# My Understanding of the Azure DE Project

## Project Title:

Lululemon Data Analysis – Azure Data Engineering

## Project Overview:

This project focuses on analyzing Lululemon-related data by building a cloud-based data pipeline using Azure services and other modern data engineering tools. The goal is to ingest, process, and analyze data efficiently to support business insights.

## Objective:

To develop a foundational understanding of data engineering workflows using Azure tools and to analyze data in both batch and real-time modes. It follows the ELT process that the data extraction from the Kafka to load into Snowflake for transformations.

**Technologies Involved:**

* **Apache Kafka:**  
  Used for real-time data ingestion from various sources such as sales platforms, store systems, or third-party APIs.
* **Azure Data Factory (ADF):**  
  A key tool for orchestrating the entire data pipeline. It manages data movement and transformation across different stages and systems.
* **Azure Databricks:**  
  An Apache Spark-based platform used for scalable data processing and advanced transformations.
* **Snowflake:**  
  A cloud-based data warehouse used for storing and querying large volumes of structured data efficiently.
* **Azure DevOps:**  
  Enables CI/CD pipelines for automating deployment, testing, and version control in the data engineering lifecycle.

**Flowchart: Lululemon Sportswear Data Pipeline**

**Lululemon Data source**

**Apache Kafka**

**Azure Data Factory (ADF)**

**Snowflake**

**Analysis & Reporting or view**

## Final Understanding:

• Kafka collects and streams real-time data from Lululemon systems.

• Azure Data Factory orchestrates data pipelines, scheduling and monitoring movement between Kafka, Databricks, and Snowflake.

• Processed data is loaded into Snowflake for querying and analysis.

• Azure DevOps manages source code, CI/CD pipelines, and deployment automation for the data pipeline.

• This entire architecture allows efficient, scalable, and automated handling of both real-time and batch data for analytical purposes. And we also have some data pipelines and data sets