



# Data Engineering Interview Questions



Ankita Gulati

Shubh Goyal



# Job Details

- **Position:** Data Engineer II
- **Experience:** 3+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹30+ 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

Ankita Gulati

Shubh Goyal

# Round 1

## Technical: DSA & SQL

### DSA Questions

1. Solve the Rainwater Trapping Problem, optimizing for both time and space complexity.
2. Solve a Priority Queue problem related to task prioritization (details varied, but required strong grasp of heaps/queues).

### SQL Questions

3. Explain the use cases of window functions in SQL.
4. Write and optimize queries using window functions for ranking, aggregations, and trend analysis.
5. What strategies do you use to optimize SQL queries on large datasets?

# Round 2

# Design & Spark Coding

## Project Discussion

1. Walk me through your previous projects – focus on:
  - Best practices in software and data engineering.
  - Why you chose specific tools/services.
  - How you solved challenges in real-world deployments.

## Design Question

2. Scenario: Design a solution to migrate data from Hadoop, S3, and Oracle DB into a consolidated S3 bucket.
  - Which services/tools would you select?
  - How would you implement error logging, scalability, and fault tolerance?

## **Spark Coding Challenge**

3. Given two DataFrames, write PySpark code to perform transformations and store the result in another DataFrame.

- Demonstrate proficiency in PySpark APIs (joins, aggregations, column manipulations).
- Optimize for performance and readability.

# Round 3

## Managerial

### Projects & Challenges

1. Deep-dive into your past projects:
  - Why did you choose certain tools/services?
  - What trade-offs did you consider?
  - How did you ensure scalability and reliability?

### Real-Life Scenario Handling

2. Describe how you handled pipeline overload situations (e.g., sudden spikes in data volume).
3. Give an example of resolving service downtimes or pipeline failures.
4. Explain how you build resiliency and monitoring into data systems.

### Behavioral Questions

5. Share examples of teamwork and collaboration in delivering a large-scale project.

6. Describe a time when you demonstrated adaptability under changing priorities.
7. How do you approach problem-solving when faced with limited information?



Ankita Gulati

Shubh Goyal

# Round 4

## HR Discussion

1. Discussion of offer details and PayPal's compensation structure.
2. Standard HR behavioral questions around:
  - Fit with PayPal's company culture.
  - Alignment with PayPal's values and expectations.
3. Career motivation: Why PayPal, and how do you see yourself contributing to its data-driven initiatives?

Ankita Gulati

Shubh Goyal

# Round 5

## System Design

1. Design a data platform to ingest, process, and analyze telemetry data from millions of Nutanix clusters worldwide.
  - What ingestion strategies would you use (batch vs streaming)?
  - Which storage solutions would you recommend (HDFS, S3, NoSQL, data lake)?
  - How would you structure real-time vs offline analytics?
  - How would you integrate machine learning for predictive maintenance?

Thank You

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



Ankita Gulati

Shubh Goyal