

A leading American multinational computer software firm improved its customer scoring engine performance by 2X through its transition from its on-prem Hadoop to Azure Databricks



Customer

The customer is an American multinational computer software company in the content creation product development space

Business Challenge

The customer had embarked on a multiyear initiative focused on moving their on-prem Cloudera Hadoop instance to Azure.

The existing customer scoring and cross sell recommendation solutions built on **Hadoop Map Reduce engine and Hive Queries (HQL)** had following challenges:

1. Delay in the recommendations reaching the sales team
2. Infrastructure limited the scope of data to be evaluated for the scoring
3. Difficult to manage the scoring parameter change requests from the business

WinWire Solution

WinWire, in collaboration with the customer has envisioned to address the existing challenges by adopting a **Spark based architecture with Azure Databricks**.

WinWire team did an end-to-end assessment of the **existing MapReduce code, Hive queries** and designed a modern data pipeline to Spark code seamlessly. This transition enabled the customer to process data faster and improved the overall performance of the job by reducing the **executing time by more than 50%**.

We migrated **Hive HQL/MapReduce to Spark/Scala code** based on design pattern analysis automation. Data between the old and new process was validated as part of integration testing at a target table level. Metadata validation was also performed like the data type of all columns and number of columns.

Environment:

Legacy: Cloudera, PIG, Hive, Oozie 1000+ nodes, 20 TB data

Azure: Databricks, Data Lake Store, Scala

Business Value

- Improved experience through on time delivery of insights to the sales team
- Better opportunity identification and planning for improved business performance
- Eliminated manual effort of monitoring and reporting the scoring process for business through automated alerts