



Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



Job Details

- **Position:** Data Engineer II
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹25+ LPA
- **Total Rounds:** 4
- **Top Required Skills:**
 1. SQL
 2. PySpark / Python
 3. Cloud Data Engineering
 4. ETL / Data Modeling
 5. Big Data & Streaming
 6. System Design

Round 1

SQL & Data Modeling

1. Duplicate Detection

- Write a SQL query to identify duplicate records in a table based on one or more columns.
- Output should list the duplicate records along with the count of occurrences.

2. Running Totals

- Given a sales table, write a SQL query to calculate the running total of sales per month for the current year.

3. Data Modeling Scenario (E-commerce)

- Design a dimensional model for an e-commerce business with focus on products, sales, and customers.
- How would you handle Slowly Changing Dimensions (SCDs) in this model?

Round 2

Azure & ETL Design

1. Azure Data Factory (ADF)

- Explain how you would build an incremental ETL pipeline in ADF to load data from an on-premises SQL database into Azure Synapse Analytics.

2. Azure Functions

- Describe how Azure Functions can be used in a data pipeline for event-driven workflows.

3. Error Handling in ETL Pipelines

- What mechanisms would you implement to handle failures in a multi-stage ETL pipeline?
 - (e.g., retries, dead-letter queues, alerts, checkpointing).

Round 3

System Design & Architecture

1. Real-Time Ingestion Case Study

- Design a pipeline to process millions of log entries per second in real-time.
- Which tools and architecture would you recommend (e.g., Kafka, Event Hubs, Spark Streaming, Databricks, Synapse)?

2. Data Partitioning Strategy

- For a 100TB dataset, how would you design a partitioning strategy to ensure efficient querying and minimal latency?

3. Schema Trade-Offs

- Compare and discuss the trade-offs between star schema and snowflake schema in a data warehouse.
- Which one would you choose in different scenarios and why?

Round 4

Behavioral & Leadership

1. Ownership Under Pressure

- Describe a situation where you had to resolve a critical data pipeline failure under tight deadlines.
- How did you diagnose, fix, and prevent it from happening again?

2. Cross-Functional Collaboration

- Share an example of how you collaborated with cross-functional teams (e.g., product managers, analysts, engineers) to deliver a large-scale data project.
- What challenges did you face and how did you overcome them?

Thank You

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

