fractal

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Senior Data Engineer
- **Experience:** 7.5+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹32-36 LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1. SQL
2. PySpark / Python
3. Cloud Data Engineering
4. ETL / Data Modeling
5. Big Data & Streaming
6. System Design

Ankita Gulati

Shubh Goyal

Round 1
Technical Discussion

1. Walk me through a recent data engineering project you have worked on, focusing on architecture, tools, and overall data flow.
2. What ETL processes did you implement, and how did you ensure performance and reliability?
3. Explain fact and dimension tables with examples. How would you design a star schema vs. snowflake schema?
4. What are Slowly Changing Dimensions (SCDs)? Describe different types and their use cases.
5. What is the difference between schema-on-read and schema-on-write? Which one does Hive follow?
6. Write a Python function to reverse a string without using built-in methods.
7. Find the first non-repeating character in a string using a dictionary. Example: "abxabyz" → Output: "x".
8. Explain lambda functions. When would you use them?

9. What are generators in Python? Why are they memory-efficient?

10. Python coding: Group words by anagrams. Example: [“eat”, “tea”, “ate”] → [[“eat”, “tea”, “ate”]].

11. Write a SQL query to fetch the 3rd highest salary without using TOP or LIMIT.

12. Given a sales table, find products with strictly increasing sales across months. (Use LAG() or window functions.)

13. Write a query to identify customers who purchased the same product in consecutive months.

14. Explain the difference between UNION and UNION ALL with examples.

15. SQL performance tuning: How would you optimize a slow query with multiple joins?

16. What is the difference between repartition() and coalesce() in Spark?

17. Explain Spark’s lazy evaluation. What happens during execution plan generation?

18. Explain all types of joins in Spark. Specifically, how do left anti join and semi join differ?

19. What is data skewness in Spark, and how do you handle it?
20. What Spark optimizations have you used (e.g., broadcast joins, partition pruning, caching)?

Round 2
Technical Deep Dive

1. What are your current responsibilities? Have you taken ownership of design or mentoring junior engineers?
2. Share an example of handling client escalations. How did you resolve them?
3. How do you balance technical depth vs. delivery deadlines in your projects?
4. What is the difference between map() and flatMap() transformations in Spark?
5. Explain Unity Catalog in Databricks. How would you implement it in production?
6. What are Delta Lake features like time travel and versioning? How do you get records present in one version but not in another?
7. What is Delta Live Tables? How is it different from batch pipelines?
8. SparkContext vs SparkSession — what’s the difference?

Ankita Gulati

Shubh Goyal

9. Explain Photon vs Liquid Clustering in Databricks.
10. How would you design a data lakehouse architecture on AWS or Azure?
11. When would you choose Parquet over CSV/ORC? Explain trade-offs.
12. What’s the difference between clustered and non-clustered indexes? Which scenarios require them?
13. How do you implement data lineage, auditability, and metadata management in cloud platforms?
14. Compare Snowflake, BigQuery, and Redshift in terms of performance and cost.
15. How would you detect fraudulent activity in a payments pipeline using streaming + batch data?

Round 3
HR & Behavioral

1. Why do you want to join Fractal, and what excites you about this role?
2. If you receive multiple offers, what factors will make you choose Fractal?
3. What are your career aspirations for the next 5 years?
4. Tell me about a time you resolved conflict within your team.
5. How do you motivate yourself and your team during high-pressure deadlines?
6. Which do you enjoy more — solving technical challenges or leading a team?

Thank You

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

Ankita Gulati

Shubh Goyal