



Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



Job Details

- **Position:** Data Engineer
- **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

Technical Interview 1

SQL Questions

1. Identify the top 5 products with the highest increase in weekly sales over the last quarter.
2. Calculate the percentage of out-of-stock items for each store and list stores that exceed 20% out-of-stock on average.
3. Find products that were consistently sold in every store across a region, but had zero sales in at least one store last month.
4. Write a query to track customer retention by finding how many customers purchased every month for the last 6 months.
5. Explain how indexing works in SQL. How would you use it to optimize queries involving multiple joins on a large sales dataset?
6. How would you manage and query a database containing billions of rows of sales data across multiple time zones?

6. How would you manage and query a database containing billions of rows of sales data across multiple time zones?

7. Write an SQL query to measure the effectiveness of seasonal promotions on specific product categories.

8. Explain the difference between OLTP and OLAP databases, with examples of how Target might use each.

Python Questions

9. Write a function to calculate the total sales of a product given a list of daily sales data.

10. Write a script to read a dictionary of product IDs and stock counts and flag products with stock below a threshold.

11. Write a Python function to generate a weekly sales report from a CSV file
(date, product_id, sales_amount), aggregating total sales per product.

12. Explain how Python's pandas library can handle missing data in large sales datasets. Provide an example of filling missing values using the average sales of the same weekday.

13. Write a function that takes two lists (store IDs and revenues) and returns a dictionary mapping store IDs to revenues.

14. What are the key differences between a list and dictionary in Python? How would you use each for retail data analysis?

DSA Question

15. Solve a basic data structures/algorithms problem (e.g., array manipulation, string operations, or hash-based lookups).

PySpark/Spark

16. PySpark coding exercise — perform data transformations and aggregations on a given dataset.

17. Explain core Spark concepts: RDDs, DataFrames, lazy evaluation, actions vs transformations.

Round 2

Technical Interview II

1. How do you handle data skew in Spark jobs?
2. Explain different partitioning strategies in Spark and when to use them.
3. How would you implement incremental data loading in PySpark?
4. What are some strategies for optimizing Spark jobs running on Azure?
5. **Scenario:** You need to process real-time sales data on Azure. Which services (Data Factory, Databricks, Synapse, Event Hubs) would you use and why?
6. Write an SQL query for a business problem (scenario-based). Explain how you would tune it for performance.

Round 3

Hiring Manager

1. Walk me through your current project in detail.
 - What were the main challenges and how did you solve them?
 - Which tools and frameworks did you use?
2. Explain your experience with Kafka:
 - How does Kafka ensure fault tolerance and durability?
 - How would you design a Kafka consumer group for real-time transaction processing?
 - How do you handle exactly-once processing semantics?

Round 4

HR Interview

1. Why do you want to join Target?
2. What are your career growth goals over the next 3–5 years?
3. How do you handle team conflicts and disagreements?
4. Tell me about a time you worked under tight deadlines.
5. Discuss your leadership or team management experience.

Thank You

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

