Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 2+ years
- **Location:** Bengaluru
- **Work mode:** Hybrid
- **Compensation:** ₹25-30 LPA
- **Total Rounds:** 5
- **Top Required Skills:**

1. SQL (Joins, Top-K, window functions)
2. DSA (arrays, strings, grids, prefix sums)
3. System Design (streaming services)
4. Python for problem solving
5. Communication & Fitment

Ankita Gulati

Shubh Goyal

Round 1
Karat Platform

SQL Questions

1. LEFT JOIN + Sorted Output
    a. Write a SQL SELECT statement with a LEFT JOIN between employees and departments.
    b. Output should list employees with their department name, sorted by department_id.
2. Top-K Records
    a. Modify the above query to return only the Top K employees with the highest salaries per department.
    b. Expected: Use ROW_NUMBER() or DENSE_RANK() inside a CTE.
    c. Follow-up: Discuss ROW_NUMBER vs DENSE_RANK in handling ties.

DSA Questions

1. Employee Entry & Exit Logs
    - Input: Logs like
    - (1, enter), (2, exit), (1, exit)
    - Task: Find employees who exited without entering and those who entered without exiting.
    - Approach: Maintain a dictionary of status; O(n) solution.
2. Logic-Only Problem
    - Candidate had to explain steps for a tougher problem:
    - Define approach clearly, discuss time & space complexity.
    - Interviewer accepted logical explanation without code.

Ankita Gulati

Shubh Goyal

Round 2
Technical

DSA Questions

1. Word Occurrence Count
   - Input: String s = "indeedindeedjobs", valid words = ["indeed","jobs"].
   - Task: Count occurrences of all valid words in s.
   - Approach:
     - Brute Force $O(n^2)$.
     - Optimized with Rabin-Karp substring search (rolling hash).

SQL Questions

1. Kth Largest Salary
   - Input: employee(emp_id, salary).
   - Task: Find the 3rd highest salary.
   - Approach: Use DENSE_RANK() in CTE.
   - Follow-up: Solve without OFFSET → Use window functions.

System Design

1. YouTube-like Video Streaming Service
   - Task: Design a system for uploading, storing, and streaming videos.
   - Key points:
     - Use object storage for videos (e.g., S3).
     - CDN for delivery.
     - Metadata DB for video info.
     - Scaling challenges: concurrent viewers, caching.

Ankita Gulati

Shubh Goyal

Round 3

Technical (Pair Programming on HackerRank)

1. Count Odd Numbers in Range
   - Input: Range [0, 5]
   - Output: 3 (numbers are 1, 3, 5).
   - Approach: Formula → (R+1)//2 - (L//2).
2. Steps to Reduce Number to Zero
   - Rule: If n is even → n/2; if odd → n-1.
   - Input: 8
   - Output: 5 steps (8→4→2→1→0).
   - Approach: Loop until zero. O(log n).
3. Sum of Even Fibonacci Terms
   - Input: n=5 → Fib = [1,2,3,5,8].
   - Output: 10 (2+8).
   - Approach: Generate Fibonacci numbers; accumulate even terms.

Ankita Gulati

Shubh Goyal

Round 4
Technical

1. Word Search in 2D Grid
    - Input: 2D grid of letters + dictionary = {"job","indeed"}
    - Task: Count frequency of words in the grid.
    - Constraint: Move in 8 directions (up, down, left, right, diagonals).
    - Approach:
    - DFS/Backtracking per cell → O(NMlen(word)).
    - Optimization: Build Prefix Tree (Trie) → Prune invalid searches early.

Ankita Gulati

Shubh Goyal

Round 5
HR & Fitment

1. Self-Introduction & Thesis Work
    a. Discuss AI/ML background and why moving to Data Engineering.
2. Why Indeed?
    a. Motivation to work in data-focused engineering vs ML.
3. Role Suitability
    a. Strengths in SQL + pipeline logic for DE role.
4. Cultural Fit
    a. Working style, collaboration, adaptability.

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal