**Part1: Batch E-commerce Data Processing and Analysis**

* **Objective**: Examine e-commerce customer behavior using the dataset available on Kaggle, which provides comprehensive e-commerce customer behavior data​​.
* **Dataset**: https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset
* **Questions**:
  + Can we segment customers based on their demographic information (Age, Gender, City) and shopping behaviors (Total Spend, Number of Items Purchased, Membership Type)?
  + Which customers are at risk of not making future purchases based on their Days Since Last Purchase and Satisfaction Level?
  + Can we predict a customer’s Satisfaction Level based on their demographic and purchase history data?

U should use HADOOP, SPARK and YARN (optional) for this implementation.

* **Hadoop Role**: Store and preprocess the large datasets of user logs and transaction data.
* **Spark Role**: Use machine learning algorithms to analyze customer behavior and predict future buying patterns.
* **YARN Role**: Efficiently manage resources for complex analytics tasks. (optional)