



Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



Job Details

- **Position:** Data Engineer
- **Experience:** 2+ years
- **Location:** Pan India
- **Work mode:** Remote
- **Compensation:** ₹8-12 LPA
- **Total Rounds:** 2
- **Top Required Skills:**
 1. SQL
 2. PySpark / Python
 3. Cloud Data Engineering
 4. ETL / Data Modeling
 5. Big Data & Streaming
 6. System Design

Round 1

Technical + Managerial

1. Can you introduce yourself and give a brief overview of your professional background?
2. Can you walk me through your current project, including the data pipeline architecture, technologies used, and your specific responsibilities?
3. What were the biggest challenges you faced in your projects, and how did you overcome them?
4. How do you ensure the scalability, reliability, and cost-effectiveness of your data pipelines?
5. What are the different types of SQL joins? Can you explain their differences with examples?
6. Given a dataset, write queries using RANK, DENSE_RANK, ROW_NUMBER, FIRST_VALUE, and LAST_VALUE to solve business problems.
7. Write a SQL query to find the second highest salary in a table without using LIMIT or TOP.
8. Given a sales transaction table, write a query to calculate the cumulative revenue by month.

9. How would you identify and remove duplicate records in a SQL table?
10. Can you explain the concept of window functions and provide real-world use cases?
11. What AWS services have you used in your projects?
12. Can you explain the difference between AWS Lambda and AWS Glue? In which scenarios would you use each?
13. How do you manage and monitor your ETL jobs in AWS?
14. What are the different S3 storage classes, and when would you use them?
15. How do you move a data pipeline from development to production in AWS?
16. If your data pipeline fails during production, how would you identify the issue and ensure minimal downtime?
17. Suppose business stakeholders request a data change that conflicts with existing pipeline logic—how would you handle it?
18. How do you prioritize tasks when working under tight deadlines in Agile projects?

19. Can you describe a situation where you had to collaborate with cross-functional teams (e.g., data scientists, business analysts)?
20. Explain Star Schema vs Snowflake Schema. Which one is better for reporting and why?
21. What are Slowly Changing Dimensions (SCDs)? How would you implement SCD Type 2 in SQL?
22. What is partitioning and bucketing in Spark or Hive?
23. Explain the CAP theorem in distributed systems and how it applies to databases like DynamoDB or Cassandra.
24. How do you handle schema evolution in a data lake?
25. Write a Python function to check if two strings are anagrams.
26. Given a list of numbers, write a function to return the top 3 largest numbers without using sort().
27. How would you use Python to perform data validation before loading data into a database?

Round 2

HR Discussion

1. Can you briefly introduce yourself and summarize your professional background?
2. Why are you interested in joining Wipro?
3. Why do you want to switch from your current company?
4. What are your salary expectations?
5. What is your notice period?
6. Describe a situation where you had a conflict at work. How did you resolve it?
7. How do you keep yourself updated with the latest data engineering trends?

Ankita Gulati

Shubh Goyal

Thank You

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

