📄 *E-Commerce Data Engineering Pipeline – Project Summary*

Objective:
To build an end-to-end data pipeline that extracts raw sales data from AWS S3, cleans and transforms it using AWS Glue (PySpark), models it into a star schema on Databricks, and visualizes insights through Power BI.

Tools & Technologies:

- **AWS S3** – Data lake storage
- **AWS Glue (PySpark)** – ETL transformation and incremental loading
- **Databricks** – Fact & Dimension modeling (Star Schema)
- **Power BI** – KPI visualization and dashboard creation

Pipeline Flow:

1. Raw data stored in S3 (raw/)
2. Processed via AWS Glue → staging/
3. Clean, incremental data in curated/
4. Databricks creates Fact & Dim tables in warehouse/
5. Power BI reads sales_summary from analytics/ layer for insights

Key Outcomes:

- Automated incremental data pipeline
- Optimized data storage using Parquet
- Star schema enabling quick analytics
- Business-ready Power BI dashboard