

A MODERNIZED DATA ARCHITECTURE DRIVES BUSINESS VALUE IN BANKING

IMPACT

- Reduces daily data integration processes by 66%
- 4x faster querying time
- Helps bank expand its market position with analyses on current data
- Enables deployment of new applications Data Sources

Established in 1986, mBank S.A. is one of the leading Polish financial institutions and a member of Commerzbank Group (originally BRE – Export Development Bank). It is the fourth largest bank in Poland measured by total assets, servicing over 5.6 million retail clients and 26,500 corporate clients in Poland, and over 958,000 retail clients in Czech Republic and Slovakia. It offers retail, corporate, and investment banking, as well as other financial services.

Currently, the bank’s main focus is set on mobile and online banking development. Cloudera’s platform offers mBank integrated storing, processing, and analysis for all data. This will support new methods of data utilization that will enable the company to improve competitiveness, as well as allow for more efficient use of the company’s existing IT infrastructure.

Challenge

mBank has over 50 applications and systems, more than 5,000 workstations, and millions of end users producing data. Established in 1986, they had an underlying legacy IT system that was no longer capable of keeping pace with the ever-increasing volume of data. The existing enterprise data warehouse infrastructure was overwhelmed when it came to data science. Staff were accustomed to using data from the previous day because of the time required to integrate and to process the data. This made using data science a question of finding enough raw compute power.

Solution

mBank succeeded in building a modern IT infrastructure with Cloudera’s platform that has given it a handle on its big data. mBank now has the ability to integrate all its data and rapidly populate its data warehouses with more than 150 TB of data, with all data sources being accessed simultaneously. This means analyses and decisions can be made with the most current data, and queries are able to be completed much faster. These capabilities have helped mBank to further expand its market position. “We’ve observed a 4-8 times increase in speed for certain queries in comparison to our Oracle Data Warehouse,” said Artur Szymanski, responsible for IT Infrastructure, Data Warehousing and Data Science departments.

After the initial migration of its data to back in 2014, the extract, transform, and load (ETL) process was then modified to work hand-in-glove with Cloudera. In the meantime, the platform has become the central data hub at mBank and has improved performance enormously. Since then its data architecture has been constantly growing, with more and more business users using data in their daily work. This technology has allowed mBank to process 10x more data than Oracle, in less time.

Additionally, mBank has been building up its Data Science department to take advantage of the centralized Enterprise Data Hub. “Data Science runs comprehensive analyses, which make large demands on computing power and require fast access to big data,” said Szymanski.

66%

Reduction in daily data integration processes

"Having access to all the data means that we can deploy many applications for the first time. We can also save money, because much of the data will be available in Cloudera instead of the legacy data warehouse."

Artur Szymanski, responsible for IT Infrastructure, Data Warehousing and Data Science departments, mBank

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises. Learn more at cloudera.com

mBank has made use of Cloudera's stream processing technologies such as Kafka and Apache NiFi, which make it possible to continuously ingest data from source systems. This has opened up new possibilities for real-time data reporting and analysis including:

- The possibility of a quick reaction on the market, which it has been able to integrate into its marketing campaigns
- Optimal content management on its website, where sales campaigns are conducted
- Analysis of sales processes to ensure there are no bottlenecks
- Faster response to emerging complaint
- Real-time monitoring of the bank's financial liquidity

Using Spark streaming / Spark structured streaming, mBank has managed to create on-line cash flow alerting. It has built an early warning system of unusual situations, e.g. increased cash flow. This has been very important in the times of COVID-19.

Results

Cloudera's platform has become the central data hub at mBank—improving performance enormously. For example, mBank reduced its daily data integration processes by 66 percent.

"Instead of 24 hours, we only need eight hours now for the entire process, and we also believe that performance can be enhanced even further to make it possible to access the data in just four hours," said Szymanski. "Having access to all the data means that we can deploy many applications for the first time. We can also save money, because much of the data will be available in Cloudera instead of the legacy data warehouse."

"We chose to partner with Cloudera for a number of reasons. We were impressed with Cloudera Data Warehouse tools like Apache Impala [incubating] and Apache Spark, and Cloudera Manager makes cluster management much simpler than the alternatives, which entail 40 pages of administrative setup. Cloudera simply has the most well-developed enterprise platform."

mBank is currently using CDH 6.3.1. However, they are in discussion regarding plans for migration to CDP, which offers many possibilities and new components, in particular Hive 3.0, Apache Ranger, Apache Atlas and comprehensive cloud solutions.