CUSTOMER

Myntra

INDUSTRY

Mobile E-Commerce

USE CASE

Cloud Analytics

Data Pipeline

APPLICATION STACK

Spark + Kafka+ Presto

Alluxio

AWS S3 (object storage)

BENEFIT HIGHLIGHTS

- Shared data layer for higher performance analytics pipeline
- Faster customized response to mobile application customers
- Interactive reporting for business analysts resulting in faster time to insight and lower cost.

Accelerating Analytics in the Cloud for Customized Mobile E-Commerce

Myntra, a division of Flipkart, is a leading fashion retailer in India offering customers a wide range of merchandise through a mobile application. An analytics pipeline in Amazon Web Services (AWS) cloud processes customer data to make recommendations, present ads, and deliver other aspects of a tailored experience. Myntra deployed Alluxio to provide a virtual data layer connecting AWS S3 to the analytics pipeline to accelerate data access and enable faster customer response and interactive business intelligence.

The Challenge

Myntra is at the forefront of the Indian retail market with their mobile application. Taking advantage of mobile device penetration and demographic trends, the company generates 75 percent of sales through smartphones. Providing a tailored experience is the key to customer satisfaction and revenue growth. To achieve this, a data processing pipeline analyzes how customers interact with the application. Understanding shopping patterns, ad responses, and reporting on clicks provides the insight required to provide customers relevant information, recommendations, and products.

The Myntra analytics solution is deployed in the AWS cloud using S3 (object storage) as the storage layer. Initially, clickstream data was processed in a pipeline using Spark as the compute engine, Kafka as the messaging bus, and Presto for internal reporting and metrics data. Spark reads data from AWS S3, processes the data, and writes it back to S3.
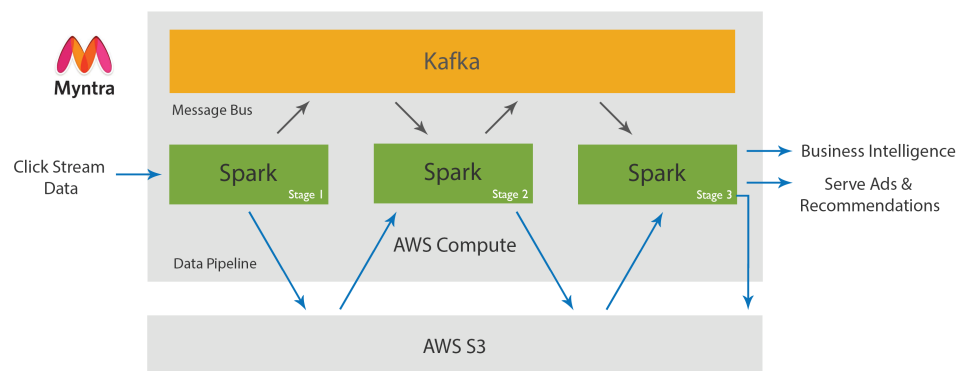

Figure 1: Initial data pipeline limited by direct data access to AWS S3 storage

The Solution

Myntra addressed their challenges by deploying Alluxio as the shared data layer for all of the intermediate stages in the data pipeline. By reading and writing data in Alluxio, the data can be read concurrently and stay in memory for the next stage of the pipeline. This increased the performance by speeding up the entire pipeline, increasing overall throughput, and enabling superior interactive response to application

users. With Alluxio in place, Myntra increased the time spent by customers on the application, a primary measure of customer success.

Additionally, most business intelligence reports were generated using the same data set. In order to speed this up, Myntra decided to store the data in Alluxio and AWS S3, with the reports being generated against the data in Alluxio. This greatly reduced the time required to generate reports and provide valuable insights into the business.
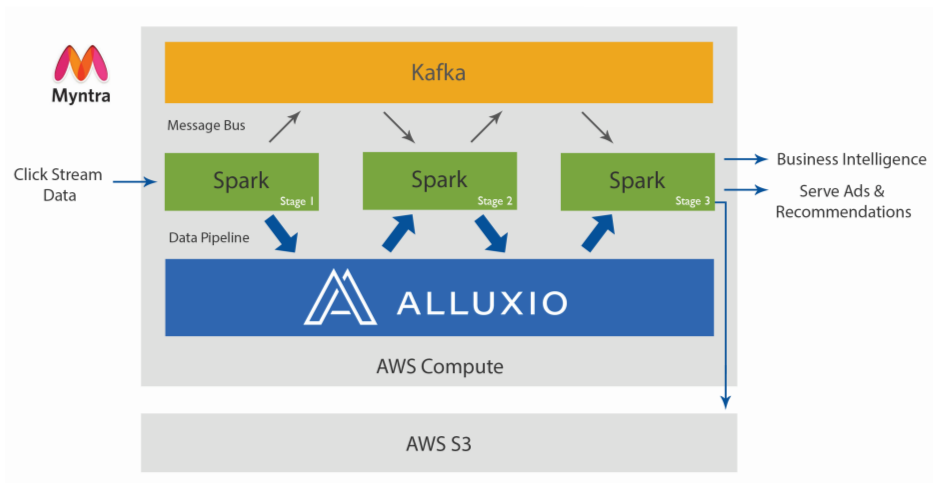

Figure 2: Data pipeline with Alluxio for increased performance and interactive response

The Results

Alluxio is in production as a critical component of the Myntra data processing pipeline architecture, significantly improving customer satisfaction, increasing revenue, and accelerating the ability to generate actionable business intelligence from customer data. With Alluxio, the company is now able to:

- Provide interactive response and improved customer experience to users which in turn leads to increased revenue.
- Accelerate interactive reporting for business analysts, enabling faster, higher quality insights and lower operational costs.

Myntra also gave back to the Alluxio community by contributing the documentation for deploying Alluxio with AWS S3 so other companies can benefit from their experience. Learn more at alluxio.org.

Stay Connected

Twitter: @Alluxio
LinkedIn: linkedin.com/company/alluxio-inc-/
Meetup: meetup.com/Alluxio/
Slack: alluxio.io/slack

About Alluxio

Proven at global web scale in production for modern data services, Alluxio is the developer of open source data orchestration software for the cloud. Alluxio moves data closer to big data and machine learning compute frameworks in any cloud across clusters, regions, clouds and countries, providing memory-speed data access to files and objects. Intelligent data tiering and data management deliver consistent high performance to customers in financial services, high tech, retail and telecommunications. Alluxio is in production use today at seven out of the top ten internet companies. Venture-backed by Andreessen Horowitz and Seven Seas Partners, Alluxio was founded at UC Berkeley's AMPLab by the creators of the Tachyon open source project. For more information, contact info@alluxio.com or follow us on LinkedIn, or Twitter.