

TOP 10

Must-Know Azure
Databricks & Spark
Interview Questions



Abhishek Agrawal
Azure Data Engineer



Q. What is the difference between managed and unmanaged tables in Databricks?

Ans. Managed tables handle both data and metadata, so dropping them removes metadata and data. Unmanaged tables only handle metadata, so dropping them retains data and removes metadata.

Q. What are the different types of clusters in Databricks?

Ans. General Purpose Compute: Provisioned compute used to analyze data in notebooks. You can create, terminate, and restart this compute using the UI, CLI, or REST API.

Job compute: Provisioned compute used to run automated jobs.

Q. What are the different types of views in Databricks?

Ans. Databricks offers Global views, which are cluster-level and not persistent, Views, which are standard, and Temp views, limited to a notebook and not accessible across notebooks.

Q. What are transformations and actions in Databricks?

Ans. Transformations in Databricks are operations on data that produce new datasets, while actions are operations that trigger computation and return results.

Q. What is left semi and left anti join in Databricks?

Ans. Left semi-join returns rows from the left table that have a match in the right table, while left anti-join returns rows from the left table that have no match in the right table.

Q. What is a Delta table and Delta Lake?

Ans. Delta tables are parquet tables with transaction logs for ACID transactions, while Delta Lake is an open-source storage layer that brings ACID transactions to Apache Spark and big data workloads.

Q. How can you mount Azure Data Lake in Databricks?

Ans. Azure Data Lake can be mounted in Databricks using the `dbutils.fs.mount` command.

Q. How do you read a CSV file in Databricks?

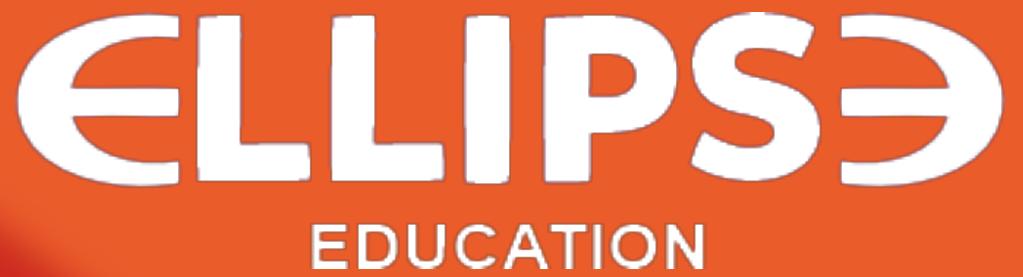
Ans. To read a CSV file in Databricks, you can use the `spark.read.csv` function.

Q. What is the vacuum process in Databricks?

Ans. The vacuum process in Databricks removes files that are no longer needed from the table's transaction log, helping to reclaim storage space.

Q. What is lazy evaluation in Databricks?

Ans. Lazy evaluation in Databricks means that Spark transformations are not executed immediately but instead are queued up and executed only when an action is called.



**Follow for more
content like this**



Abhishek Agrawal
Azure Data Engineer

