Platform for Marketing Datasets Processing and Distribution

Industry
Marketing

Location
US

Technologies

Azure   BLOB Storage   Apache Spark

.NET   HD Insight   Data Lake   ClickHouse

CLIENT

In this case, the client provides IT services for its clients as well as perfects the use of Big Data to do its goals. The company targets small and medium-sized businesses (SMBs) that generate data volumes of up to hundreds of thousands every day and uses these volumes to create valuable marketing/advertising insights.

CHALLENGE

Our team joined the web application when it was already an MVP. The main task was to carry out the migration to the cloud, which took place on 2 levels.

Initially, we aimed to migrate from on-premises to the cloud, since hundreds of millions of data pass through the system every day, and it would be problematic to increase the amount of hardware for such conditions with a classical approach.

The second migration was from SQL-based services to technological ones related to big data. This is, from a database project to a cloud project since new modern technologies require modern solutions. These main tasks were our top priorities.

SOLUTION

The project is a classic Big Data solution aimed at simplifying the sorting of millions of data loads daily. The project follows an ETL process scenario for decision support suitable for data warehouses and Business Intelligence tools. Users can upload huge chunks of data loads from various sources and transform them to get some valuable marketing/advertisement insights.

A lot of our data transformations and aggregations are performed with Azure Databricks (Spark). There is a variety of transformations we have performed, ranging from standard joins and column updates to more sophisticated integrations with third parties. Since Azure Databricks (Spark) scales horizontally, we have gained a significant performance boost from moving to it.

Data input and primary data storage are typically handled by Azure Blobs. This feature offers simple setup and maintenance, seamless integration with other Azure services, and, most importantly, extremely fast read-write speeds when accessed from Azure Databricks.

Using the personal data, we helped gain valuable insights for the consuming company through our work building a modern Big Data project and Data Warehouse.

FEATURES:

1 Automated data extraction from various sources

2 Azure Databricks powered processing with 3rd party integrations

3 Extract, Transform, and Load (ETL) of marketing datasets to provide useful insights

4 Output results into different destinations such as Azure Blob, SFTP, SQL Database depending on client needs