

Problem Statement: ABC Corp, a leading retail US based fashion company, faces significant challenges in optimizing pricing strategies, managing orders efficiently, forecasting demand accurately, and streamlining distribution processes. These challenges hinder ABC Corp's ability to remain competitive in the fast-paced retail industry. We provide the analytics using the big data technologies to enhance and grow the business.

Solution:

Utilizing advanced technologies such as Hadoop and Kafka, ABC Corp can transform its operations. By leveraging Hadoop for robust data storage, processing, and analysis, coupled with Kafka for real-time data streaming and event processing, ABC Corp can optimize pricing strategies, enhance order management efficiency, accurately forecast demand, and streamline distribution processes. Additionally, integrating DevOps technologies such as Docker and Kubernetes enables seamless deployment, scaling, and management of applications, ensuring agility and reliability in ABC Corp's operations. This comprehensive approach empowers ABC Corp to stay ahead of market trends and maintain a competitive edge in the fast-paced retail industry.

Capacity Planning:

Cloudera Version :

Project Duration: 2 yrs

Daily ingestion : 25 GB/ Replication Factor is 3 So $25 * 3 = 75\text{GB}$

Hardware Selection:

OS-Ubuntu Server 20.04

HDD Size: 4 TB (2 TB's two HDD on each DN) 75% for HDFS usable i.e, 3TB

CPU Core Size: 8

RAM Size: 32 GB

Formula: $75 * (365 * 2) / 3000 \sim 19 \text{ Nodes}$

Average Node Size: 20

Master Nodes: 3

Worker Nodes: 16

Edge Node: 1

Overall Working:

By combining the capability to accurately track each consumer's full path in the store with a high-performance and scalable analytics engine, We can provide real-time analytics that helps brands, marketers, and retailers understand and learn how consumers behave while they shop in store. Our's analytics engine, built on Cloudera Enterprise, is able to process and deliver the massive amount of data that is collected—analyzing the millions of routes shoppers take throughout the store and their stops in each aisle—to help retailers understand shopping patterns and trends, and apply predictive models to suggest in-store improvements.

For example, something as simple as the direction in which the shopper walks the aisle can now help marketers determine where to place promotions.

Additionally, measuring and capitalizing on the conversion rates for products and marketing campaigns—metrics that have been successfully used by online retailers for years—is now possible in store.

How Do We Collect Data from Various Sources ?

Source 1: IOT sensors

In this system architecture, Bluetooth sensors positioned in the store's ceiling receive signals emitted by small beacons attached to shopping carts and baskets, enabling real-time tracking of shoppers' movements and behaviors. The collected data undergoes processing for analysis, including tracking which aisles shoppers visit, their approach direction, and the duration they spend in front of various items. Insights derived from this analysis inform decision-making processes such as optimizing store layout, improving inventory management, and providing real-time recommendations to enhance the overall shopping experience.

Source 2: CRM (Customer Relationship Management)

Centralized CRM systems offer insights into customer preferences, purchase history, and interaction patterns. This data drives personalized marketing campaigns, communications, and promotions. Integration of CRM data enhances customer engagement, satisfaction, and loyalty towards the brand.

Source 3: Store Applications(AI models)

Enhanced customer behavior insights are achieved through the integration of footfall counting applications and AI-driven person tracking algorithms, seamlessly linked with store cameras and DVRs. These advanced technologies measure foot traffic, determine customer demographics like gender and age, and analyze entry and exit times, empowering targeted marketing strategies and streamlined store operations.

Source 4: Suppliers API

Utilizing insights from diverse sources, including data from suppliers, we effectively manage inventories and logistics

Implementation:

We consolidate data from diverse sources into a centralized Apache Hadoop storage system.

We use Apache Kafka to handle data streaming data from IoT sensors.

We utilize Sqoop to transfer CRM data to the HDFS cluster, preparing it for future processing.

Also, we used the client's traditional database for Store Application and Suppliers data.

We have integrated various tools for the various teams like for the operations, data analyst, Business Intelligence, AI/ML and Data Scientist.

For the Data Analyst team we have the Hive , Impala tools.

For Business Intelligence we have the Tableau tool.

For AI/ML/DS we use the Sparks.

Also we have a complete devops CI/CD pipeline using Git,Jenkins,Dockers etc.

Cluster Architecture:

