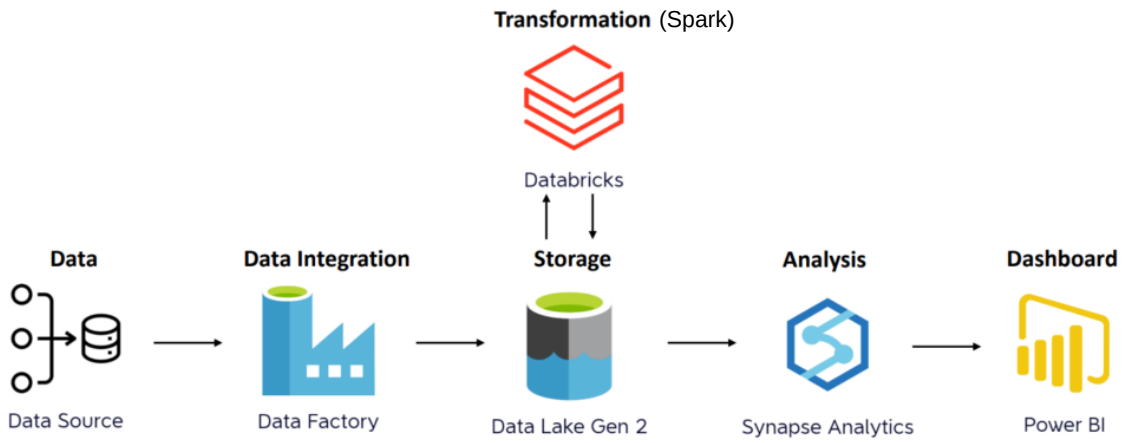
Azure Data Engineering Project

Introduction

The Sports Retail Data Azure Data Engineering Pipeline project aims to create a robust data pipeline for a sports retail business. This end-to-end solution leverages Azure services and Power BI to enable data-driven decision-making and enhance operational efficiency.



Task

The primary objective of this project is to design and implement a data engineering pipeline that can ingest, process, store, and visualize data from various sources, including transactional databases, third-party APIs, web scraping, and internal data systems. The pipeline comprises the following key components and tasks:

Steps

1. **Data Ingestion:**
    - Data from multiple sources is ingested into Azure Data Lake Gen 2.
    - Azure Data Factory is used to schedule and coordinate data ingestion activities.
2. **Data Transformation:**
    - Azure Databricks is employed for data cleaning, transformation, and enrichment.
    - Databricks notebooks are developed to ensure data quality and prepare it for downstream analytics.
3. **Data Storage:**
    - Cleaned and transformed data is stored in Azure Data Lake Gen 2 for easy accessibility.
    - The hierarchical structure of Data Lake aids efficient data organization.

4. **Data Warehousing:**
    - Azure Synapse Analytics is utilized for data warehousing and advanced analytics.
    - Data is curated and aggregated within Synapse Analytics for complex queries.
5. **Report Development:**
    - Power BI connects to Azure Synapse Analytics to create interactive dashboards and reports.
    - These reports provide actionable insights into sales performance, inventory levels, customer behaviour, and other key performance indicators (KPIs).

Dashboard