# **Walmart Inventory Optimization Project**

## **1. Business Use Case: Inventory Management and Demand Forecasting**

### **Business Challenge**

Walmart faces significant challenges in:

* Optimizing inventory levels across multiple stores
* Minimizing stockout and overstock situations
* Reducing inventory carrying costs
* Improving product availability and customer satisfaction

### **Business Impact**

* Enhanced operational efficiency
* Improved cash flow management
* Better customer experience
* Data-driven inventory strategies

## **2. Project Overview**

### **Project Goal**

Develop a comprehensive, real-time inventory management solution using advanced big data technologies to:

* Process historical sales data (Batch Processing)
* Capture and analyze real-time sales transactions (Streaming Processing)
* Provide actionable insights for inventory optimization

### **Scope**

* Integrate multiple data sources
* Implement real-time and batch data processing
* Create a scalable, flexible inventory analytics platform

## **3. Detailed Technology Stack**

### **Batch Processing Components**

1. **Apache Hadoop (HDFS)**
   * Distributed storage
   * Large-scale data storage
   * Fault-tolerant file system
2. **Apache Spark (PySpark)**
   * Batch data processing
   * Complex transformations
   * Machine learning preprocessing
3. **Apache Hive**
   * Data warehousing
   * SQL-like queries
   * Structured data management

### **Streaming Processing Components**

1. **Apache Kafka**
   * Real-time data ingestion
   * Event streaming
   * Message queue management
2. **Apache Spark Streaming**
   * Real-time data processing
   * Stream analytics
   * Micro-batch processing

### **Supporting Technologies**

* **Apache Sqoop**: Data integration
* **Zookeeper**: Cluster coordination
* **Airflow**: Workflow orchestration

## **4. End-to-End Data Flow**

### **Batch Processing Workflow**

Transactional Databases

→ Apache Sqoop → HDFS Storage → Apache Spark Batch Processing → Apache Hive Data Warehouse → Analytics and Reporting

### **Streaming Processing Workflow**

Point-of-Sale Systems

→ Apache Kafka → Apache Spark Streaming → Real-time Inventory Updates → Immediate Analytics → Decision Support System