Swiggy Data Engineer Interview Guide – Experienced 2+

Swiggy, a leader in the food delivery industry, offers dynamic challenges for data engineers along with a competitive package. Here's an in-depth roadmap to help you prepare and ace the interview process:

1. Master Data Engineering Fundamental

Key Topics: ETL vs. ELT, pipeline scalability, batch vs. real-time processing

Sample Questions:

- What's the difference between ETL and ELT, and when would you choose one over the other?
- How do you ensure the scalability of a data pipeline handling rapidly growing data volumes
- Explain the trade-offs between batch and real-time data processing. Provide examples of when each is appropriate.
- Describe the role of a workflow orchestrator like Airflow in a data pipeline.
- How would you handle schema changes in a production ETL pipeline?

Tools to Focus On: Airflow, dbt, streaming frameworks, data pipeline design

2. Sharpen Your SQL Skills

Key Topics: Complex queries, window functions, large dataset optimization

Sample Questions:

- Write a SQL query to find the top 5 most ordered dishes in the last 30 days.
- Calculate a 7-day moving average of orders for each city in the Swiggy database.
- Write a query to identify duplicate customer entries based on email and phone number.
- Optimize a slow SQL query for a large orders table containing billions of rows.
- Explain the use of indexing and partitioning to improve query performance.
- How do you handle NULL values in a SQL query to avoid incorrect results?

Focus Areas: Window functions, common table expressions (CTEs), indexing, query optimization

3. Dive into Data Modeling

Key Topics: Schema design, dimensional modeling, data warehouse optimization

Sample Questions:

- Compare the star schema and snowflake schema. Which one would you use for reporting at Swiggy, and why?
- Design a data warehouse schema to track orders, customers, delivery partners, and payments.
- Explain the concept of surrogate keys and natural keys. Which would you prefer in a data warehouse?
- What are Slowly Changing Dimensions (SCD), and how would you implement them for tracking customer data changes?

Tools to Study: Dimensional modeling, data warehouse systems like BigQuery, Redshift, and Snowflake

4. Explore Big Data Tools

Key Topics: Spark, Kafka, HDFS, distributed processing

Sample Questions:

- Explain how Kafka handles real-time data streaming and guarantees message delivery.
- Describe the stages of a Spark job and strategies to optimize Spark performance for large datasets.
- How do you ensure data consistency and fault tolerance in a distributed data pipeline?
- Compare HDFS and cloud-based storage systems in terms of scalability and performance.
- How would you handle late-arriving data in a real-time stream processing pipeline?

Key Tools: Apache Kafka, Spark, Hadoop, cloud storage systems (AWS S3, GCS, Azure Blob Storage)

5. Python for Data Engineering

Key Topics: Data transformations, scripting, automation

Sample Questions:

- Write a Python function to deduplicate a list while preserving the order of elements.
- How do you clean missing values in a pandas DataFrame?
- Implement a function to parse and analyze log files for error rates.
- Write a script to automate daily ingestion of data from an API into a data lake.
- Describe how you would use PySpark to aggregate and summarize large transaction datasets.

Libraries to Focus On: pandas, PySpark, NumPy, built-in Python functions

6. System Design

Key Topics: Scalable architectures, fault tolerance, data partitioning

Sample Questions:

- Design a system to handle 1M daily transactions with real-time analytics for Swiggy.
- How would you ensure fault tolerance and data consistency in a distributed data pipeline?
- What are the trade-offs between consistency, availability, and partition tolerance (CAP theorem)?
- How do you handle schema evolution in a system with multiple data sources and consumers?
- Design a logging and monitoring solution for a mission-critical data pipeline.

Important Concepts: CAP theorem, scalable architectures, schema evolution, monitoring and alerting

7. Scenario-Based Problem Solving

Key Topics: Debugging, performance bottlenecks, stakeholder collaboration

Example Scenario: Swiggy faces delivery delays in a region—how would you identify and fix pipeline bottlenecks?

- Walk through how you would debug the data ingestion process to identify slow stages.
- How would you use monitoring tools to detect and resolve pipeline failures proactively?
- Describe a situation where you prioritized business needs over technical elegance. How did you manage trade-offs?

Additional Key Questions to Prepare:

- How do you ensure data quality in an automated pipeline?
- What techniques do you use to balance compute costs and performance in cloud-based data solutions?
- Explain how you would design a pipeline for streaming real-time order status updates.
- How do you secure sensitive customer data in a data warehouse?
- Discuss trade-offs between serverless and traditional cloud data architectures.
- Provide strategies for handling data deduplication and cleaning in Spark jobs.
- What are the best practices for CI/CD implementation in data engineering workflows?

Pro Tips for Cracking Swiggy Data Engineering Interviews:

- Understand Real-World Use Cases: Prepare to connect technical solutions with business outcomes, especially in e-commerce and food delivery.
- Brush Up on Core Tools: Master Spark, Kafka, Airflow, dbt, and cloud platforms (AWS, GCP, Azure).
- Stay Ready for Trade-Off Questions: Practice articulating trade-offs between performance, cost, and complexity.
- Communicate Clearly: Be ready to explain technical designs and decisions to non-technical stakeholders.

With these insights and a solid study plan, you'll be better equipped to secure your dream role as a Data Engineer at Swiggy!

Glassdoor Swiggy Review –

https://www.glassdoor.co.in/Reviews/Swiggy-Reviews-E952680.htm

Swiggy Careers –

https://careers.swiggy.com/#/

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