Yash Technologies Data Engineer Interview Guide – Experienced 3+

Round 1 & Round 2 – Technical

Key Questions and Topics:

1. **Introduction and Experience**

   - Briefly introduce yourself and share your journey as a Data Engineer.

   - Describe your current project, including technologies, architecture, and responsibilities.

   - Discuss the average data volume handled and strategies used for efficient processing.

   - Highlight challenges faced in your projects and how you overcame them.

2. **Azure Data Factory (ADF)**

   - What is ADF? Is it an ETL or ELT tool, and why?

   - Explain linked services and how they are created.

   - Difference between linked services and datasets in ADF.

   - Integration Runtimes (IR): Types and their use cases.

   - Explain tumbling window triggers in ADF.

   - How do you move ADF pipelines from development to production using ARM templates?

3. **SQL**

   Write a query to find the second-highest salary in a table.

4. **Python**

   Write a function to find the top 3 largest numbers in a list.

5. **Data Validation and Schema Management**

   - How do you handle data validation using SQL or Python?

   - Explain strategies for managing schema changes in PySpark over time.

6. **PySpark and Spark**

   - Why is RDD considered resilient and fault-tolerant?

   - What is lazy evaluation in Spark, and how does it impact performance?

   - Compare persist() and cache() in Spark.

   - Difference between reduceByKey() and groupByKey() in Spark.

   - DataFrames vs. RDDs: Key differences and use cases.

7. **Azure Databricks**

   - What is a mount point in Azure Databricks? How do you mount ADLS Gen2 to Databricks?

Round 3 – Managerial

Key Questions:

1. **Professional Background**

   - Provide a brief overview of your career.

   - Describe your project team size and the roles of other members.

2. **Project Architecture**

   Explain the architecture of your current project and your role in its implementation.

3. **Data Warehouse Design**

   - Star Schema vs. Snowflake Schema: Which one you implemented and why?

   - Discuss differences between ROW_NUMBER(), RANK(), and DENSE_RANK(), and provide examples from your projects.

4. **Stakeholder Collaboration**

   How do you ensure smooth communication between data scientists, business teams, and developers?

Round 4 – HR Discussion

Key Topics:

- Salary negotiations.

- Notice period and availability.

Glassdoor Yash Technologies Review –

https://www.glassdoor.co.in/Reviews/YASH-Technologies-Reviews-E14704.htm

Yash Technologies Careers –

https://www.yash.com/about-us/career/

Subscribe to my YouTube Channel for Free Data Engineering Content –

https://www.youtube.com/@shubhamwadekar27

Connect with me here –

https://bento.me/shubhamwadekar

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar