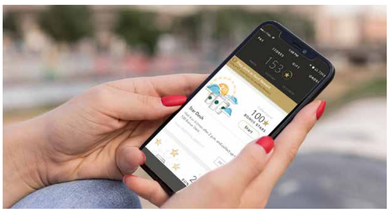
Case Study | Intelligence

Data Lake Modernization to drive cost-efficiency & faster decision-making for a retail customer

Our client is one of the largest and most valuable free membership loyalty programs in the world, connecting savvy shoppers with more than 3,500 online merchants and services to give them a hassle-free way to save on everyday purchases.

The client performs various insights and analysis on shopper behaviour, pricing, commissions and much more to create compelling offers and satisfying service for their customers.



How Did It All Start

As multiple business units started running various workloads and on-demand analytics on the platform it started to take a toll on disks and processors, reaching its critical mass limit/maximum threshold. The increase in data volume created more demand from platform resources such as hard disk access, memory, and CPU time for resource-intensive processes. The client's current environment couldn't manage the needs and demand for scalability.

Operations engineers were spending most of their time in troubleshooting hardware and services failures. The client needed a new environment for their data and a new way to query their data that would ensure efficient and responsive analysis during peak demand periods.

TVS Next did a 4 weeks POC for the customer to transfer 15TBs of on-premise Hadoop clusters data to S3 while supporting current Hadoop environment.

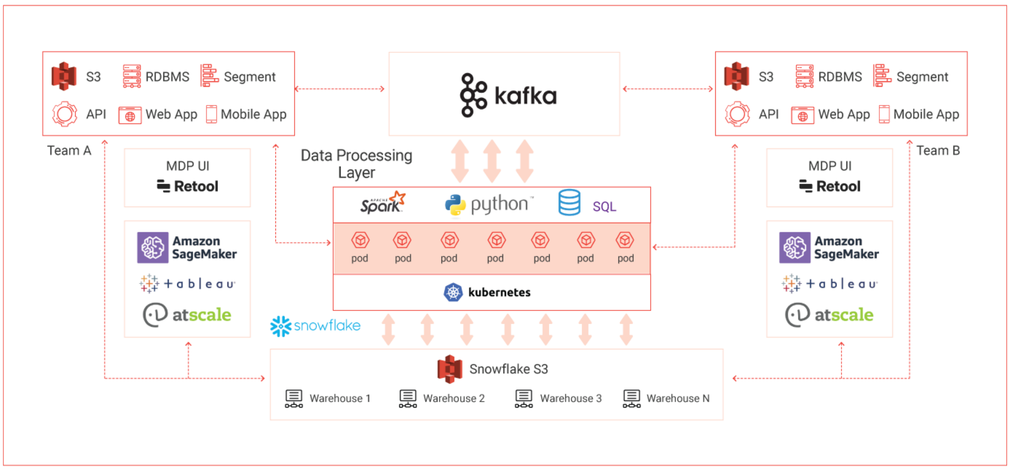
What Did We Do

TVS Next team restructured the client's data infrastructure and migrated the client data warehouse workloads from their on-premises Cloudera Hadoop cluster to AWS using a combination of services like Amazon S3, AWS EMR, and Snowflake. The data marts were created using Snowflake on AWS. AtScale was used to provide query performance optimization and a single, virtualized view of the data delivered as Data-as-a-Service.

TVS Next team leveraged its Elastic Data Platform solution accelerators to build a job orchestration platform to ingest data (original data store layer), process data (data warehouse layer) and store in data warehouse. This involved near real time jobs and batch jobs. The data ingestion layer ingested data from various sources to Snowflake on AWS. The data processing layer read data from Snowflake and then it was processed and written to a final data warehouse layer.

A new user to the system could now easily schedule and configure a job. Approximately 600 jobs were migrated and now run in the cloud using the job orchestration framework.

Architecture



The Business Outcome

30% drop in load on the on-premises computing cluster

68% improvement in real-time feedback & live monitoring

98% speedy delivery & release cycles

Zero downtime due to infrastructure updates with no negative impact on business

270 TB data migrated to a hybrid cloud platform with Snowflake

At TVS Next, we re-imagine, design, and develop software to enable our clients deliver differentiated experiences.
To know more, email us at marketing@tvsnext.io.