

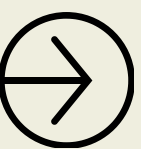
**adidas**

# Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



# Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Gurgaon
- **Work mode:** Hybrid
- **Compensation:** ₹15–20LPA
- **Total Rounds:** 3
- **Top Required Skills:**
  - 1.SQL & Database Design
  - 2.Python & ETL Development
  - 3.Big Data & Cloud Technologies
  - 4.System Design for Scalability
  - 5.Behavioral & Communication Skills

# Round 1

## Technical Assessment

### SQL & Database Design

1. Write an SQL query to find the top five most-sold Adidas products in the last month. Be sure to handle ties correctly.
2. Design a database schema to store customer transactions. Your schema should include attributes such as region, product category, product details, transaction timestamp, and transaction amount.
3. How would you optimize a query that fetches sales data across multiple countries where the dataset contains billions of rows? Discuss indexes, partitioning, and query rewriting.
4. Explain the differences between OLTP and OLAP databases. Which is better suited for Adidas's analytical workloads, and why?

### Python & ETL Pipelines

5. Write a Python script that processes raw JSON files containing sales data and loads the results into a relational database. Describe the libraries you would use.
6. How would you approach debugging a failing ETL pipeline in production? Give examples of checks and monitoring strategies you would use.
7. How would you identify and handle duplicate or corrupted data records during a batch ETL job?
8. Write a Python function that detects anomalies in sales trends using Pandas and NumPy. Explain how you would define an "anomaly."

# Round 2

## System Design & Big Data

### Big Data & Cloud Computing

1. Compare Apache Spark and AWS Glue for large-scale data processing. Which would you choose for Adidas's workloads and why?
2. How would you design a scalable data lake for Adidas's global e-commerce operations? Include details about data partitioning, schema evolution, and security.
3. Explain how you would implement real-time analytics using a streaming platform such as Kafka or Amazon Kinesis.
4. What are the benefits of using a cloud data warehouse like Amazon Redshift or Snowflake for Adidas's analytics?

### System Design

5. Design a data pipeline that collects, processes, and visualizes customer feedback from Adidas stores worldwide. Which tools and storage formats would you use?
6. How would you architect a recommendation system for Adidas's e-commerce platform? Describe the data sources, feature engineering, and model-serving layer.
7. Propose a solution for monitoring and maintaining data quality across multiple geographic regions and data sources.
8. Describe how you would design a system to track inventory and sales in real time across Adidas's global operations.

# Round 3

## Behavioral & Leadership

1. Share an example of when you had to explain a complex technical issue to a non-technical stakeholder. How did you ensure they understood?
2. Describe a project where you optimized an existing process or pipeline. What approach did you take, and what measurable impact did it have?
3. Tell us how you manage competing priorities or tight deadlines in a collaborative team environment.
4. What is your approach to continuous learning, especially in keeping up with evolving data engineering technologies?

*Thank You*

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

