Data Engineering
Interview Questions

Ankita Gulati

Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 3+ years
- **Location:** Pan India
- **Work mode:** Hybrid
- **Compensation:** ₹15-20 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 Evaluation

1. Can you explain your recent projects in detail, highlighting the data pipelines, tools, and technologies used?

→ Follow-up: How did you handle performance bottlenecks or scalability challenges?

2. Write a Python function to rotate an array by K positions.

3. Write a Python function to find the length of the longest substring without repeating characters.

4. Explain different data manipulation and cleaning techniques in Pandas. Provide examples.

5. Which data structure is used under the hood of a Python dictionary? Why is it efficient?

6. What is the importance of Data Engineering in the current data ecosystem? How does it differ from Data Science?

7. SQL Question: Write a query to display the second highest salary in each department.

Ankita Gulati   Shubh Goyal

8. SQL Question: Given an orders table with columns (order_id, customer_id, order_date, amount), write a query to:

→ Find the top 3 customers by total spending.

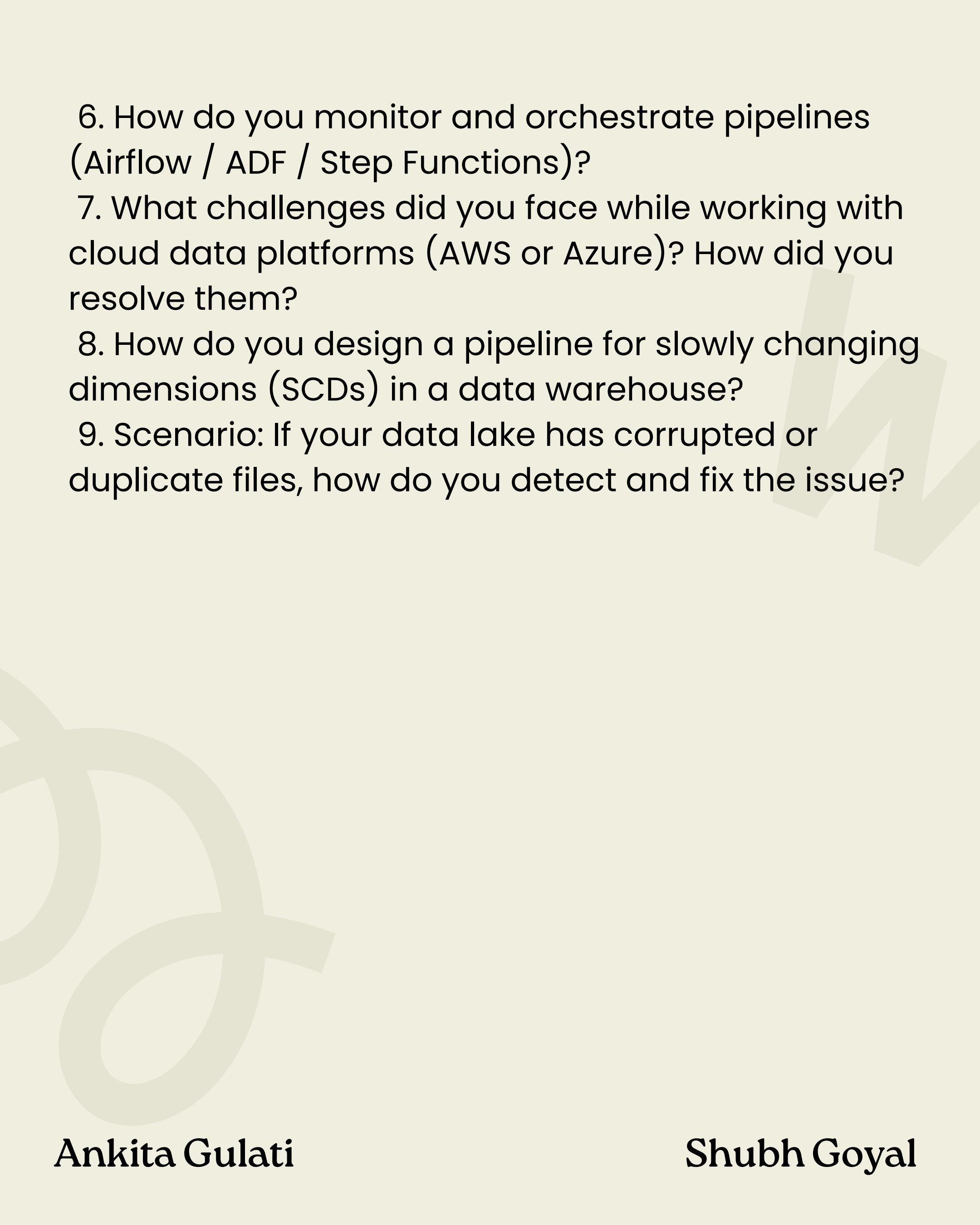→ Get the running total of amount per customer ordered by date.

9. What are the key differences between RDBMS and NoSQL databases? Give examples of when to use each.

10. Explain data modeling concepts (star schema vs snowflake schema). Which one have you used in your projects?

Ankita Gulati   Shubh Goyal

Round 2
Managerial & System Design

1. Walk me through the end-to-end architecture of one of your complex projects.

    → How did you design the data ingestion, transformation, and storage layers?

2. How do you ensure data quality, governance, and lineage in your pipelines?

3. Scenario: You are working on a real-time streaming pipeline. How would you design it in AWS or Azure?

    → Which services would you use (e.g., Kinesis/Kafka, Glue, Redshift, Azure Data Factory, Synapse)?

4. Explain how you would handle a sudden spike in data volume that impacts pipeline performance.

5. In Spark, how do you handle:

    → Data skew

    → Shuffle operations

    → Partitioning and caching

Ankita Gulati   **Shubh Goyal**

6. How do you monitor and orchestrate pipelines (Airflow / ADF / Step Functions)?

7. What challenges did you face while working with cloud data platforms (AWS or Azure)? How did you resolve them?

8. How do you design a pipeline for slowly changing dimensions (SCDs) in a data warehouse?

9. Scenario: If your data lake has corrupted or duplicate files, how do you detect and fix the issue?

Ankita Gulati

Shubh Goyal

Round 3
HR & Behavioral

1. Walk me through your resume and highlight key achievements.

2. Why do you want to join Accenture?

3. What is your long-term career goal in Data Engineering?

4. Discussion on salary expectations and variable components.

5. Availability to join and preferred location.

Ankita Gulati   Shubh Goyal

Thank You

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

HIRED

Ankita Gulati

Shubh Goyal