



POWERING SEARCH WITH A MODERN DATA PLATFORM

IMPACT

- Cross data center replication abilities globally
- Meeting SLA requirements with robust security and high availability
- Increased productivity for DevOps team with data search and discovery

Challenge

This global information technology services company provides one of the largest e-commerce platforms in the world to its customers. As more internet transactions are completed, the volume of data to be processed has grown exponentially. The e-commerce increase drove the company's need to expand capacities for its global data centers.

One of the key use cases is related to improving the relevance of information for discovery. The company needed a semantic search engine to power the search function for all the applications on its platform. This would help it understand the user's intent through the search context and improve relevance of the results.

To provide the foundation for this, the company needed to create a modern architecture framework to enable better searches, replicate data, and do experimental customization analytics. Moving data was important to allow efficient search-ability, which the company was struggling to do effectively. The platform would also be designed to handle the increased search traffic and data volumes expected in the future - which is an important step towards an intelligent enterprise. Adhering to the highest security and compliance standards plus avoiding high costs were additional challenges.

Solution

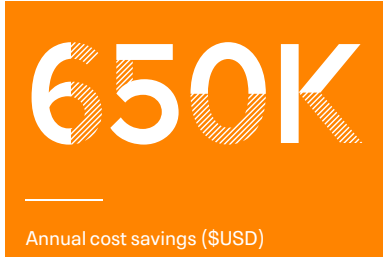
To originally address these challenges, the company chose Cludera's CDH platform a few years ago. With CDH coming to end-of-life and the company's requirements for keeping all customer data on premises, it recognized the need to upgrade - beginning a migration of 900+ nodes to CDP Private Cloud Base for all data processing and storage.

The company also needed a solution for streaming workloads, selecting Cludera Data Flow's (CDF) Streams Messaging - including Apache Kafka, Streams Replication Manager (SRM), and Streams Messaging Manager (SMM) - replacing Confluent. This decision was made for a number of reasons. The company wanted the flexibility to replicate data across data centers globally. The data would be secured, backed up, and then replicated to another data center to meet SLA requirements while maintaining high availability of the data. It needed Kafka support to buffer streaming data for their use cases, and going with Cludera led to cost savings, additional features, and the highest level of security for the company's transactional data.

They preferred the monitoring and deployment management using Cludera Manager compared to Confluent Control Center. Additionally, SRM (for data replication) and SMM (for data monitoring) were easy and provided a great deal of observability right out of the box.

Results

Through this data journey, the company delivered a complete scalable platform with extensibility and unified integration for managing all business transactions. The latest open source streams messaging capability, including Kafka, SMM and SRM, enable DevOps to experiment with new opportunities that better serve their customers. Kafka helps categorize data efficiently to make it more relevant to do specific catalog searches. It easily separates the different types of transactional data out by topic so that the search engine can focus on the data it cares about, enabling it to handle increases in search traffic volume and making things much faster.



About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow.

We empower people to transform complex data into clear and actionable insights.

Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI.

Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

The platform provides comprehensive process capability with immersive end-user experience. Now there is visibility through the entire Kafka lifecycle - of all data coming in and out - along with latency and throughputs. SMM provides monitoring capabilities to avoid things getting stuck, helping to pinpoint and isolate problems along the pipeline so that they can immediately be addressed. Moving to Cloudera has also resulted in \$650k+ in yearly savings from Confluent licensing costs.