Delivery Hero

Data Engineering Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 7+ Years
- **Location:** Berlin, Germany
- **Work mode:** Hybrid
- **Compensation: €65k - €75k**
- **Total Rounds:** 4
- **Top Required Skills:**

1.AdvancedSQL
2. Python
3.System Design
4.Apache Spark
5. Airflow
6.Kafka
7.Behavioral Skills

Ankita Gulati

Shubh Goyal

Round 1
Get To Know

- Can you walk me through your career journey and past projects?
  → What was your role and contribution in each?
  → Which project are you most proud of, and why?
- What challenges did you face in your data engineering work?
  → Can you share one example of a technical roadblock and how you solved it?
- Why did you apply to Delivery Hero?
  → What excites you about our business and data platform?
- How do you continue learning as a data engineer?
  → Do you attend conferences, workshops, certifications, or rely on blogs/online communities?
  → How do you apply what you learn into your projects?

Ankita Gulati

Shubh Goyal

Round 2
Hiring Manager Interview

Team & Collaboration

1. What does an ideal team look like to you?
2. What kind of team environment do you struggle in?
3. If you disagree with feedback from your lead/peer, how do you handle it?
4. Share a challenge you faced with a cross-functional team — how did you resolve it?
5. Tell me about a mistake you made in a project.
    - How did you detect it?
    - What steps did you take to fix it?

Decision-Making & Impact

1. Howdoyou compare timevs. value when working on tasks?
    - Example: Choosing between quick fixes vs long-term scalable solutions.
2. Describe a project you successfully accomplished.
    - What were the key success factors?

Ankita Gulati

Shubh Goyal

Technical Add-ons

1. Whatmetrics do you use to evaluate if a Spark job is running efficiently?
    - Task completion time? Shuffle read/write? Executor memory?
2. How do you handle schema evolution when new files with extra/missing columns arrive?
    - Do you use schema-on-read, versioned schemas, or enforce schema validation?

Round 3

Technical (Coding+System Design)

Python

1.Theatre Visibility Problem:

- Write Python code to check if everyone in a theatre can see thescreen.
- Input isa2D matrix where each number represents a person's height.
- A personcan see the screen if their height is greater thantheperson in front.
- Example (valid):
  [[1,2, 3, 2, 1, 1],
  [2,4, 4, 3, 2, 2],
  [5,5, 5, 5, 4, 4],
  [6,6, 7, 6, 5, 5]]
- Modifycode for failing case.
- Discusstime complexity (O(n*m)) and space complexity.

SQL

1.Write a query to find the second highest salary per department.

- If a department has less than two employees, return NULL.
- Show solutions using RANK() and also LIMIT.

Ankita Gulati

Shubh Goyal

System Design

1. Design a cost-efficient, scalable data pipeline for an e-commerce website.
    - How would you ingest user clickstream data? (Kafka, Kinesis)
    - How would you process it? (Batch vs Streaming)
    - Which storage would you choose? (Data Lake, Delta Lake, Data Warehouse)
    - How do you ensure fault tolerance, partitioning, and monitoring?

Spark

1. Explain narrow vs wide transformations in Spark with examples.
2. What causes data skew in Spark jobs?
    - How would you mitigate it (salting, repartitioning, skew join optimization)?
3. Which join strategies optimize performance in Spark?
    - Broadcast join vs Sort-merge join.

Airflow

1. Explain Airflow fundamentals.
    - What is a DAG?
    - What are operators?
    - How do XComs work for data sharing?

Ankita Gulati Shubh Goyal

Kafka

1. Explain Kafka basics.
    - How are messages written to topics/partitions?
2. What are consumer groups in Kafka?
    - How do they ensure parallelism?
3. How does offset management work?

- Difference between automatic and manual offset commits.

Ankita Gulati

Shubh Goyal

Round 4
Technical (Bar Raiser)

- Walk me through one of your past projects in detail.

→ What was the scale of data processed?

→ What optimizations did you implement?

- Write a program to calculate the frequency of each character in a string.

Example input: "asrfrfgg"

Expected output: "a1s1r2f2g2"

- Discuss time complexity (O(n)) and space complexity (O(1) if ASCII, O(k) if Unicode).

Thank You

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

Ankita Gulati

Shubh Goyal