accenture

Data Engineering
Interview
Questions

Ankita Gulati   Shubh Goyal

Job Details

- **Position:** Data Engineer
- **Experience:** 4+ years
- **Location:** Bangalore
- **Work mode:** Hybrid
- **Compensation:** ₹20-25 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

Ankita Gulati

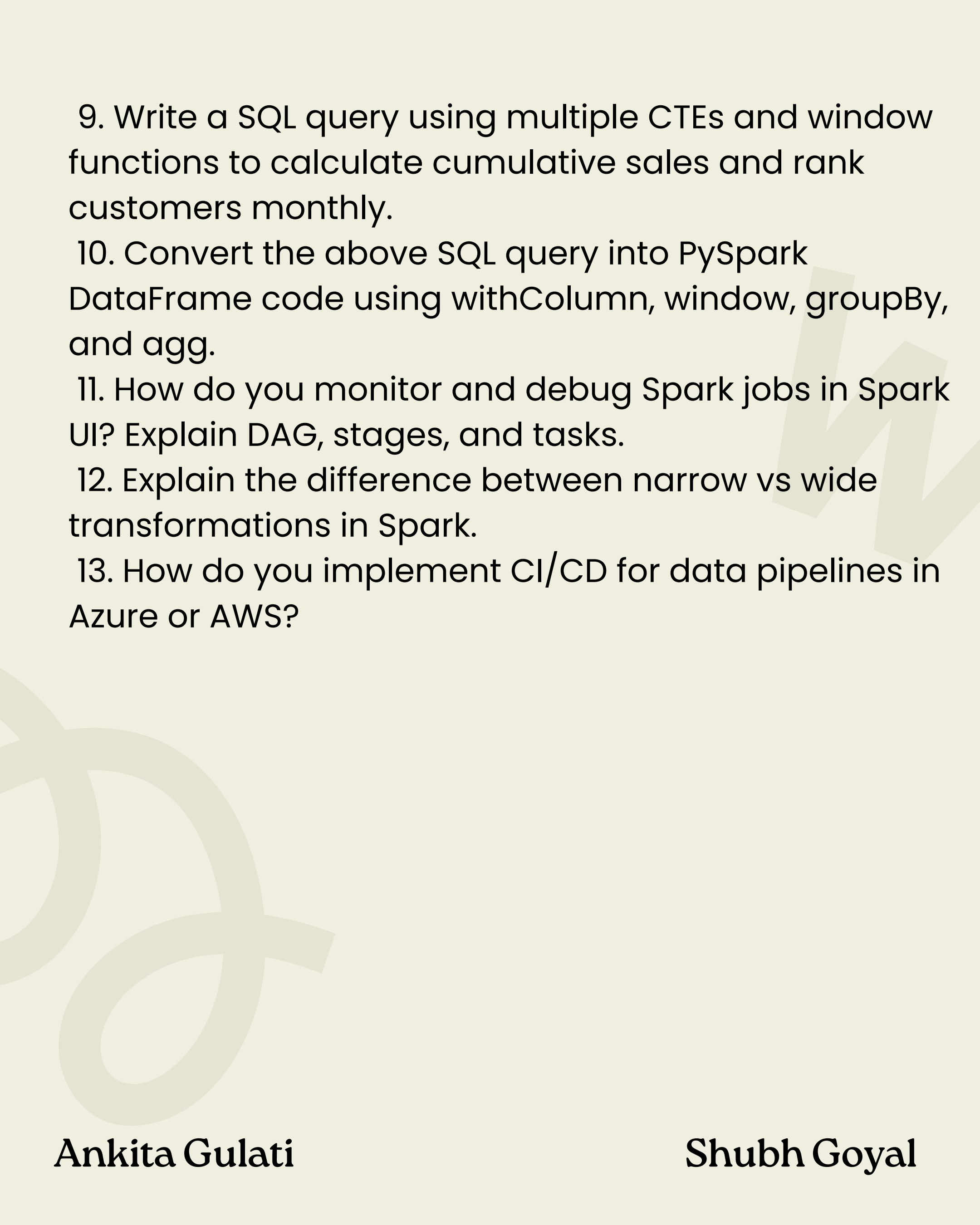Shubh Goyal

Round 1
Core Data Foundations

1. Explain the key concepts of RDBMS (tables, primary key, foreign key, normalization).

2. What is normalization and denormalization? Give examples of when you would use each.

3. Explain ACID properties in databases. Why are they important?

4. Differentiate between OLTP and OLAP systems with examples.

5. How would you design a data model for a ride-hailing app (Uber/Ola) that captures trips, drivers, and customers?

6. Explain the difference between star schema and snowflake schema in data modeling. Which one would you choose for reporting and why?

7. Write an SQL query to fetch the nth highest salary from an employee table.

8. Write an SQL query to list departments with more than 5 employees.

Ankita Gulati   Shubh Goyal

9. Write an SQL query to find duplicate records in a table and remove them while keeping only one instance.

10. Write an SQL query using INNER JOIN, LEFT JOIN, and FULL OUTER JOIN and explain their differences.

11. Write a query using window functions (ROW_NUMBER, RANK, DENSE_RANK) to find the top 3 salaries in each department.

12. Explain the use of LEAD() and LAG() functions with a real-world example.

Round 2
Spark, SQL & Project Discussion

1. Walk me through one of your end-to-end data engineering projects involving Azure and Spark.
2. Explain the architecture of the project (data lake, ETL pipelines, orchestration, monitoring).
3. How do you handle real-time streaming vs batch processing in your pipelines?
4. What Spark optimization techniques have you used? (partitioning, caching, broadcast joins).
5. Explain shuffle operations in Spark. How can you minimize them?
6. What is data skew in Spark? How do you identify and resolve it?
7. Explain schema evolution in Spark and how to handle it in ETL pipelines.
8. What are the advantages and limitations of using UDFs in Spark? When should you avoid them?

Ankita Gulati   **Shubh Goyal**

9. Write a SQL query using multiple CTEs and window functions to calculate cumulative sales and rank customers monthly.

10. Convert the above SQL query into PySpark DataFrame code using withColumn, window, groupBy, and agg.

11. How do you monitor and debug Spark jobs in Spark UI? Explain DAG, stages, and tasks.

12. Explain the difference between narrow vs wide transformations in Spark.

13. How do you implement CI/CD for data pipelines in Azure or AWS?

Ankita Gulati

Shubh Goyal

Round 3
Hands-On Coding & Optimization

1. Write a SQL query to find employees who earn above the average salary in their department.
2. Write a SQL query to pivot rows into columns (e.g., monthly sales pivoted by product category).
3. Write a SQL query to calculate running totals and moving averages using window functions.

4. Write PySpark code to:
- Remove null values from specific columns.
- Handle missing data by imputation.
- Apply schema evolution when a new column is added.

5. Convert a complex SQL query with multiple joins and aggregations into PySpark DataFrame transformations.
6. Explain the usage of explain() in Spark. How does it help in query optimization?

Ankita Gulati   Shubh Goyal

8. Explain how you would use broadcast joins in Spark to optimize joins with skewed data.

9. Write Python code to implement an LRU Cache.

10. Write Python code to check if a string is a valid palindrome ignoring special characters and case sensitivity.

11. How do you ensure idempotency in data pipelines (i.e., avoid duplicate loads)?

12. How would you design a data quality framework for ensuring valid, complete, and consistent data in your ETL pipelines?

Ankita Gulati   Shubh Goyal

Round 4
HR & Behavioral

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

2. What was your role and contribution in the most recent project?

3. What new skills and technologies are you currently learning?

4. Why do you want to join Accenture specifically?

5. Tell me about a challenge in your project and how you resolved it.

6. Are you open to relocation and flexible working models?

7. What are your long-term career goals?

8. Salary expectations and notice period.

Ankita Gulati   Shubh Goyal

Thank You

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

HIRED

Ankita Gulati   Shubh Goyal