# Data Design Approach (DDA) for Log Movement

## 1. Objective

Explain the purpose of the data movement and how it aligns with business needs (e.g., reduce on-premises storage costs, improve query performance using Athena).

## 2. Data Flow Overview

Create a high-level flowchart illustrating the process:  
- Extract logs from Oracle.  
- Transform them into Parquet format for Athena compatibility.  
- Load them into an S3 bucket (Standard-IA class for hot data, Glacier for archival).  
- Update the Oracle table row with a 'migrated' flag or timestamp.  
- Periodically delete migrated rows from Oracle.

## 3. Source System (Oracle)

Define:  
- Table schema and structure of log data.  
- Query for extracting logs older than 30 days.

## 4. Target System (AWS S3 and Athena)

Define:  
- S3 bucket hierarchy (e.g., `/logs/YYYY/MM/DD`).  
- Storage class transitions: Standard-IA → Glacier.  
- Parquet file format schema and compression standards for Athena queries.  
- Athena table schema mapping to the Parquet files.

## 5. Data Migration Steps

Detail the steps:  
- \*\*ETL Process\*\*: Use a tool or custom script (e.g., Python with `boto3`, or AWS Glue).  
- \*\*Update Process\*\*: Update rows in Oracle with a `migrated` status or timestamp after uploading to S3.  
- \*\*Batch Deletion\*\*: Schedule a batch job (via cron or SQL) to delete `migrated` rows periodically.

## 6. Error Handling

Plan for:  
- Failed migrations (retry mechanism or logging failed records).  
- Transaction rollbacks in Oracle if upload fails.

## 7. Access and Security

Detail access controls:  
- Oracle read access.  
- S3 bucket policies.  
- IAM roles for uploading data and querying with Athena.

## 8. Retention Policy

Document:  
- Retain logs in S3 Standard-IA for 3 months.  
- Move logs to Glacier after 3 months.  
- Set lifecycle rules to delete Glacier logs after 1 or 3 years.

## 9. Automation

Describe the tools used for automation:  
- Oracle query scripts or tools for extraction.  
- S3 upload scripts (e.g., Python + `boto3`).  
- AWS S3 lifecycle policies for retention management.  
- Athena for querying logs.

## 10. Testing and Validation

Explain how to validate:  
- Logs are uploaded and queryable in Athena.  
- Oracle table updates are accurate.  
- Logs in S3 follow the lifecycle policies.

## 11. Monitoring

Set up monitoring:  
- AWS CloudWatch for S3 events and Athena queries.  
- Oracle logs for migration status.