



Powering Data Science Use Cases at a CPG company

Consumer Packaged Goods



Overview

A multinational CPG company operating worldwide needs to grow its sales at a higher rate. They have different regions of the world like APAC, North America, South America etc. In order to increase their sales, they focused on implementing and rolling out Data Science use cases to the field. They narrowed down these use cases to product recommendations, applying price elasticity for market share gains, and using algorithmically driven promotions. The company has various datasets including customers, products and sales for the various regions.



Data Science Use cases

Need to roll out Advanced Data Science use cases to the field for increased market share and sales.



Recommend Products

Applying AI for better product recommendations to their customers.



Various datasets

The company has various datasets including customers, products and sales for the various regions.

Challenges

Building these use cases effectively involves very heavy optimizations of the algorithms, deep understanding of the customers and their behavior. It also involves continuous evolving of the results with continuous changes happening including change of seasons, changes in the market due to Covid and changing priorities of the Sales team. The massive size of the datasets and in many cases the extreme performance tuning needed for the algorithms to complete in a meaning time and using meaningful compute resources, makes building the effective solution very daunting. The variety of new technologies including Databricks, Azure, Snowflake, various ML engines and algorithms, made building the end system very challenging. It is important to enable many more people in the organization to be able to work together on building the end solution. Limiting things to a very small team of data scientists significantly slows down the process.



The massive size of the datasets and in many cases the extreme performance tuning needed for the algorithms to complete in a meaning time and using meaningful compute resources makes building the solution very effective.

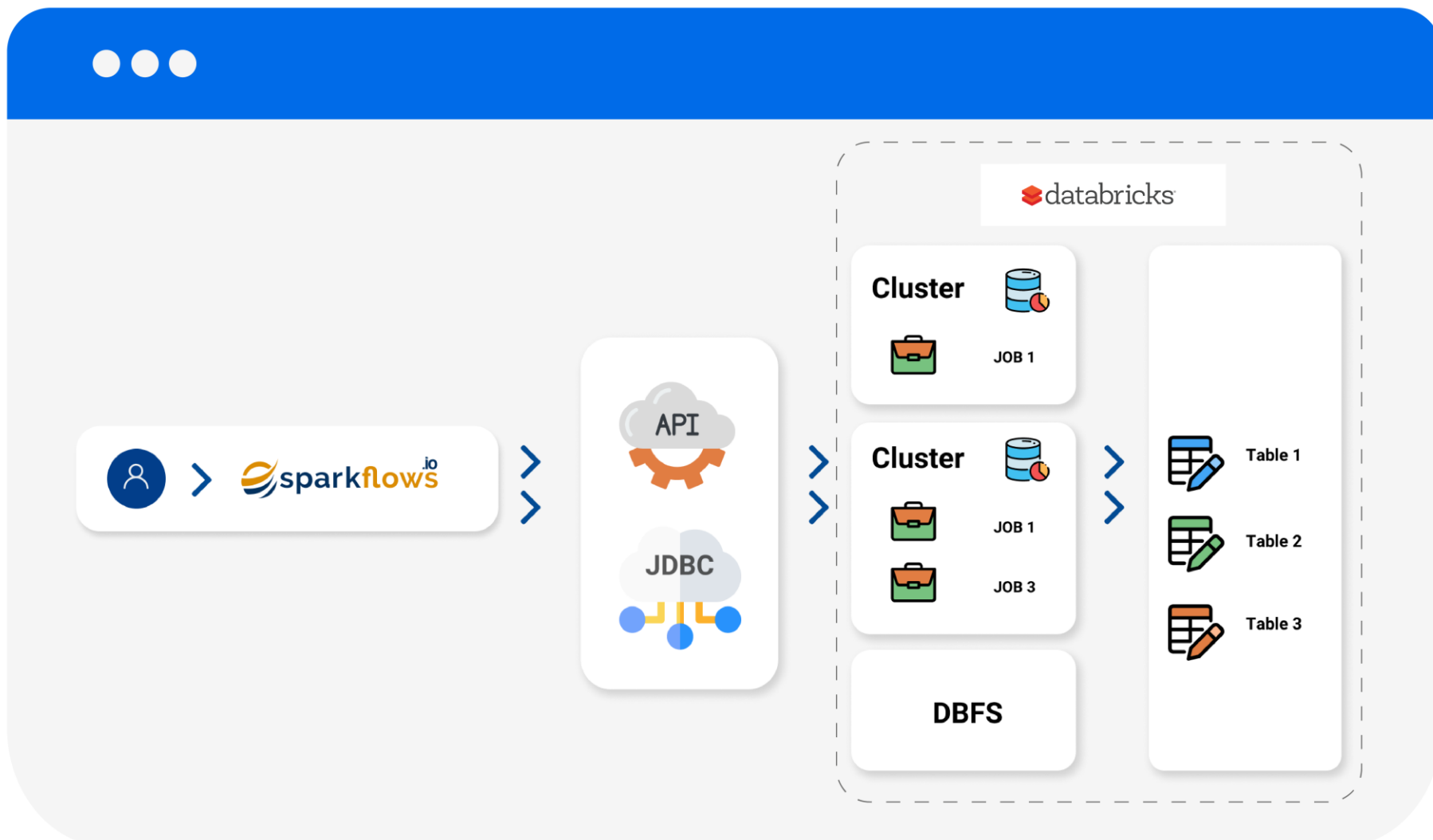
Solution

The solution involved installing Sparkflows in the current environment. The current environment has Databricks as the compute platform running on the Azure Cloud. The data is on Azure ADLS, Databricks tables and also in Snowflake.

Sparkflows was connected to execute jobs on Databricks. Connections were also defined for reading from and writing data to Snowflake.

The various users of the customers who are distributed across the world got access to

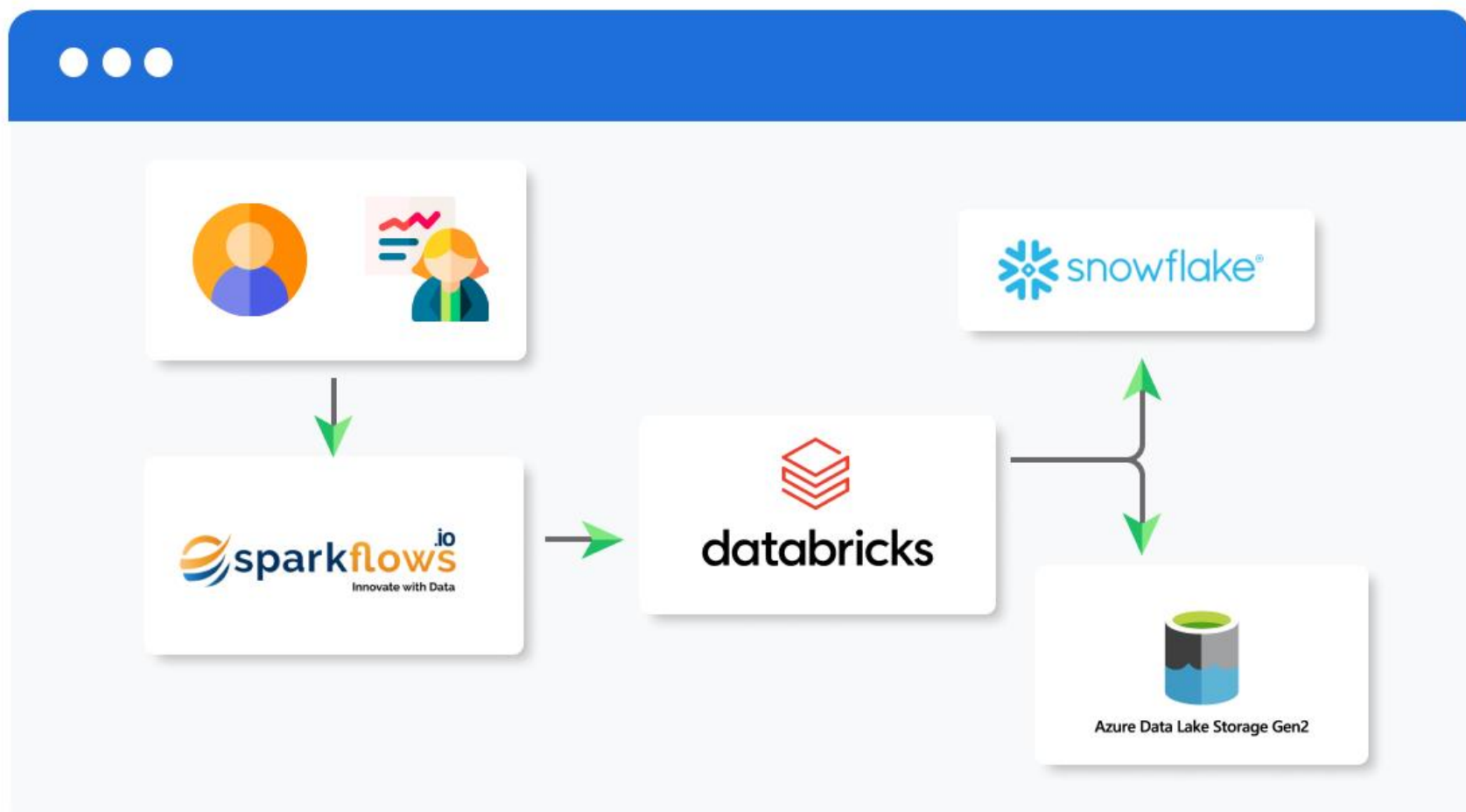
Sparkflows. They went through a training session of a few hours which enabled them to start building their data science use cases. Building the Analytics and Data Science use cases using workflows in Sparkflows along with the 330+ processors available in Sparkflows enabled the data analysts and data scientists to focus on building a very powerful solution.



Sparkflows allowed them to easily read data from Databricks, Snowflake and ADLS to prepare and clean the data and explore it using the various visualizations available in Sparkflows. Then they went on to build ML models using the 100+ ML processors available in Sparkflows. It enables Machine Learning with Apache Spark ML, Scikit Learn, Keras, Prophet, ARIMA etc. The use cases were built out and then the process of field testing started. The results of field testing were easily evaluated for AB testing by again building the workflows.

The output from Sparkflows was also integrated with external systems which were then used by the businesses for using the various recommendations, pricings etc. coming out of the ML Algorithms.

The various workflows were checked into Github for versioning. Dev, QA, and Production environments were set up. The workflows could easily move from Dev to QA to production and with continuous easy iterations of the end to end is enabling better and better algorithms and results.



Result / Benefit

The powerful ML platform provided by Sparkflows enabled users in various geographic regions of the organization to work together on a powerful Self-Serve Advanced Data Science Platform.

A common training program, a common way to access and process data, access to 330+ processors and algorithms enabled seamless execution of the projects and taking the use cases to production.

The overall process and platform was continuously upgraded and fine tuned with new learning and needed features. It brought the various teams together and gave a big momentum to the building of Data Science use cases across the organization.



10X

More users enabled

50+

Faster use cases developed

10x

More Reusability



To know more visit us at

