

1. Load Transaction Data from ADLS Gen2:

- Load transaction data from CSV files stored in ADLS Gen2 into a Spark Data Frame.

2. Extract Insights:

- Join transaction data with product information from a separate CSV file in ADLS Gen2 to obtain product details like price, description, and category.

- Perform analysis on combined data, including calculating average order value for each customer, identifying popular products and categories, and analyzing the impact of marketing campaigns on sales.

3. Create a Managed Delta Table:

- Create a managed Delta table on Azure Databricks.

- Load the extracted insights into the Delta table.

4. Optimize Data Storage:

Optimize data storage using Delta Lake's snapshot feature to store data in a compact format.

5. Monitor Data Quality:

Monitor data quality by checking for missing values and outliers.

NOTE:

Handle data from multiple channels (web, mobile, in-store) and combine it for analysis.

Utilize Spark's built-in functions and data types for efficient data processing.

Include data validation and error handling to ensure accurate processing.

Optimize code by creating reusable functions for data processing and transformation.