Nielsen

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer II
- **Experience:** 2+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹20+ 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
DSA and SQL

General Introduction

1. Walk me through your previous tech stack.
2. What data volumes have you worked with, and what was the business impact of your work?

Coding Problems

3. Longest Substring with Unique Characters
- Input: "aabcdeeefijklmno"
- Output: "fijklmno"
- Solve using the sliding window technique with a hash set to track characters.

4. Check if Two Strings are Anagrams (O(n), no sorting allowed)
- Input: s1 = "tan", s2 = "ant"
- Output: Yes
- Solve using a hash map (dictionary) to count character frequencies.

SQL Problem

5. From a product pricing table, write a query to find product names with strictly increasing prices across months.

- Schema: product_name, product_id, price, price_change_month
- Expected Output: product(s) with consistently rising prices.
- Approach: Use window functions (LAG, ROW_NUMBER) or self-joins to compare consecutive rows.

Ankita Gulati

Shubh Goyal

Round 2
Data Modeling and Spark

Project Discussion

1. Walk me through your previous project(s) and explain your key responsibilities.

Spark & Optimization

2. How do you handle data skewness in Spark?
3. What are some code optimization and partitioning strategies you have implemented in Spark?

Data Modeling Task

4. Design a ride-booking app like Uber/Ola using a dimensional model.

• Create a galaxy schema with fact and dimension tables for:

→ Users, Drivers, Rides, Payments, Locations, Ratings

Ankita Gulati Shubh Goyal

OOP Code Design Task

5. Implement a simplified Ride Booking System using classes and inheritance.

- Example components:
- User: requests rides.
- Driver: provides location and vehicle details.
- DriverAvailable: checks for available drivers.
- FareCalc: calculates fares.
- Status: tracks acceptance by driver and user.
- Payments: integrates fare calculation, driver location, and ride acceptance status.

Ankita Gulati

Shubh Goyal

Round 3
Situational + Hiring Manager

Role Discussion

1. Introduce yourself and describe your role at Nielsen — responsibilities, impact areas, and expectations.

Situational Design Problem

2. Amazon Prime wants to detect if a user logging in from different accounts and locations (e.g., India vs US) is the same person, to provide consistent movie recommendations.

- How would you design such a system?

Solution Aspects to Cover

- Probabilistic Matching using soft + hard identifiers:
  - → First & last name
  - → Date of birth
  - → City of birth
  - → Phone number (high confidence match)
  - → Email address (high confidence match)
  - → Time of birth, security questions (lower reliability)

- Matching Flow:
  → Step 1: Match on strong identifiers (phone/email).
  → Step 2: Apply layered filters with weaker attributes.
  → Step 3: Prompt user confirmation (device, last login, location).

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal