

Delivery Hero Data Engineer Interview Guide – Experienced 7+

Round 1: Get to Know

1. Can you tell me about your past experiences and projects?

Answer:

Focus on impactful projects, highlighting tools, technologies, and quantifiable outcomes.

Example: *I worked on optimizing a real-time data pipeline using Apache Kafka and Spark, which reduced data processing time by 40%. Additionally, I led a cloud migration project from on-premises Oracle databases to Google Cloud, leveraging BigQuery and Dataproc.*

2. What challenges did you face, and how did you tackle them?

Answer:

Use the STAR framework to structure your response.

Example: *In one project, we faced data skew issues with Spark jobs. I identified the issue using Spark UI and implemented salting techniques, which balanced the load across partitions and reduced job failures.*

3. Why did you apply to Delivery Hero?

Answer:

Align your motivations with their mission and the role.

Example: *I admire Delivery Hero's commitment to innovation in the food delivery space. The opportunity to work on scalable, real-time systems aligns with my expertise and passion for solving challenging data problems.*

4. How do you keep up with learning? Have you attended any conferences or engaged in other learning activities?

Answer:

Mention certifications, platforms, or events.

Example: *I recently completed the AWS Certified Data Analytics – Specialty certification and attended the Data + AI Summit 2024. I also regularly follow industry blogs and contribute to open-source projects.*

Round 2: Hiring Manager

5. What does an ideal team look like to you?

Answer:

An ideal team promotes collaboration, innovation, and knowledge sharing while maintaining clear goals and accountability.

6. What kind of team would you prefer not to work with?

Answer:

I prefer not to work in a team with poor communication or unclear priorities, as it can hinder productivity and growth.

7. How do you handle situations where you disagree with feedback from others?

Answer:

I first seek to understand their perspective, provide data-driven insights to support my view, and collaborate to find the best solution.

8. Can you describe a challenge you faced and how you resolved it?

Answer:

During a migration project, I faced schema mismatches between source and target systems. I resolved it by creating a mapping strategy and automating schema validation using Python scripts.

9. What is a mistake you made, and how did you overcome or resolve it?

Answer:

In a data pipeline project, I underestimated resource requirements, causing delays. I resolved it by introducing dynamic resource allocation and ensuring better workload testing in subsequent projects.

10. How do you compare the time investment and value of a task?

Answer:

I use the effort-impact matrix to prioritize tasks, focusing on high-impact and low-effort items first.

11. Can you describe a project you successfully accomplished? What did you do to achieve that success?

Answer:

In a recent project, I built a recommendation engine using Spark and MLlib, which improved product recommendations by 30%. The success was due to iterative improvements and effective collaboration with the team.

12. What metrics do you use to determine whether a Spark job is going well or not?

Answer:

Metrics like DAG stage duration, shuffle read/write size, executor memory usage, and task failure rates are crucial for monitoring Spark job performance.

13. How would you handle a schema change when new files arrive?

Answer:

I use schema evolution with tools like Avro or Parquet and validate new schemas using automated checks. For breaking changes, I version schemas and update downstream dependencies accordingly.

Round 3: Technical - Coding and System Design

14. Write a Python code that determines if all the people in their seats can see the screen in the theatre.

Answer:

```
def can_see_screen(seats):
    for col in range(len(seats[0])):
        for row in range(1, len(seats)):
            if seats[row][col] <= seats[row - 1][col]:
                return False
    return True
```

15. Write a SQL query to find the second highest salary in each department.

Answer:

```
SELECT department_id,
       MAX(salary) AS second_highest_salary
FROM (
    SELECT department_id, salary,
           DENSE_RANK() OVER (PARTITION BY department_id ORDER BY salary DESC) AS rank
    FROM employees
) ranked_salaries
WHERE rank = 2
GROUP BY department_id;
```

16. How would you design a cost-effective, scalable, and efficient data pipeline for an e-commerce website?

Answer:

Use Kafka for ingestion, Spark for transformations, and Snowflake for storage. Optimize cost with spot instances and monitor using Prometheus/Grafana.

17. Apache Spark Questions

- **What is data skewness?**

Uneven distribution of data across partitions causing performance issues. Resolved with salting or partition pruning.

- **Narrow vs Wide transformations**

Narrow transformations operate within a single partition (e.g., map), while wide require shuffles across partitions (e.g., reduceByKey).

18. Airflow Questions

- **What are XComs?**

XComs allow task communication by passing small amounts of data between tasks in a DAG.

19. Kafka Questions

- **What is offset management?**

Offset management tracks consumer progress in reading messages from topics, ensuring fault tolerance and no duplicate processing.

Round 4: Technical - Bar Raiser

20. How would you implement a program to determine the frequency of each letter in a string?

Answer:

```
from collections import Counter

def letter_frequency(s):
    freq = Counter(s)
    return ''.join(f"{char}{freq[char]}" for char in sorted(freq))

print(letter_frequency("asrfrfgg")) # Output: "a1f2g2r2s1"
```

21. Can you tell me about your past experiences and projects?

Glassdoor Delivery Hero Review –

<https://www.glassdoor.co.in/Reviews/Delivery-Hero-Reviews-E504556.htm?countryRedirect=true>

Delivery Hero Careers –

<https://careers.deliveryhero.com/deliveryhero-jobs>

Subscribe to my YouTube Channel for Free Data Engineering Content –

<https://www.youtube.com/@shubhamwadekar27>

Connect with me here –

<https://bento.me/shubhamwadekar>

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar

© Shubham Wadekar