



# Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



# Job Details

- **Position:** Senior Data Engineer
- **Experience:** 5–8 years
- **Location:** Pune
- **Work mode:** Hybrid
- **Compensation:** ₹28–35 LPA
- **Total Rounds:** 3
- **Top Required Skills:**
  - Advanced SQL
  - Python
  - Apache Spark
  - Databricks
  - Data Warehousing
  - Behavioral & Managerial

# Round 1

## Technical Round

### SQL Questions

1. Write a query to calculate total transactions per customer for the last 30 days.
2. Find the top 3 transactions per customer using `ROW_NUMBER()` or `DENSE_RANK()`.
3. Running sum per branch ordered by transaction date.

### Python Questions

1. Given an array, find all subarrays with sum = 0.
2. Reverse the words in a string without using built-in `split()`.
3. Optimize solution to run in  $O(n)$ .

### Spark / Project Discussion

1. How do you optimize Spark jobs?
2. Difference between `cache()` and `persist()`.
3. When would you use `repartition()` vs `coalesce()`?
4. Why choose Parquet over CSV in pipelines?

# Round 2

## Techno-Managerial Round

### SQL

1. Find the Nth highest salary without using RANK() / DENSE\_RANK().
2. Given a call log table (caller, callee, duration), compute average call duration per pair (A,B).
  - Expected: Treat (A,B) = (B,A).

### Python Questions

1. Find the first non-repeating character in a string.
2. Given n, reduce it to 0 → if even divide by 2, else subtract 1. Return steps count.

### Spark / Big Data Internals

1. Explain jobs, stages, and tasks in Spark.
2. What causes a shuffle? How to avoid it?
3. When do you use a broadcast join?
4. Explain Adaptive Query Execution (AQE) – when does it trigger?
5. Compare file formats: Parquet vs ORC vs Avro vs CSV.
6. How does Hive handle schema evolution?
7. Difference between Jobs cluster vs All-purpose cluster in Databricks.
8. How would you schedule an ETL pipeline in Databricks?

# Round 3

## HR + Behavioral Round

### General Fit

1. Tell me about yourself and your journey in data engineering.
2. Why do you want to join Mastercard?
3. What do you know about Mastercard's data ecosystem and products?
4. Why should we hire you over other candidates?

### Behavioral Scenarios

1. Tell me about a time you had a production incident at midnight. How did you handle it?
2. Describe a situation where you had conflicting priorities from two managers – what did you do?
3. Give an example of when you mentored a junior team member.
4. Tell me about a project you're most proud of and why.
5. Have you ever disagreed with your manager's decision? What happened?
6. If you join Mastercard, what will be your first 90-day goals?

*Thank You*

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

