, 9.4)
- B. INSERT VALUES ('a1', 6, 9.4) INTO my_table
- C. UPDATE my_table VALUES ('a1', 6, 9.4)
- D. UPDATE VALUES ('a1', 6, 9.4) my_table

Answer: A

Timestamp: May 25, 2024, 8:13 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 101 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 101 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has realized that the data files associated with a Delta table are incredibly small. They want to compact the small files to form larger files to improve performance. Which keyword can be used to compact the small files? Suggested Answer: A

- A. OPTIMIZE
- B. VACUUM
- C. COMPACTION
- D. REPARTITION

Answer: A

Timestamp: May 25, 2024, 8:13 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 102 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 102 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to create a data entity from a couple of tables. The data entity must be used by other data engineers in other sessions. It also must be saved to a physical location. Which of the following data entities should the data engineer create? Suggested Answer: A

- A. Table
- B. Function
- C. View
- D. Temporary view

Answer: A

Timestamp: June 25, 2024, 3:10 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 103 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 103 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer runs a statement every day to copy the previous day's sales into the table transactions. Each day's sales are in their own file in the location "/transactions/raw". Today, the data engineer runs the following command to complete this task: After running the command today, the data engineer notices that the number of records in table transactions has not changed. What explains why the statement might not have copied any new records into the table? Suggested Answer: C

- A. The format of the files to be copied were not included with the FORMAT_OPTIONS keyword.
- B. The COPY INTO statement requires the table to be refreshed to view the copied rows.
- C. The previous day's file has already been copied into the table.
- D. The PARQUET file format does not support COPY INTO.

Answer: C

Timestamp: May 25, 2024, 8:14 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 104 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 104 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which command can be used to write data into a Delta table while avoiding the writing of duplicate records? Suggested Answer: C

- A. DROP
- B. INSERT
- C. MERGE
- D. APPEND

Answer: C

Timestamp: May 25, 2024, 8:13 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 105 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 105 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data analyst has created a Delta table sales that is used by the entire data analysis team. They want help from the data engineering team to implement a series of tests to ensure the data is clean. However, the data engineering team uses Python for its tests rather than SQL. Which command could the data engineering team use to access sales in PySpark? Suggested Answer: B

- A. `SELECT * FROM sales`
- B. `spark.table("sales")`
- C. `spark.sql("sales")`
- D. `spark.delta.table("sales")`

Answer: B

Timestamp: May 25, 2024, 8:14 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 106 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 106 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has created a new database using the following command: `CREATE DATABASE IF NOT EXISTS customer360;` In which location will the customer360 database be located? Suggested Answer: B

A. `dbfs:/user/hive/database/customer360`

B. `dbfs:/user/hive/warehouse`

C. `dbfs:/user/hive/customer360`

D. `dbfs:/user/hive/database`

Answer: B

Timestamp: May 25, 2024, 8:14 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 107 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 107 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is attempting to drop a Spark SQL table my_table and runs the following command: DROP TABLE IF EXISTS my_table; After running this command, the engineer notices that the data files and metadata files have been deleted from the file system. What is the reason behind the deletion of all these files? Suggested Answer: A

- A. The table was managed
- B. The table's data was smaller than 10 GB
- C. The table did not have a location
- D. The table was external

Answer: A

Timestamp: May 9, 2024, 6:19 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 108 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 108 Topic #: 1

[All Certified Data Engineer Associate Questions]

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? Suggested Answer: B

A. FROM "path/to/csv"

B. USING CSV

C. FROM CSV

D. USING DELTA

Answer: B

Timestamp: Sept. 12, 2024, 5:36 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 109 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 109 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is a benefit of creating an external table from Parquet rather than CSV when using a CREATE TABLE AS SELECT statement? Suggested Answer: C

- A. Parquet files can be partitioned
- B. Parquet files will become Delta tables
- C. Parquet files have a well-defined schema
- D. Parquet files have the ability to be optimized

Answer: C

Timestamp: May 14, 2024, 2:12 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 110 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 110 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which SQL keyword can be used to convert a table from a long format to a wide format? Suggested

Answer: B

A. TRANSFORM

B. PIVOT

C. SUM

D. CONVERT

Answer: B

Timestamp: Aug. 26, 2024, 10:30 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 111 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 111 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a Python variable `table_name` that they would like to use in a SQL query. They want to construct a Python code block that will run the query using `table_name`. They have the following incomplete code block: `_____ (f"SELECT customer_id, spend FROM {table_name}")` What can be used to fill in the blank to successfully complete the task? Suggested Answer: B

A. `spark.delta.sql`

B. `spark.sql`

C. `spark.table`

D. `dbutils.sql`

Answer: B

Timestamp: Sept. 20, 2024, 3:46 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 112 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 112 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is working with two tables. Each of these tables is displayed below in its entirety. The data engineer runs the following query to join these tables together: Suggested Answer: C

A.

B.

C.

D.

Answer: C

Timestamp: May 22, 2024, 10:07 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 113 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 113 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to apply custom logic to identify employees with more than 5 years of experience in array column employees in table stores. The custom logic should create a new column exp_employees that is an array of all of the employees with more than 5 years of experience for each row. In order to apply this custom logic at scale, the data engineer wants to use the FILTER higher-order function. Which code block successfully completes this task? Suggested Answer: A

A.

B.

C.

D.

Answer: A

Timestamp: Sept. 12, 2024, 5:50 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 114 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 114 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer that is new to using Python needs to create a Python function to add two integers together and return the sum? Which code block can the data engineer use to complete this task?

Suggested Answer: D

A.

B.

C.

D.

Answer: D

Timestamp: May 9, 2024, 6:21 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 115 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 115 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table. The code block used by the data engineer is below: Which line of code should the data engineer use to fill in the blank if the data engineer only wants the query to execute a micro-batch to process data every 5 seconds? Suggested Answer: D

- A. `trigger("5 seconds")`
- B. `trigger(continuous="5 seconds")`
- C. `trigger(once="5 seconds")`
- D. `trigger(processingTime="5 seconds")`

Answer: D

Timestamp: Aug. 27, 2024, 12:45 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 116 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 116 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is maintaining a data pipeline. Upon data ingestion, the data engineer notices that the source data is starting to have a lower level of quality. The data engineer would like to automate the process of monitoring the quality level. Which of the following tools can the data engineer use to solve this problem? Suggested Answer: D

- A. Auto Loader
- B. Unity Catalog
- C. Delta Lake
- D. Delta Live Tables

Answer: D

Timestamp: Nov. 21, 2024, 10:21 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 117 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 117 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has three tables in a Delta Live Tables (DLT) pipeline. They have configured the pipeline to drop invalid records at each table. They notice that some data is being dropped due to quality concerns at some point in the DLT pipeline. They would like to determine at which table in their pipeline the data is being dropped. Which approach can the data engineer take to identify the table that is dropping the records? Suggested Answer: D

- A. They can set up separate expectations for each table when developing their DLT pipeline.
- B. They can navigate to the DLT pipeline page, click on the "Error" button, and review the present errors.
- C. They can set up DLT to notify them via email when records are dropped.
- D. They can navigate to the DLT pipeline page, click on each table, and view the data quality statistics.

Answer: D

Timestamp: Aug. 27, 2024, 11:21 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 118 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 118 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is used by Spark to record the offset range of the data being processed in each trigger in order for Structured Streaming to reliably track the exact progress of the processing so that it can handle any kind of failure by restarting and/or reprocessing? Suggested Answer: A

- A. Checkpointing and Write-ahead Logs
- B. Replayable Sources and Idempotent Sinks
- C. Write-ahead Logs and Idempotent Sinks
- D. Checkpointing and Idempotent Sinks

Answer: A

Timestamp: Aug. 27, 2024, 12:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 119 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 119 Topic #: 1

[All Certified Data Engineer Associate Questions]

What describes the relationship between Gold tables and Silver tables? Suggested Answer: A

- A. Gold tables are more likely to contain aggregations than Silver tables.
- B. Gold tables are more likely to contain valuable data than Silver tables.
- C. Gold tables are more likely to contain a less refined view of data than Silver tables.
- D. Gold tables are more likely to contain truthful data than Silver tables.

Answer: A

Timestamp: Aug. 27, 2024, 12:02 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 120 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 120 Topic #: 1

[All Certified Data Engineer Associate Questions]

What describes when to use the CREATE STREAMING LIVE TABLE (formerly CREATE INCREMENTAL LIVE TABLE) syntax over the CREATE LIVE TABLE syntax when creating Delta Live Tables (DLT) tables using SQL? Suggested Answer: B

- A. CREATE STREAMING LIVE TABLE should be used when the subsequent step in the DLT pipeline is static.
- B. CREATE STREAMING LIVE TABLE should be used when data needs to be processed incrementally.
- C. CREATE STREAMING LIVE TABLE should be used when data needs to be processed through complicated aggregations.
- D. CREATE STREAMING LIVE TABLE should be used when the previous step in the DLT pipeline is static.

Answer: B

Timestamp: Aug. 27, 2024, 12:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 121 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 121 Topic #: 1

[All Certified Data Engineer Associate Questions]

A Delta Live Table pipeline includes two datasets defined using STREAMING LIVE TABLE. Three datasets are defined against Delta Lake table sources using LIVE TABLE. The table is configured to run in Production mode using the Continuous Pipeline Mode. What is the expected outcome after clicking Start to update the pipeline assuming previously unprocessed data exists and all definitions are valid? Suggested Answer: C

- A. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will persist to allow for additional testing.
- B. All datasets will be updated once and the pipeline will shut down. The compute resources will persist to allow for additional testing.
- C. All datasets will be updated at set intervals until the pipeline is shut down. The compute resources will be deployed for the update and terminated when the pipeline is stopped.
- D. All datasets will be updated once and the pipeline will shut down. The compute resources will be terminated.

Answer: C

Timestamp: Aug. 27, 2024, 12:15 p.m.

[View on ExamTopics](#)


Exam Certified Data Engineer Associate topic 1 question 122 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 122 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which type of workloads are compatible with Auto Loader? Suggested Answer: A 

- A. Streaming workloads
- B. Machine learning workloads
- C. Serverless workloads
- D. Batch workloads

Answer: A

Timestamp: Oct. 18, 2024, 11:16 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 123 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 123 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has developed a data pipeline to ingest data from a JSON source using Auto Loader, but the engineer has not provided any type inference or schema hints in their pipeline. Upon reviewing the data, the data engineer has noticed that all of the columns in the target table are of the string type despite some of the fields only including float or boolean values. Why has Auto Loader inferred all of the columns to be of the string type? Suggested Answer: B

- A. Auto Loader cannot infer the schema of ingested data
- B. JSON data is a text-based format
- C. Auto Loader only works with string data
- D. All of the fields had at least one null value

Answer: B

Timestamp: July 19, 2024, 4:59 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 124 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 124 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which statement regarding the relationship between Silver tables and Bronze tables is always true?

Suggested Answer: D

- A. Silver tables contain a less refined, less clean view of data than Bronze data.
- B. Silver tables contain aggregates while Bronze data is unaggregated.
- C. Silver tables contain more data than Bronze tables.
- D. Silver tables contain less data than Bronze tables.

Answer: D

Timestamp: May 10, 2024, 1:31 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 125 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 125 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which query is performing a streaming hop from raw data to a Bronze table? Suggested Answer: D

A.

B.

C.

D.

Answer: D

Timestamp: Sept. 12, 2024, 6:08 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 126 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 126 Topic #: 1

[All Certified Data Engineer Associate Questions]

A dataset has been defined using Delta Live Tables and includes an expectations clause: `CONSTRAINT valid_timestamp EXPECT (timestamp > '2020-01-01') ON VIOLATION DROP ROW` What is the expected behavior when a batch of data containing data that violates these constraints is processed? Suggested Answer: C

- 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

Timestamp: May 22, 2024, 10:03 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 127 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 127 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a Job with multiple tasks that runs nightly. Each of the tasks runs slowly because the clusters take a long time to start. Which action can the data engineer perform to improve the start up time for the clusters used for the Job? Suggested Answer: D

- A. They can use endpoints available in Databricks SQL
- B. They can use jobs clusters instead of all-purpose clusters
- C. They can configure the clusters to autoscale for larger data sizes
- D. They can use clusters that are from a cluster pool

Answer: D

Timestamp: Aug. 26, 2024, 8:16 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 128 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 128 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a single-task Job that runs each morning before they begin working. After identifying an upstream data issue, they need to set up another task to run a new notebook prior to the original task. Which approach can the data engineer use to set up the new task? Suggested

Answer: B

- A. They can clone the existing task in the existing Job and update it to run the new notebook.
- B. They can create a new task in the existing Job and then add it as a dependency of the original task.
- C. They can create a new task in the existing Job and then add the original task as a dependency of the new task.
- D. They can create a new job from scratch and add both tasks to run concurrently.

Answer: B

Timestamp: May 10, 2024, 2:56 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 129 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 129 Topic #: 1

[All Certified Data Engineer Associate Questions]

A single Job runs two notebooks as two separate tasks. A data engineer has noticed that one of the notebooks is running slowly in the Job's current run. The data engineer asks a tech lead for help in identifying why this might be the case. Which approach can the tech lead use to identify why the notebook is running slowly as part of the Job? Suggested Answer: C

- A. They can navigate to the Runs tab in the Jobs UI to immediately review the processing notebook.
- B. They can navigate to the Tasks tab in the Jobs UI and click on the active run to review the processing notebook.
- C. They can navigate to the Runs tab in the Jobs UI and click on the active run to review the processing notebook.
- D. They can navigate to the Tasks tab in the Jobs UI to immediately review the processing notebook.

Answer: C

Timestamp: Aug. 26, 2024, 8:10 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 130 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 130 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data analysis team has noticed that their Databricks SQL queries are running too slowly when connected to their always-on SQL endpoint. They claim that this issue is present when many members of the team are running small queries simultaneously. They ask the data engineering team for help. The data engineering team notices that each of the team's queries uses the same SQL endpoint. Which approach can the data engineering team use to improve the latency of the team's queries? Suggested Answer: B

- A. They can increase the cluster size of the SQL endpoint.
- B. They can increase the maximum bound of the SQL endpoint's scaling range.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can turn on the Serverless feature for the SQL endpoint.

Answer: B

Timestamp: Aug. 14, 2024, 6:11 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 131 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 131 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has been using a Databricks SQL dashboard to monitor the cleanliness of the input data to an ELT job. The ELT job has its Databricks SQL query that returns the number of input records containing unexpected NULL values. The data engineer wants their entire team to be notified via a messaging webhook whenever this value reaches 100. Which approach can the data engineer use to notify their entire team via a messaging webhook whenever the number of NULL values reaches 100? Suggested Answer: C

- A. They can set up an Alert with a custom template.
- B. They can set up an Alert with a new email alert destination.
- C. They can set up an Alert with a new webhook alert destination.
- D. They can set up an Alert with one-time notifications.

Answer: C

Timestamp: Aug. 26, 2024, 8:07 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 132 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 90 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer wants to schedule their Databricks SQL dashboard to refresh once per day, but they only want the associated SQL endpoint to be running when it is necessary. Which approach can the data engineer use to minimize the total running time of the SQL endpoint used in the refresh schedule of their dashboard? Suggested Answer: C

- A. They can ensure the dashboard's SQL endpoint matches each of the queries' SQL endpoints.
- B. They can set up the dashboard's SQL endpoint to be serverless.
- C. They can turn on the Auto Stop feature for the SQL endpoint.
- D. They can ensure the dashboard's SQL endpoint is not one of the included queries' SQL endpoint.

Answer: C

Timestamp: Aug. 26, 2024, 8:07 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 132 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 132 Topic #: 1

[All Certified Data Engineer Associate Questions]

A company uses Delta Sharing to collaborate with partners across different cloud providers and geographic regions. What will result in additional costs due to cross-region or egress fees?

Suggested Answer: B

- A. Sharing data within the same cloud provider and region
- B. Transferring data via Delta Sharing across clouds and across different geographic regions
- C. Accessing Delta Sharing data using a VPN within the same data center
- D. Utilizing Delta Sharing for internal data analytics within a single cloud environment

Answer: B

Timestamp: Oct. 4, 2025, 9:01 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 133 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 86 Topic #: 1

[All Certified Data Engineer Associate Questions]

An engineering manager wants to monitor the performance of a recent project using a Databricks SQL query. For the first week following the project's release, the manager wants the query results to be updated every minute. However, the manager is concerned that the compute resources used for the query will be left running and cost the organization a lot of money beyond the first week of the project's release. Which approach can the engineering team use to ensure the query does not cost the organization any money beyond the first week of the project's release? Suggested Answer: C

- A. They can set a limit to the number of DBUs that are consumed by the SQL Endpoint.
- B. They can set the query's refresh schedule to end after a certain number of refreshes.
- C. They can set the query's refresh schedule to end on a certain date in the query scheduler.
- D. They can set a limit to the number of individuals that are able to manage the query's refresh schedule.

Answer: C

Timestamp: Aug. 26, 2024, 8:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 134 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 85 Topic #: 1

[All Certified Data Engineer Associate Questions]

A new data engineering team has been assigned to work on a project. The team will need access to database customers in order to see what tables already exist. The team has its own group team. Which command can be used to grant the necessary permission on the entire database to the new team? Suggested Answer: D

- A. GRANT VIEW ON CATALOG customers TO team;
- B. GRANT CREATE ON DATABASE customers TO team;
- C. GRANT USAGE ON CATALOG team TO customers;
- D. GRANT USAGE ON DATABASE customers TO team;

Answer: D

Timestamp: Aug. 26, 2024, 8:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 135 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 83 Topic #: 1

[All Certified Data Engineer Associate Questions]

A new data engineering team has been assigned to an ELT project. The new data engineering team will need full privileges on the table sales to fully manage the project. Which command can be used to grant full permissions on the database to the new data engineering team? Suggested

Answer: A

- A. GRANT ALL PRIVILEGES ON TABLE sales TO team;
- B. GRANT SELECT CREATE MODIFY ON TABLE sales TO team;
- C. GRANT SELECT ON TABLE sales TO team;
- D. GRANT ALL PRIVILEGES ON TABLE team TO sales;

Answer: A

Timestamp: Aug. 26, 2024, 8:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 136 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 77 Topic #: 1

[All Certified Data Engineer Associate Questions]

Differentiate between all-purpose clusters and jobs clusters. A data engineering team has created a python notebook to load data from cloud storage, this job has been tested and now needs to be scheduled in production. Which would be the best cluster to be used in this case? Suggested

Answer: C

- A. All purpose cluster
- B. Any Unity Catalog-enabled cluster
- C. Jobs Cluster
- D. Serverless SQL warehouse

Answer: C

Timestamp: Dec. 1, 2024, 3 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 136 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 136 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to develop integration tests for an ETL process and deploy a version-controlled, packaged workflow into production using an external job scheduler. Which tool should the data engineer use for this job? Suggested Answer: B

- A. Databricks Connect
- B. Databricks Asset Bundles
- C. Databricks Command Line Interface
- D. Databricks Software Development Kit

Answer: B

Timestamp: Oct. 4, 2025, 9 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 137 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 137 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which Databricks asset bundle format is valid? Suggested Answer: A

A. resources: jobs: hello-job: name: hello-job tasks: - task_key: hello-task existing_cluster_id: 1234-567890-abcde123 notebook_task: notebook_path: ./hello.py

B. "resources":{ "jobs":{ "name":"hello-job", "tasks":{ "task_key":"hello-task", "existing_cluster_id":"1234-567890-abcde123", "notebook_task":{ "notebook_path": ".hello.py" } } }

C. configuration = { "resources":{ "jobs":{ "name":"hello-job", "tasks":{ "task_key":"hello-task", "existing_cluster_id":"1234-567890-abcde123", "notebook_task":{ "notebook_path": ".hello.py" } } } }

D. resources { jobs { name = "hello-job" tasks{ task_key = "hello-task" existing_cluster_id = "1234-567890-abcde123" notebook_task{ notebook_path = ".hello.py" } } } }

Answer: A

Timestamp: Sept. 10, 2025, 10:32 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 138 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 74 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which two components function in the DB platform architecture's control plane? (Choose two.)

Suggested Answer: BE

- A. Virtual Machines
- B. Compute Orchestration
- C. Serverless Compute
- D. Compute
- E. Unity Catalog

Answer: B

Timestamp: Oct. 19, 2024, 6:32 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 139 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 70 Topic #: 1

[All Certified Data Engineer Associate Questions]

In a healthcare provider organization using Delta Lake to store electronic health records (EHRs), a data analyst needs to analyze a snapshot of the patient_records table from two weeks ago before some recent data corrections were applied. What approach should the Data Engineer take to allow the analyst to query that specific prior version? Suggested Answer: B

- A. Truncate the table to remove all data, then reload the data from two weeks ago into the truncated table for the analyst to query.
- B. Identify the version number corresponding to two weeks ago from the Delta transaction log, share that version number with the analyst to query using VERSION AS OF syntax, or export that version to a new Delta table for the analyst to query.
- C. Restore the table to the version from two weeks ago using the RESTORE command, and have the analyst query the restored table.
- D. Use the VACUUM command to remove all versions of the table older than two weeks, then the analyst can query the remaining version.

Answer: B

Timestamp: Oct. 21, 2024, 2:09 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 140 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 69 Topic #: 1

[All Certified Data Engineer Associate Questions]

What can be used to simplify and unify siloed data architectures that are specialized for specific use cases? Suggested Answer: D

- A. Delta Lake
- B. Data lake
- C. Data warehouse
- D. Data lakehouse

Answer: D

Timestamp: Dec. 1, 2024, 3:03 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 141 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 67 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has configured a Structured Streaming job to read from a table, manipulate the data, and then perform a streaming write into a new table. The code block used by the data engineer is below: The data engineer only wants the query to process all of the available data in as many batches as required. Which line of code should the data engineer use to fill in the blank? Suggested Answer: A

- A. `trigger(availableNow=True)`
- B. `trigger(processingTime= "once")`
- C. `trigger(continuous= "once")`
- D. `trigger(once=True)`

Answer: A

Timestamp: Oct. 27, 2024, 8:49 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 142 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 65 Topic #: 1

[All Certified Data Engineer Associate Questions]

Data engineer and data analysts are working together on a data pipeline. The data engineer is working on the raw, bronze, and silver layers of the pipeline using Python, and the data analyst is working on the gold layer of the pipeline using SQL. The raw source of the pipeline is a streaming input. They now want to migrate their pipeline to use Delta Live Tables. Which of the following changes will need to be made to the pipeline when migrating to Delta Live Tables? Suggested Answer: A

- A. The pipeline can have different notebook sources in SQL & Python
- B. The pipeline will need to be written entirely in SQL
- C. The pipeline will need to use a batch source in place of a streaming source
- D. The pipeline will need to be written entirely in Python

Answer: A

Timestamp: Nov. 2, 2024, 12:08 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 142 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 142 Topic #: 1

[All Certified Data Engineer Associate Questions]

An organization has data stored across multiple external systems, including MySQL, Amazon Redshift, and Google BigQuery. The data engineer wants to perform analytics without ingesting directly into Databricks, ensuring unified governance and minimizing data duplication. Which feature of Databricks enables querying these external data sources while maintaining centralized governance? Suggested Answer: B

- A. Delta Lake
- B. Lakehouse Federation
- C. MLflow
- D. Databricks Connect

Answer: B

Timestamp: Oct. 4, 2025, 8:52 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 143 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 64 Topic #: 1

[All Certified Data Engineer Associate Questions]

Identify a scenario to use an external table. A Data Engineer needs to create a parquet bronze table and wants to ensure that it gets stored in a specific path in an external location. Which table can be created in this scenario? Suggested Answer: A

- A. An external table where the location is pointing to specific path in external location.
- B. An external table where the schema has managed location pointing to specific path in external location.
- C. A managed table where the catalog has managed location pointing to specific path in external location.
- D. A managed table where the location is pointing to specific path in external location.

Answer: A

Timestamp: Dec. 1, 2024, 3:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 144 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 144 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer streams customer orders into a Kafka topic (orders_topic) and is currently writing the ingestion script of a DLT pipeline. The data engineer needs to ingest the data from Kafka brokers to DLT using Databricks. What is the correct code for ingesting the data? Suggested

Answer: A

A.

B.

C.

D.

Answer: A

Timestamp: Sept. 21, 2025, 11:16 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 144 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 63 Topic #: 1

[All Certified Data Engineer Associate Questions]

Identify the impact of ON VIOLATION DROP ROW and ON VIOLATION FAIL UPDATE for a constraint violation. A data engineer has created an ETL pipeline using Delta Live table to manage their company travel reimbursement detail, they want to ensure that if the location details has not been provided by the employee, the pipeline needs to be terminated. How can the scenario be implemented? Suggested Answer: B

- A. CONSTRAINT valid_location EXPECT (location = NULL)
- B. CONSTRAINT valid_location EXPECT (location != NULL) ON VIOLATION FAIL UPDATE
- C. CONSTRAINT valid_location EXPECT (location != NULL) ON DROP ROW
- D. CONSTRAINT valid_location EXPECT (location != NULL) ON VIOLATION FAIL

Answer: B

Timestamp: Oct. 26, 2024, 9:52 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 145 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 89 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which two conditions are applicable for governance in Databricks Unity Catalog? (Choose two.)

Suggested Answer: A, E

- A. You can have more than 1 metastore within a databricks account console but only 1 per region.
- B. Both catalog and schema must have a managed location in Unity Catalog provided metastore is not associated with a location
- C. You can have multiple catalogs within metastore and 1 catalog can be associated with multiple metastore
- D. If catalog is not associated with location, it's mandatory to associate schema with managed locations
- E. If metastore is not associated with location, it's mandatory to associate catalog with managed locations

Answer: A

Timestamp: Oct. 26, 2024, 9:51 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 146 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 61 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to access the view created by the sales team, using a shared cluster. The data engineer has been provided usage permissions on the catalog and schema. In order to access the view created by sales team. What are the minimum permissions the data engineer would require in addition? Suggested Answer: B

- A. Needs SELECT permission on the VIEW and the underlying TABLE.
- B. Needs SELECT permission only on the VIEW
- C. Needs ALL PRIVILEGES on the VIEW
- D. Needs ALL PRIVILEGES at the SCHEMA level

Answer: B

Timestamp: Nov. 16, 2024, 11:34 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 147 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 60 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which method should a Data Engineer apply to ensure Workflows are being triggered on schedule?

Suggested Answer: C

- A. Scheduled Workflows require an always-running cluster, which is more expensive but reduces processing latency.
- B. Scheduled Workflows process data as it arrives at configured sources.
- C. Scheduled Workflows can reduce resource consumption and expense since the cluster runs only long enough to execute the pipeline.
- D. Scheduled Workflows run continuously until manually stopped.

Answer: C

Timestamp: Dec. 16, 2024, 2:11 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 148 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 59 Topic #: 1

[All Certified Data Engineer Associate Questions]

The Delta transaction log for the `students` table is shown using the `DESCRIBE HISTORY students` command. A Data Engineer needs to query the table as it existed before the UPDATE operation listed in the log. Which command should the Data Engineer use to achieve this? (Choose two.) Suggested Answer: AB

- A. `SELECT * FROM students@v4`
- B. `SELECT * FROM students TIMESTAMP AS OF '2024-04-22T 14:32:47.000+00:00'`
- C. `SELECT * FROM students FROM HISTORY VERSION AS OF 3`
- D. `SELECT * FROM students VERSION AS OF 5`
- E. `SELECT * FROM students TIMESTAMP AS OF '2024-04-22T 14:32:58.000+00:00'`

Answer: A

Timestamp: Dec. 1, 2024, 3:21 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 148 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 148 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which compute option should be chosen in a scenario where small-scale ad-hoc Python scripts need to be run at high frequency and should wind down quickly after these queries have finished running? Suggested Answer: B

- A. All-purpose Cluster
- B. Job Cluster
- C. Serverless Compute
- D. SQL Warehouse

Answer: B

Timestamp: Sept. 21, 2025, 9:58 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 149 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 55 Topic #: 1

[All Certified Data Engineer Associate Questions]

An engineering manager uses a Databricks SQL query to monitor ingestion latency for each data source. The manager checks the results of the query every day, but they are manually rerunning the query each day and waiting for the results. Which of the following approaches can the manager use to ensure the results of the query are updated each day? Suggested Answer: C

- A. They can schedule the query to refresh every 1 day from the SQL endpoint's page in Databricks SQL.
- B. They can schedule the query to refresh every 12 hours from the SQL endpoint's page in Databricks SQL.
- C. They can schedule the query to refresh every 1 day from the query's page in Databricks SQL.
- D. They can schedule the query to run every 12 hours from the Jobs UI.

Answer: C

Timestamp: Dec. 16, 2024, 2:18 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 150 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 53 Topic #: 1

[All Certified Data Engineer Associate Questions]

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? Suggested Answer: A

A.

B.

C.

D.

Answer: A

Timestamp: Dec. 5, 2024, 11:22 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 151 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 52 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has realized that they made a mistake when making a daily update to a table. They need to use Delta time travel to restore the table to a version that is 3 days old. However, when the data engineer attempts to time travel to the older version, they are unable to restore the data because the data files have been deleted. Which of the following explains why the data files are no longer present? Suggested Answer: A

- A. The VACUUM command was run on the table
- B. The TIME TRAVEL command was run on the table
- C. The DELETE HISTORY command was run on the table
- D. The OPTIMIZE command was run on the table

Answer: A

Timestamp: Dec. 16, 2024, 2:26 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 152 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 50 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following describes the relationship between Bronze tables and raw data? Suggested Answer: C

- A. Bronze tables contain less data than raw data files.
- B. Bronze tables contain more truthful data than raw data.
- C. Bronze tables contain raw data with a schema applied.
- D. Bronze tables contain a less refined view of data than raw data.

Answer: C

Timestamp: Dec. 16, 2024, 2:29 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 153 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 153 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is the maximum output supported by a job cluster to ensure a notebook does not fail?

Suggested Answer: C 30MBs

A. 25MBs

B. 10MBs

C. 30MBs

D. 15MBs

Answer: C

Timestamp: Sept. 9, 2025, 1:28 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 153 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 48 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer only wants to execute the final block of a Python program if the Python variable `day_of_week` is equal to 1 and the Python variable `review_period` is `True`. Which of the following control flow statements should the data engineer use to begin this conditionally executed code block? Suggested Answer: D

- A. `if day_of_week = 1 and review_period:`
- B. `if day_of_week = 1 and review_period = "True":`
- C. `if day_of_week = 1 & review_period: = "True":`
- D. `if day_of_week == 1 and review_period:`

Answer: D

Timestamp: Dec. 5, 2024, 11:26 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 154 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 42 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following must be specified when creating a new Delta Live Tables pipeline?

Suggested Answer: B

- A. A key-value pair configuration
- B. At least one notebook library to be executed
- C. A path to cloud storage location for the written data
- D. A location of a target database for the written data

Answer: B

Timestamp: Dec. 16, 2024, 2:32 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 155 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 41 Topic #: 1

[All Certified Data Engineer Associate Questions]

In which of the following scenarios should a data engineer select a Task in the Depends On field of a new Databricks Job Task? Suggested Answer: B

- A. When another task needs to be replaced by the new task
- B. When another task needs to successfully complete before the new task begins
- C. When another task has the same dependency libraries as the new task
- D. When another task needs to use as little compute resources as possible

Answer: B

Timestamp: Dec. 16, 2024, 2:34 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 156 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 39 Topic #: 1

[All Certified Data Engineer Associate Questions]

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? Suggested Answer: B

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;`

Answer: B

Timestamp: Dec. 16, 2024, 3:51 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 157 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 35 Topic #: 1

[All Certified Data Engineer Associate Questions]

How can Git operations must be performed outside of Databricks Repos? Suggested Answer: C

A. Commit

B. Pull

C. Merge

D. Clone

Answer: C

Timestamp: Dec. 16, 2024, 3:53 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 158 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 34 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has joined an existing project and they see the following query in the project repository: `CREATE STREAMING LIVE TABLE loyal_customers AS SELECT customer_id FROM STREAM(LIVE.customers) WHERE loyalty_level = 'high'`; Which of the following describes why the `STREAM` function is included in the query? Suggested Answer: C

- A. The `STREAM` function is not needed and will cause an error.
- B. The data in the customers table has been updated since its last run.
- C. The customers table is a streaming live table.
- D. The customers table is a reference to a Structured Streaming query on a PySpark DataFrame.

Answer: C

Timestamp: Dec. 16, 2024, 3:55 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 159 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 30 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which Structured Streaming query is performing a hop from a Silver table to a Gold table?

Suggested Answer: D

A.

B.

C.

D.

Answer: D

Timestamp: Dec. 5, 2024, 11:36 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 160 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 29 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data organization leader is upset about the data analysis team's reports being different from the data engineering team's reports. The leader believes the siloed nature of their organization's data engineering and data analysis architectures is to blame. Which of the following describes how a data lakehouse could alleviate this issue? Suggested Answer: B

- A. Both teams would respond more quickly to ad-hoc requests
- B. Both teams would use the same source of truth for their work
- C. Both teams would reorganize to report to the same department
- D. Both teams would be able to collaborate on projects in real-time

Answer: B

Timestamp: Dec. 5, 2024, 11:38 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 161 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 161 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is managing a data pipeline in Databricks, where multiple Delta tables are used for various transformations. The team wants to track how data flows through the pipeline, including identifying dependencies between Delta tables, notebooks, jobs, and dashboards. The data engineer is utilizing the Unity Catalog lineage feature to monitor this process. How does Unity Catalog's data lineage feature support the visualization of relationships between Delta tables, notebooks, jobs, and dashboards? Suggested Answer: D

- A. Unity Catalog lineage visualizes dependencies between Delta tables, notebooks, and jobs, but does not provide column-level tracing or relationships with dashboards.
- B. Unity Catalog lineage only supports visualizing relationships at the table level and does not extend to notebooks, jobs, or dashboards.
- C. Unity Catalog lineage provides an interactive graph that tracks dependencies between tables and notebooks but excludes any job-related dependencies or dashboard visualizations.
- D. Unity Catalog provides an interactive graph that visualizes the dependencies between Delta tables, notebooks, jobs, and dashboards, while also supporting column-level tracking of data transformations.

Answer: D

Timestamp: Sept. 9, 2025, 1:43 a.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 161 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 28 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data analyst has developed a query that runs against Delta table. They want help from the data engineering team to implement a series of tests to ensure the data returned by the query is clean. However, the data engineering team uses Python for its tests rather than SQL. Which of the following operations could the data engineering team use to run the query and operate with the results in PySpark? Suggested Answer: C

A. `SELECT * FROM sales`

B. `spark.delta.table`

C. `spark.sql`

D. `spark.table`

Answer: C

Timestamp: Dec. 5, 2024, 11:39 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 162 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 27 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has a Job that has a complex run schedule, and they want to transfer that schedule to other Jobs. Rather than manually selecting each value in the scheduling form in Databricks, which of the following tools can the data engineer use to represent and submit the schedule programmatically? Suggested Answer: D

- A. `pyspark.sql.types.DateType`
- B. `datetime`
- C. `pyspark.sql.types.TimestampType`
- D. Cron syntax

Answer: D

Timestamp: Dec. 16, 2024, 4 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 162 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 162 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to conduct Exploratory Analysis on data residing in a database that is within the company's custom-defined network in the cloud. The data engineer is using SQL for this task. Which type of SQL Warehouse will enable the data engineer to process large numbers of queries quickly and cost-effectively? Suggested Answer: D

A. Serverless compute for notebooks

B. Pro SQL Warehouse

C. Classic SQL Warehouse

D. Serverless SQL Warehouse

Answer: D

Timestamp: Sept. 12, 2025, 12:49 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 163 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 26 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer and data analyst are working together on a data pipeline. The data engineer is working on the raw, bronze, and silver layers of the pipeline using Python, and the data analyst is working on the gold layer of the pipeline using SQL. The raw source of the pipeline is a streaming input. They now want to migrate their pipeline to use Delta Live Tables. Which of the following changes will need to be made to the pipeline when migrating to Delta Live Tables? Suggested Answer: D

- A. The pipeline will need to be written entirely in Python
- B. The pipeline will need to stop using the medallion-based multi-hop architecture
- C. The pipeline will need to be written entirely in SQL
- D. The pipeline will need to use a batch source in place of a streaming source

Answer: D

Timestamp: Dec. 5, 2024, 11:43 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 164 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 164 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is debugging a Python notebook in Databricks that processes a dataset using PySpark. The notebook fails with an error during a DataFrame transformation. The engineer wants to inspect the state of variables, such as the input DataFrame and intermediate results, to identify where the error occurs. Which tool should the engineer use to debug the notebook and inspect the values of variables like DataFrames? Suggested Answer: B

- A. Use the Databricks CLI to download and analyze driver logs for detailed error messages
- B. Use the Python Notebook Interactive Debugger to set breakpoints and inspect variable values in real-time
- C. Use the Ganglia UI to monitor cluster resource usage and identify hardware issues
- D. Use the Spark UI to analyze the execution plan and identify stages where the job failed

Answer: B

Timestamp: Sept. 20, 2025, 2:15 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 164 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 22 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs access to a table `new_table`, but they do not have the correct permissions. They can ask the table owner for permission, but they do not know who the table owner is. Which of the following approaches can be used to identify the owner of `new_table`? Suggested Answer: C

- A. Review the Permissions tab in the table's page in Data Explorer
- B. There is no way to identify the owner of the table
- C. Review the Owner field in the table's page in Data Explorer
- D. Review the Owner field in the table's page in the cloud storage solution

Answer: C

Timestamp: Dec. 16, 2024, 4:02 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 165 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 21 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to create a table in Databricks using data from their organization's existing SQLite database. They run the following command: Which of the following lines of code fills in the above blank to successfully complete the task? Suggested Answer: A

- A. `org.apache.spark.sql.jdbc`
- B. `autoloader`
- C. `org.apache.spark.sql.sqlite`
- D. `sqlite`

Answer: A

Timestamp: Dec. 16, 2024, 4:03 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 166 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 20 Topic #: 1

[All Certified Data Engineer Associate Questions]

In which of the following scenarios should a data engineer use the MERGE INTO command instead of the INSERT INTO command? Suggested Answer: D

- A. When the location of the data needs to be changed
- B. When the target table is an external table
- C. When the source is not a Delta table
- D. When the target table cannot contain duplicate records

Answer: D

Timestamp: Dec. 16, 2024, 4:03 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 166 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 166 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is developing a small proof of concept in a notebook. When running the entire notebook, the Cluster usage spikes. The data engineer wants to keep the development requirements and get real-time results. Which Cluster meets these requirements? Suggested

Answer: A

- A. All Purpose Cluster with autoscaling
- B. Job Cluster with Photon enabled and autoscaling
- C. Job Cluster with autoscaling enabled
- D. All-Purpose Cluster with a large fixed memory size

Answer: A

Timestamp: Sept. 20, 2025, 2:20 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 167 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 17 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is designing a data pipeline. The source system generates files in a shared directory that is also used by other processes. As a result, the files should be kept as is and will accumulate in the directory. The data engineer needs to identify which files are new since the previous run in the pipeline, and set up the pipeline to only ingest those new files with each run. Which of the following tools can the data engineer use to solve this problem? Suggested Answer: D

- A. Unity Catalog
- B. Delta Lake
- C. Databricks SQL
- D. Auto Loader

Answer: D

Timestamp: Dec. 16, 2024, 4:04 p.m.

[View on ExamTopics](#)


Exam Certified Data Engineer Associate topic 1 question 168 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 15 Topic #: 1

[All Certified Data Engineer Associate Questions]

What is stored in the Databricks customer's cloud account? Suggested Answer: D 

- A. Databricks web application
- B. Cluster management metadata
- C. Notebooks
- D. Data

Answer: D

Timestamp: Dec. 16, 2024, 4:05 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 169 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 10 Topic #: 1

[All Certified Data Engineer Associate Questions]

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? Suggested Answer: D

A. Spark SQL Table

B. View

C. Delta Table

D. Temporary view

Answer: D

Timestamp: Dec. 16, 2024, 4:06 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 170 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 8 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following commands will return the number of null values in the member_id column?

Suggested Answer: C

- A. `SELECT count(member_id) FROM my_table;`
- B. `SELECT count(member_id) - count_null(member_id) FROM my_table;`
- C. `SELECT count_if(member_id IS NULL) FROM my_table;`
- D. `SELECT null(member_id) FROM my_table;`

Answer: C

Timestamp: Dec. 16, 2024, 4:07 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 171 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 7 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which tool is used by Auto Loader to process data incrementally? Suggested Answer: B

- A. Checkpointing
- B. Spark Structured Streaming
- C. Databricks SQL
- D. Unity Catalog

Answer: B

Timestamp: Dec. 16, 2024, 4:07 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 172 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 2 Topic #: 1

[All Certified Data Engineer Associate Questions]

Which of the following benefits is provided by the array functions from Spark SQL? Suggested

Answer: D

- A. An ability to work with data in a variety of types at once
- B. An ability to work with data within certain partitions and windows
- C. An ability to work with time-related data in specified intervals
- D. An ability to work with complex, nested data ingested from JSON files

Answer: D

Timestamp: Dec. 16, 2024, 4:32 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 173 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 1 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer is working with two tables. Each of these tables is displayed below in its entirety. The data engineer runs the following query to join these tables together: Which of the following will be returned by the above query? Suggested Answer: C

A.

B.

C.

D.

Answer: C

Timestamp: Dec. 9, 2024, 7:11 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 175 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 175 Topic #: 1

[All Certified Data Engineer Associate Questions]

A Databricks single-task workflow fails at the last task due to an error in a notebook. The data engineer fixes the mistake in the notebook. What should the data engineer do to rerun the workflow?
Suggested Answer: A

- A. Repair the task
- B. Rerun the pipeline
- C. Restart the cluster
- D. Switch the cluster

Answer: A

Timestamp: Sept. 20, 2025, 2:39 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 177 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 177 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer has written a function in a Databricks Notebook to calculate the population of bacteria in a given medium. Analysts use this function in the notebook and sometimes provide input arguments of the wrong data type, which can cause errors during execution. Which Databricks feature will help the data engineer quickly identify if an incorrect data type has been provided as input? Suggested Answer: B

- A. The Spark User interface has a debug tab that contains the variables that are used in this session.
- B. The Databricks debugger enables breakpoints that will raise an error if the wrong data type is submitted.
- C. The Databricks debugger enables the use of a variable explorer to see at a glance the value of the variables.
- D. The Data Engineer should add print statements to find out what the variable is.

Answer: B

Timestamp: Sept. 20, 2025, 2:53 p.m.

[View on ExamTopics](#)

Exam Certified Data Engineer Associate topic 1 question 179 discussion

Actual exam question from

Databricks's Certified Data Engineer Associate

Question #: 179 Topic #: 1

[All Certified Data Engineer Associate Questions]

A data engineer needs to parse only png files in a directory that contains files with different suffixes. Which code should the data engineer use to achieve this task? Suggested Answer: B

A. `df = spark.readStream.format("cloudFiles") \ .option("cloudFiles.format", "binaryFile") .append("/*.png")`

B. `df = spark.readstream. format("cloudFiles") .option("cloudFiles.format", "binaryFile") .option("pathGlobfilter", "*.png") .load()`

C. `df = spark.readStream.format("cloudFiles") \ .option("cloudFiles.format", "binaryFile") .option("pathGlobfilter", "*.png") \ .append()`

D. `df = spark.readstream.format("cloudFiles") \ .option("cloudFiles.format", "binaryFile") .load("/*.png")`

Answer: B

Timestamp: Sept. 9, 2025, 5:01 a.m.

[View on ExamTopics](#)

