A leading American multinational computer software firm transitioned its Hadoop MapReduce to Apache Spark to improve business performance

Customer

The customer is an American multinational computer software company with game-changing innovations that are redefining the possibilities of digital experiences. The customer connects content and data and introduces new technologies that democratize creativity, shapes the next generation of storytelling, and inspires new categories of business.

Business Challenge

The customer had embarked on a multiyear initiative focused on moving their Big Data platform from Cloudera Hadoop On-Prem instance to Cloudera Data Platform (CDP) on Azure. As a first step, they wanted to explore the prioritized MapReduce jobs in the current state and consider migration to Spark before moving the workloads to Azure Cloud.

They had initially created a solution with Hadoop Map Reduce engine and Hive Queries (HQL). The current setup had the following challenges:

1. Slower code execution speed
2. Higher storage requirement
3. Difficult to maintain workflows

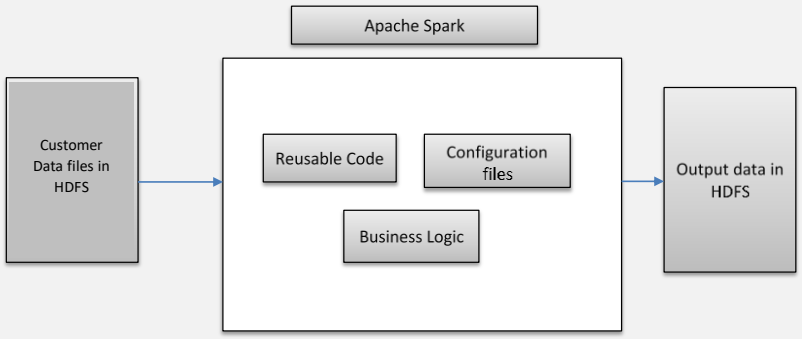
The newer solution they envisioned should address all issues mentioned above and wanted a revamped approach in processing Big Data. They were looking for a partner that could support them in converting identified MapReduce Jobs to Spark as they wanted **to reduce the execution and processing time** of Jobs as it was **impacting their business performance.**

Eventually it will enable them to move their Big Data platform from Cloudera Hadoop On-Prem instance to Cloudera Data Platform (CDP) on Azure.

WinWire Solution

WinWire, in collaboration with the customer has taken two prioritized jobs [LTV & AES] to convert MapReduce jobs to Spark. These were categorized as high complexity jobs.

WinWire team transitioned MapReduce code 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%.



Technologies used: Hive, Spark -2.4, Scala – 2.11, IntelliJ Idea Community Edition – 2021.1, Unravel, Hive Shell, Spark2-shell, CDH – 5.16, GitHub

Business Value

- Reduced the execution and processing time of job by 50%
- Greater customer satisfaction through better project execution
- Better opportunities and improved business performance

WinWire Technologies

WinWire Technologies Inc. 5201 Great America Pkwy,
5201 Great America Pkwy, Suite 320, Santa Clara, CA 95054; Phone/FAX: +1 (888) 849 7339;
Email: info@winwire.com Website: www.winwire.com