Hashedin
by Deloitte

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹16-20LPA
- **Total Rounds:** 2
- **Top Required Skills:**

1. SQL
2. PySpark / Python
3. Databricks
4. Azure
5. System Architecture

Ankita Gulati

Shubh Goyal

Round 1
ADF, SQL & Python Assessment

Azure Data Factory

1. Can you introduce yourself?
2. Can you explain your recent project in detail?
3. What is Azure Data Factory (ADF), and why is it used?
4. What is an Integration Runtime (IR) in ADF, and what are its different types?
5. How do you create a Self-hosted Integration Runtime?
6. How many activities have you used in your ADF pipelines?
7. From how many different sources have you extracted data in your project?
8. How did you implement parameterization in your pipelines?
9. What is a Linked Service in ADF?

Ankita Gulati Shubh Goyal

SQL

10. Can you write a SQL query to fetch the second highest employee salary?
11. What is a window function in SQL? Can you give an example?
12. What is the difference between RANK() and DENSE_RANK() functions?
13. What is a Common Table Expression (CTE), and can you write a query using it?
14. What are some SQL query optimization techniques you have used?
15. Can you write a SQL query to fetch the top three products by revenue for each region and category?

Databricks / PySpark

16. How do you establish a connection between Azure Data Lake and Databricks?
17. Can you write a PySpark code snippet to read data from Azure Data Lake?

18. What are the different write modes in Spark, and when would you use them?
19. What is an Anti Join in Spark? Can you provide an example?
20. Can you explain the Spark Architecture?
21. What are some PySpark optimization techniques you have applied in your projects?

Python

22. What is the difference between a List and a Tuple in Python?
23. What is a static method in Python, and when would you use it?
24. What is a Lambda function in Python? Can you provide an example?
25. What is the difference between an Object and a Class in Python? Please provide an example.
26. What are the different data types available in Python?

AWS (Other Cloud Exposure)

27. What is AWS Glue, and how is it used?
28. What is an AWS Lambda function, and where have you used it?
29. Can you explain an end-to-end project you built using AWS services?
30. How do you create an AWS Glue Crawler?
31. Have you worked with Amazon Redshift? If yes, can you explain your experience?

Round 2
DBT, ETL & Cloud Integration

SQL, Python, Databricks / Spark

1. Can you introduce yourself and explain your project?
2. Can you write a SQL query to fetch those customers who placed an order exactly one month ago?
3. Can you write Python code to find all pairs of numbers whose sum is 7 and
----> input_list = [1, 2, 3, 4, 5, 6]
4. What is Databricks, and can you explain its architecture?
5. How does Apache Spark work? Can you explain its architecture?
6. Can you write PySpark code to demonstrate some transformations you used in your project?
7. What are Actions in Spark? Can you provide examples?

Ankita Gulati Shubh Goyal

8. What is an Anti Join in Spark? Can you explain with an example?
9. What is a Delta Live Table (DLT), and why would you use it instead of Delta Lake or Data Lake?
10. What are some of the important features of Delta Lake?

Power BI & Integration

11. How do you connect Databricks to Power BI?

ETL & ADF

12. Can you explain an ETL pipeline you implemented to move data from an On-Premises database to Azure Data Lake Storage (ADLS)?
13. How do you handle incremental data loading in your pipelines?
14. What are two limitations of the Copy Activity in ADF?
15. What is Dynamic Data Masking in SQL, and can you provide an example?

Ankita Gulati Shubh Goyal

16. What is Encrypted Data, and can you give an example?
17. What is the difference between Data Masking and Data Encryption?
18. What are the different types of triggers available in ADF?

Synapse Analytics

19. How do you access Azure Data Lake Storage (ADLS) data from Synapse Analytics?
20. What is the difference between a Dedicated SQL Pool and a Serverless SQL Pool in Synapse?

Thank You

Best of luck with your upcoming interviews — you've got this!

Ankita Gulati

Shubh Goyal