EMAG: ONLINE MARKETPLACE INTEGRATES WEB DATA TO PROVIDE SUPERIOR CONNECTED SERVICES

IMPACT

- Datasets that used to take weeks to build are generated with queries that now run in hours or even minutes
- Free from time constraints, eMAG data scientists can experiment with data solutions and different indicators can be tried on the same day
- Time to insight down 75% – from over 40 days to 10 days – and still improving as more use cases are brought to the new infrastructure
- Data science has been moved out of the lab and into the production line
- More teams can use the platform for scalable storing and processing of data

eMAG is a fast-growing online marketplace and ecommerce leader in south-eastern Europe headquartered in Romania.

From its early days as a start-up, eMAG understood that in order to speed up its growth it needed enterprise-grade tools and support. The company also understood the importance of freeing up its data scientists so that they could spend most of their time focusing on results and solutions rather than building data sets.

Increasing web traffic

First and foremost, growing web traffic was proving problematic. From an architectural perspective, driving insight and value from the growing data estate was becoming increasingly unfeasible. Keen on applying more Artificial intelligence and Machine Learning to mature its analytics, eMAG found itself hindered by silos and a fragmented landscape.

Previously, eMAG's workflows and applications were based on open-source components which meant each element had its own lifecycle without integration. The company also had instances of shadow IT within the organization which were creating divergent processes, applications and forecasting methods – causing a further misalignment between business intelligence (BI) and the business team.

Another challenge for eMAG was that its data was kept in silos and could only be viewed in BI reports based on extracts. As such, anyone wanting to analyse the data in a way not covered by existing reports had to export it to Excel and then use VLOOKUP and various other processes to utilize it, making for a protracted and tedious process.

With bottlenecks throughout its BI deployment, and data and teams siloed, eMAG's BI department went on a mission to modernize its data and analytics landscape.

Led by the head of BI, and supported by data engineers, BI developers, data administrators and data stewards, the team began developing a centralized data lake process and platform in order to harness value from the data eMAG already had stored but was not using efficiently.

Complete platform migration in 48 hours

In January 2021, eMAG phased out its self-supported deployment, in favour of Cloudera Data Platform (CDP) Private Cloud Base. Over the course of just one weekend, all tools and data were migrated onto a single platform.

By switching to CDP, eMAG can perform different types of data processing to enrich the information it receives. It uses Kafka for streaming, Impala and Hive as SQL engine, HDFS and Kudu for storage, and Spark with Scala for data processing.

48

Hours to complete migration

"As a small BI department preferring to invest most of our resources in delivering business value rather than maintaining a big data platform, CDP has helped level the playing field."

– Laurentiu Matei, Business Intelligence Director, eMAG

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

Now, eMAG has a categorization system featuring real-time, near real-time and historical data, with different dashboard feeds for various types of data. With data being more accessible and easier to discover, queries can be run in hours, not weeks. Furthermore, the platform enables data to be entered into formats that are simpler to apply to use cases.

By leveraging CDP's ability to carry out seamless electronic data capture (EDC), the eMAG team can focus its attention on creating ad-hoc analysis and predictive machine learning applications.

Being able to focus on the development of value-adding applications has grown eMAG's business. Not only have the number of use cases significantly increased, but also the company's business value from its enterprise data platform.

Less time building data sets, more time for data science

Before introducing CDP, the most common complaint amongst eMAG data scientists was that most of their time was being spent carrying out the time-intensive task of building data sets, which left little time for actual data science. By streamlining its data processing using CDP to generate suitable data sets, eMAG data scientists have been able to move data science from a lab environment to a factory production line. With CDP, eMAG has gained time for insight – an issue that was previously a key pain point.

Using CDP to categorize large amounts of data more efficiently, eMAG is building complex customer cohorts quicker than ever. With an increased number of cohorts available, it can better discern consumer behaviour – like when and how often certain items are purchased. Access to this type of information allows eMAG to provide a higher calibre and better aligned service, as the company can now show customers better targeted product offerings. It also means marketing budgets are spent more cost-effectively.

"As a small BI department preferring to invest most of our resources in delivering business value rather than maintaining a big data platform, CDP has helped level the playing field," commented Laurentiu Matei, Business Intelligence Director, eMAG.

The impact of CDP can be felt throughout eMAG's operations having fundamentally changed the company's data and analytics landscape. By investing in the platform and all its capabilities, eMAG has enriched its entire data lifecycle – allowing its data scientists to move away from time-consuming data processing tasks and focus on achieving insights and results.

Additionally, eMAG adopted CDP to keep itself open to the option of using either private cloud, public cloud, or hybrid environments in the future. Through this, the organization is able to continually adjust its deployments of data and analytics to provide the optimal fit for changing markets as well as business imperatives.

CLOUDERA

Cloudera, Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA. cloudera.com