Expedia

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** DataEngineer
- **Experience:** 2–3 Years
- **Location:** Bangalore, Gurgaon
- **Work mode:** Hybrid
- **Compensation:** ₹18–₹25 LPA
- **Total Rounds:** 3
- **Top Required Skills:**

1.SQL (deep dive)
2.DSA (BFS, arrays)
3.Data Governance
4. Indexing/Partitioning
5.BI practices,
6. Behavioral

Ankita Gulati

Shubh Goyal

Round 1
Technical Screening

DSA Questions:

1.Array Manipulation Problem

- Question: Givenanarrayofintegers, return the maximum difference betweentwoelements such that the larger element comesafterthesmaller one.
- Example:

  Input:[2,3,10,6,4,8,1]

  Output: 8 (10 - 2)
- ExpectedApproach:Keeptrack of min element seen so far, compute max diff in O(n).

2. Matrix BFSProblem

- Question:Givenabinarymatrix (0 = water, 1 = land), find the shortestdistancefromtop-left to bottom-right using BFS.
- Example:

  Input:

  1 0
  1
  1 1
  0
  0 1

SQL Problem:

- Scenario: Write a query to return the average order amount per customer after applying multiple transformations.
- Required:
  - Use multiple CTEs.
  - Do filtering in one CTE, aggregation in another, then join results.
- Focus: readability, correctness, optimization.

Round 2
SQL Deep Dive(Technical Interview)

1.Rolling MeanperProduct

- Question:Foreach product, calculate a rolling 3-day average of sales .Example Table:

product_id|sale_date | sales

A |2025-01-01 | 100
A |2025-01-02 | 200
A |2025-01-03 | 300
A |2025-01-04 | 400

2. CTE + Join + RANK()
   - Example:Findthe top-selling product per category.
   - Required:UseCTE + RANK().
   - Follow-up:Modify using LEFT JOIN to include categories with no sales.
3. Cities with MoreCustomers than Average
   - Tables:
     - country_table(id, country_name)
     - city_table(id, city_name, country_id)
     - customer_table(id, city_id)
   - Requirement: Return all cities with customer count > average customer count.

SQL Concepts Discussion:

- Indexing:Clusteredvs Non-clustered, Bitmap, B-Tree, covering indexes.
- Partitioning: Horizontal vs Vertical, pruning techniques.
- File Formats: CSV vs Parquet vs ORC (pros/cons for analytics).

Ankita Gulati Shubh Goyal

Round 3

Data Governance&Practices (Behavioral + Conceptual)

1.Expedia Value – “Force Simplicity”
   a.Share an example where you simplified a pipeline/system.
   b.Example: Replacing multiple Python scripts with a config-driven Spark job, reducing maintenance by 60%.
2. User-Centricity
   a.How do you ensure your pipeline design considers end users (analysts, product teams)?
3.Data Governance
   a.How would you implement access control in a data lake?
   b.How do you maintain data lineage?
   c.What practices ensure auditability of sensitive data (PII)?
4.Data Cleaning & BI Practices
   a. How do you handle nulls, duplicates, inconsistent formats before data reaches dashboards?
   b.Example: Created an Airflow DAG that validates data quality before loading into Snowflake.
5.Challenges & Learnings
   a.Share a challenge faced in maintaining pipelines under heavy load.
   b.Example: Schema evolution in Kafka → solved using schema registry + backward compatibility rules.

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati

Shubh Goyal