Ford Motor Data Engineer Interview Guide – Experienced 3+

Technical Round 1

1. **Introduction**
   The interview began with a standard introduction where I briefly talked about my professional journey, current role, key projects, and technologies I work with.
2. **Pipeline Design and Explanation**
   I was asked to explain a data pipeline from my project. I described an end-to-end GCP-based pipeline handling data ingestion from Oracle into Google Cloud Storage (GCS), processing with Dataproc (PySpark jobs), and storing results in BigQuery for reporting.
3. **GCP Tools**
   I highlighted key tools:
   - Dataproc for running Spark jobs
   - BigQuery for analytical querying
   - Pub/Sub for messaging
   - Firestore for semi-structured data
   - GCS for storage
4. **Day-to-Day Activities**
   I shared a typical day involving data pipeline monitoring, developing new transformations, resolving production issues, and collaborating with data analysts to improve query performance.
5. **Clustering in Real-Time**
   I explained clustering with a real-time example of customer segmentation in retail data using k-means for grouping similar purchasing behaviors.
6. **Production Issues and Approach**
   I described how I resolved a production issue involving delayed Pub/Sub messages by debugging logs, identifying bottlenecks, and optimizing Spark job configurations.
7. **Agile: Adding New Requirements**
   Explained how new requirements are added via change requests reviewed during backlog grooming sessions.
8. **Sprint Duration**
   Sprint duration in my project is two weeks.
9. **Source and Target Systems**
   Source: Oracle database
   Target: BigQuery
10. **Joins in SQL**
    Explained INNER, LEFT, RIGHT, and FULL joins with examples.
11. **Types of Views**
    Covered Simple, Materialized, and Indexed views.
12. **Native vs. External Tables**
    Native tables store data within the database, while external tables reference data stored externally (e.g., in GCS).

Technical Round 2

13. **Access Granting for Views**
    I mentioned that access is managed by a DBA or Security Team based on roles and policies.

14. **Source Issue Resolution**
    Explained interaction with the source team, using tools like Jira to raise issues and Slack for real-time communication.

15. **Script Implementation**
    Described deployment through CI/CD pipelines using Git and Jenkins.

16. **P1 Issues**
    Shared experience handling a P1 issue involving data loss by restoring from a backup and reprocessing missed data.

17. **P2 Ticket SLA**
    P2 tickets typically have a resolution time of 24-48 hours.

18. **Removing Duplicates with Partitioning**
    Using ROW_NUMBER() over (PARTITION BY...) to remove duplicates.

19. **Inserting Columns Between Others**
    Mentioned it requires creating a new table with desired schema and migrating data (as direct insertion isn't supported).

20. **Additional Calls**
    Apart from standard daily stand-ups, we have weekly sync-ups with stakeholders.

21. **Ticket Beyond Story Point Duration**
    Explained how we request an extension with proper justification during sprint reviews.

22. **User Interaction**
    Yes, I interact directly with business users for understanding requirements and gathering feedback.

23. **Notice Period and Relocation**
    Shared my notice period and willingness to relocate.

Glassdoor Ford Review –

https://www.glassdoor.co.in/Reviews/Ford-Motor-Company-Reviews-E263.htm

Ford Careers –

https://www.careers.ford.com/en/home.html

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