stripe

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 4+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹50+ LPA
- **Total Rounds:** 4
- **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
Data Structures & Algorithms

1. Given a string, find the length of the longest substring without repeating characters.
2. Write a function that takes an array of integers and a target value, and returns the two numbers that add up to the target.
3. Given a string, return the longest palindromic substring.
4. Given a graph, write a function to count the number of connected components.

5. General DSA Concepts

- Be ready to work with arrays, hash tables, linked lists, trees, and graphs.
- Common patterns include sliding windows, sorting, searching, pathfinding, dynamic programming.

Ankita Gulati

Shubh Goyal

Round 2
Querying & Data Analysis

1. Write a query to find customers who placed more than three orders in the past year.

2. Write a query to rank employee salaries by department using window functions.

3. Write a query to aggregate sales data by product and month.

4. Write a query to filter outliers in transaction data using window functions.

5. Explain when to use RANK() vs. DENSE_RANK() vs. ROW_NUMBER().

Ankita Gulati

Shubh Goyal

Round 3
System Design

1. Real-Time Data Pipeline (Payments Use Case)

• Design a system to process payment transactions in real-time and update customer accounts.

• Key considerations:

• Ingestion: Kafka (events)

• Processing: Flink / Spark Streaming

• Storage: Cassandra/DynamoDB + Snowflake/BigQuery

• Monitoring: Prometheus/Grafana

2. Recommendation System for E-commerce

• Design a scalable recommendation system using real-time data.

• Components: ingestion, feature store, ML serving, storage.

Ankita Gulati

Shubh Goyal

3. General System Design Themes

• How would you design to scale for billions of records?

• How do you decide which big data tools (Spark, Kafka, Flink, Snowflake, DynamoDB) to use?

• How would you ensure low latency and fault tolerance in a real-time system?

Ankita Gulati

Shubh Goyal

Round 4
Behavioral

1. Tell me about a time you solved a challenging technical problem.
• Use the STAR (Situation, Task, Action, Result) method.

2. Give an example of a team conflict. How did you resolve it?

3. How do you ensure cross-functional alignment in data projects?

4. Why Stripe?
• How does your experience align with Stripe's products, mission, and business model?

5. How do you stay updated with the latest technologies in data engineering?

Ankita Gulati Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal