J.P.Morgan
Data Engineering
Interview
Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Senior Data Engineer
- **Experience:** 6–9 Years
- **Location:** Bengaluru
- **Work mode:** Hybrid
- **Compensation:** ₹38-50 LPA
- **Total Rounds:** 5
- **Top Required Skills:**
1. Java + Concurrency
2. Big Data Tools
3. SQL
4. Python
5. AWS Cloud
6. System Design
7. Behavioral

Ankita Gulati   Shubh Goyal

Round 1
Preliminary Screening Call

1. Walk me through your work experience with cloud platforms, data pipelines, databases, and programming languages.
2. Why are you interested in JP Morgan? What are your career aspirations?
3. Java Fundamentals:
4. Which version introduced lambda expressions?
5. Differences between primitive data types and wrapper classes.
6. Have you worked with Infrastructure as Code tools like Terraform? Explain your use cases.
7. How have you used CI/CD pipelines in your projects? Example with Jenkins.

Ankita Gulati   Shubh Goyal

Round 2
Java & System Design

Terraform, Jenkins & Deployments
1. Write Terraform configuration for provisioning an EC2 machine.
2. Explain the Terraform lifecycle for deploying a cluster on AWS. Compare with CloudFormation stacks.
3. Walk through how you deploy code to production using Jenkins pipelines.

Java Concepts
1. Default value for:
   - float primitive type
   - Float wrapper class
2. Write a Singleton class implementation in Java.
3. How is HashMap implemented internally?
4. What happens if we don't override the Thread class run() method?
5. Difference between stubs and skeletons in RMI (Remote Method Invocation).
6. Write example code using Java Concurrency API (forEach(), forEachEntry(), forEachKey()).

Ankita Gulati   **Shubh Goyal**

Big Data — Hadoop, Spark, Databricks

1. Explain use cases of Bloom Filters in Spark.
2. How do you optimize Spark jobs?
3. What is data locality in Hadoop and why is it important?
4. Explain Databricks Job Clusters and Photon in SQL Warehouse.
5. How do you calculate Databricks cluster cost using DBUs?
6. How would you handle custom data types in Spark that are unsupported natively?
7. How do you control the number of mappers in MapReduce?
8. SQL + Spark problem:
9. Two DataFrames:
   a. df1(id, name)
   b. df2(id, country, address, city, count)
10. Perform join → filter where country = 'Singapore' → pivot → order cities by highest count

Round 3
SQL, Python, AWS, Data Structures

SQL Challenges

1. Find the consecutive number appearing at least 3 times without interruption.
2. Employee hierarchy table:

    employee_id | employee_name | manager_id
    - Write a query to find employees and their reporting managers (recursive CTE).

Python & Data Structures

1. Write a program to replace all characters in a list except a given character.
2. Given two sets of strings, find pairs that form a complete set of 26 alphabets on concatenation.
3. Read data from 3 files into Pandas DataFrame. Perform:
    → Column removal
    → Row filtering by index
    → String search
4. Convert a Binary Search Tree into a skewed tree (increasing or decreasing).

Ankita Gulati   **Shubh Goyal**

AWS & General CS

1. What AWS services have you worked on?
   Examples: EC2, S3, EMR, Glue, Redshift.
2. How do you manage file handling in Python?
3. Explain Decorators and __init__ usage in Python.

Ankita Gulati   Shubh Goyal

Round 4
Techno-Managerial Interview

1. Difference between Spark Streaming vs Structured Streaming (with real-world example).
2. How would you design an ETL pipeline for:
   - Kafka event data
   - RDBMS data (CDC with Kafka Connect).
3. Does Delta Lake support upserts?
4. Optimize AWS costs in a scenario of unused S3 data and high on-demand instances.
   - S3 tiering
   - Spot vs On-demand instances
   - IAM restrictions

Scenario Questions

1. If a business user requests a non-value-add column that slows reporting, what would you do?
2. You're assigned 3 sprint tasks, but a critical urgent task comes in — how do you handle workload?

Ankita Gulati   Shubh Goyal

Round 5
Executive Director Interview

1. Walk me through your projects with Databricks, Spark, and AWS.
2. Explain your day-to-day responsibilities with SQL, Python, and Databricks.
3. How does your team use Kubernetes clusters?
4. Steps to migrate on-prem ETL jobs to AWS.
5. Compare Spark vs Hadoop.
6. Scenario: How do you manage conflict resolution in teams?
7. Are you open to relocation for this hybrid role?

Ankita Gulati   Shubh Goyal

Thank You

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

HIRED

Ankita Gulati   Shubh Goyal