

Discover What Was
Asked in Abhishek's
Interview at




Abhishek Agrawal
Azure Data Engineer

Swipe Right



Q. What is the difference between DataFrame, RDD, and Dataset in Databricks?

Q. How do you handle null values in Databricks?

Q. What is lazy evaluation in Databricks?

Q. What is the difference between Persist and Cache in Databricks?

Q. What is the difference between a Global Temp View and a Temp View in Databricks?



Q. Why do we move from a relational database to a blob, data lake, delta lake, and finally a lakehouse?

Q. What's the difference between DataFrame, RDD, and Dataset?

Q. What is the difference between Persist and Cache?

Q. What is a Broadcast Join and a Sort Merge Join?

Q. What are Global Temp View and Temp View?



Q. What is Unity Catalog in Databricks and why is it important?

Q. What is the Catalyst Optimizer in Spark?

Q. What is Adaptive Query Execution in Databricks?

Q. What is meant by Data Skewness and how do you deal with it in Databricks?

Q. What are the roles and responsibilities of worker and driver in Spark architecture?



Q. What are some Spark optimization techniques you have applied?

Q. What is Ganglia Metrics in Databricks and how do you use it?

Q. What is the difference between transformations and actions in Spark?

Q. How do you implement Slowly Changing Dimension (SCD) Type 2 in Databricks?

Q. What is the difference between a relational database, blob storage, data lake, data warehouse, delta lake, and a lakehouse architecture?



**Follow for more
content like this**



Abhishek Agrawal

Azure Data Engineer

