Nielsen

Data Engineering
Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer II
- **Experience:** 2+ years
- **Location:** Bangalore
- **Work mode:** Office
- **Compensation:** ₹25+ LPA
- **Total Rounds:** 3
- **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

Cloud & Infrastructure
1. Explain the concept of Infrastructure as Code in AWS.
2. How do you create and deploy an AWS Lambda function using the CLI? What are the available Lambda runtimes?

Data Concepts
3. What is data smoothing? Provide an example of when it is useful.

Python Programming
4. What is the difference between is vs. == in Python?
5. How would you remove duplicate elements from a list in Python?
6. How are the startswith() and endswith() methods used in Python? Provide examples.
7. Write a Python program to count the frequency of each character in a string, treating uppercase and lowercase characters separately.

Ankita Gulati   Shubh Goyal

Docker & Containerization

8. What is Docker and why is it useful in data engineering pipelines?

Apache Spark & PySpark

9. Explain Spark Context and Streaming Context.

10. What are RDDs (Resilient Distributed Datasets) in Spark?

11. Explain the difference between Map and Reduce operations.

12. What are the different types of joins in PySpark?

13. What is the role of the DAG Scheduler in Spark?

Ankita Gulati   Shubh Goyal

Round 2
Techno-Managerial

AWS EMR & Big Data Services

 1. Have you used Amazon EMR in your work (or portfolio projects)?

 2. What are Bootstrap actions in EMR, and how are they used?

 3. Walk through the steps to create an Amazon EMR cluster.

 4. How can you add arguments to Bootstrap actions in EMR?

 5. How do you make an EMR cluster resilient for large-scale data processing? Explain the use of EMR Fleets.

AWS Services & Security

 6. Explain the purpose of AWS KMS (Key Management Service).

 7. What is the general use of an Elastic IP in AWS?

 8. How would you restrict access to an S3 bucket so that only a specific EC2 instance can access it?

Ankita Gulati   Shubh Goyal

Scenario-Based Cloud Design

9. Design an end-to-end AWS data pipeline using EC2, S3, Glue, EMR, Lambda, Redshift, and Athena.
   • Explain the flow of data from ingestion to analytics.

Portfolio Review

10. Walk me through your LinkedIn/GitHub portfolio and certifications. Which projects best demonstrate your cloud and big data expertise?

Ankita Gulati   Shubh Goyal

Round 3
Hiring Manager

1. Discussion on team culture and collaboration at Nielsen.
2. Review of leave policies and holidays.
3. Salary negotiation — expectations vs. company structure.
4. Discussion on variable compensation components.

Ankita Gulati   Shubh Goyal

- Matching Flow:
  → Step 1: Match on strong identifiers (phone/email).
  → Step 2: Apply layered filters with weaker attributes.
  → Step 3: Prompt user confirmation (device, last login, location).

Ankita Gulati

Shubh Goyal

Thank You

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

Ankita Gulati   Shubh Goyal