# Matrix Data Engineer Interview Guide – Experienced 3+

## Round 1: Technical Round

1. **Introduction**

   The interviewer began with a brief "Introduce yourself" question, focusing on my background and relevant experience in data engineering.

2. **Conceptual Differences**

   Explained the differences between Data Warehouse, Data Lake, and Delta Lake, highlighting their use cases:

   - Data Warehouse for structured data and analytics.
   - Data Lake for storing large volumes of raw data.
   - Delta Lake as a hybrid solution providing ACID compliance for data lakes.

3. **Spark Execution Flow**

   - Described the execution flow of a Spark job, focusing on how transformations are lazy and executed only when an action is called.
   - Discussed the creation of DAGs (Directed Acyclic Graphs) and how Spark stages jobs.

4. **Memory Management in Spark**

   - Explained Spark's memory management, including executor memory, storage memory, and shuffle memory.
   - Differentiated between broadcast variables (used for sharing a read-only variable across tasks) and accumulators (used for aggregating values).
   - Real-Life Example: Used broadcast variables to distribute a small lookup table across nodes for faster joins.

5. **Handling Skewed Data**

   Discussed strategies to handle skewed data:

   - Salting keys to evenly distribute data.
   - Using repartition() or coalesce() to optimize partitioning.

6. **Salting Implementation**

   Provided an example of salting keys to mitigate skew:

```
from pyspark.sql.functions import col, concat, lit, rand

df = df.withColumn("salted_key", concat(col("key"), lit("_"), (rand() * 10).cast("int")))
```

© Shubham Wadekar

For personal use only. Redistribution or resale is prohibited. © Shubham Wadekar

7. **Spark Configurations for Large-Scale Jobs**

   Explained the process of deciding **executor memory, cores, and driver memory**:

   - Consider data size, complexity of operations, and cluster capacity.
   - Highlighted tuning using Spark UI for optimal performance.

## Round 2: Technical + Problem-Solving Round

1. **Scala Basics**

   Explained the difference between var, val, and def in Scala:

   - var: Mutable variable.
   - val: Immutable variable.
   - def: Method definition.

2. **SOLID Principles in Scala**

   Described the **SOLID principles** with examples in Scala:

   - Single Responsibility, Open-Closed, Liskov Substitution, Interface Segregation, and Dependency Inversion.

3. **Traits in Scala**

   Explained how **traits** are used for multiple inheritance and reusable code.

4. **Monads in Scala**

   - Defined **Monad** as a design pattern for chaining operations.
   - Example: Option in Scala for handling null safety.

5. **Spark Scala Code**

   Task: Calculate a 7-day moving average of clicks for each user_id.

```
import org.apache.spark.sql.expressions.Window

import org.apache.spark.sql.functions._


val windowSpec = Window.partitionBy("user_id").orderBy("click_date").rowsBetween(-6, 0)

val result = df.withColumn("moving_avg", avg("clicks").over(windowSpec))

result.show()
```

© Shubham Wadekar

For personal use only. Redistribution or resale is prohibited. © Shubham Wadekar

6. **SQL Query**

Task: Calculate cumulative sales for each product in each store, ordered by sale_date.

```
SELECT
    store_id,
    product_id,
    sale_date,
    SUM(sales_amount) OVER (PARTITION BY store_id, product_id ORDER BY sale_date) AS cumulative_sales
FROM
    sales;
```

7. **SQL Sorting and Partitioning**

Differentiated between SORT BY, ORDER BY, DISTRIBUTE BY, and CLUSTER BY:

- ORDER BY: Global ordering of results (single reducer).
- SORT BY: Local ordering within partitions.
- DISTRIBUTE BY: Partitioning data based on column values.
- CLUSTER BY: Combines DISTRIBUTE BY and SORT BY.

© Shubham Wadekar

**Glassdoor Matrix Review –**

https://www.glassdoor.co.in/Reviews/Matrix-Reviews-E832341.htm

**Matrix Careers –**

https://www.matrixcomsec.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

For personal use only. Redistribution or resale is prohibited. © Shubham Wadekar