



eMAG accelerates time to insight and leaves no opportunities uncovered with Dremio

Challenges	Solution	Results
eMAG's online business had outgrown their IT infrastructure. The main issue was time to insight. Reports arrived too often too late to be relevant.	As the key component of eMAG's analytics infrastructure, the Dremio data lake engine offers speed, usability and security.	Once the BI team has fully automated the workflows behind the existing reports, this may result in a 50-75% reduction of generation time.
Data was siloed in all kinds of different sources. For a report data had to be moved from system to system in a time-consuming process.	With Dremio, data stored in multiple media like data warehouses, the marketplace and big data can now be combined for the first time.	Report creation takes only a few hours, instead of weeks and marketplace KPI analytics are down from 1 day to 20 minutes.
The cumbersome report creation practice also undermined timely ad-hoc analysis and data discovery.	Dremio enables ad hoc analysis through live data connections and direct queries for fast BI and dashboard analyses.	Data analysts can now create their own analyses and reports. Plus, Dremio offers analytics that have not been possible before.

CUSTOMER



emag.ro

GEO

Romania, eastern Europe

INDUSTRY

Retail, e-commerce

OBJECTIVES

eMAG wanted to be able to quickly and securely join data from multiple sources, accelerate queries for dashboarding, provide users with easier data discovery and enable ad-hoc data analysis through live connections and direct queries.

DATA ENVIRONMENT

- **Data Sources:** MySQL, SQL Server
- **Storage:** Amazon S3, Hadoop HDFS, Hive
- **Compute:** Dremio
- **Applications:** Qlik, Tableau, cnvrg.io, KNIME

Summary

eMAG is using Dremio as their query engine to make data from disparate data sources quickly available for analysis. Time to insight has been reduced dramatically allowing business users and data scientists to uncover business opportunities in near time. Furthermore, Dremio supports the BI team in its quest to establish a lean, responsive and user-friendly analytics infrastructure that ensures both excellent performance and security.

The Business

Romanian company founded in 2001, eMAG is a pioneer of the Romanian online retail with a presence in Romania, Bulgaria and Hungary. For over 20 years, the company has been constantly investing in services based on technologies developed in Romania, which help customers save time and money. Since its foundry in 2001, eMAG has developed a client-centered business model by constantly investing in offering clients a better shopping experience every day.

In 2011, eMAG launched eMAG Marketplace thus opening its platform to today's more than 36,000 retailers. Thus, customers gained access to an even wider range of products, while retailers could boost their businesses by addressing a larger consumers pool. Thanks to constant investments and innovative services and products, eMAG became one of growth drivers for the Romanian online retail sector.

eMAG Group announced revenues of 8.93 billion lei (1.8 billion euro) in 2020, of which eMAG revenues in Romania accounted for 5.42 billion lei (1.1 billion euro). Operations in Hungary and Bulgaria recorded a significant increase of 87%, reaching 2.1 billion lei (0.4 billion euro).

The Challenge

eMAG's quickly expanding online business had outgrown their IT infrastructure. The main issue was time to insight. Data was siloed in all kinds of different sources and was usually accessed via reports. For a report data had to be moved from system to system in a time-consuming process. If one of these complex ETL jobs failed, the delay got even bigger as the whole process had to be re-started from scratch. As a result, reports arrived too often too late to be relevant.

The cumbersome report creation practice also undermined timely ad-hoc analysis. Every time a business user or data scientist wanted to follow up on a result or zoom into details, the BI group had to create a new report. For this the developers either needed to assemble completely new datasets or use the data from existing reports and join them.

“Our teams were split by technology. They relied on reports but creating a report took so long that when it was finally available, the information was outdated. With Dremio, all users have access to the latest data and can act on opportunities in an instant.”

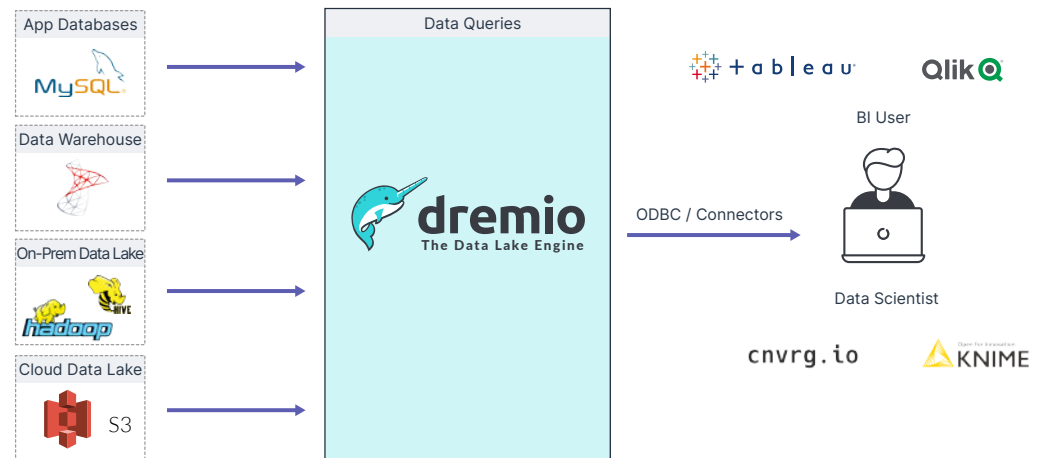
Laurentiu Matei
BI & ERP Director
eMAG

Though the whole process placed a heavy burden on resources, important reports were still not available in time. To stay on top of their increasing business requirements and to support their decision makers, eMAG wanted access to the latest information 24/7. But modernizing an infrastructure and introducing a completely new architecture takes time. eMAG needed a solution that could provide all users with the latest data now and in future. After online research for fast query engines and after attending some conferences eMAG discovered Dremio.

The Solution

Before eMAG finally chose Dremio, they tested a couple of other data warehouse and query engine solutions but none was able to meet their stringent criteria. As the primary goal was to accelerate time to insight, speed, performance and ease of use were the top priorities. Dremio convinced the eMAG team with its:

- Query speed
- Ability to accelerate queries
- Ability to document, catalogue and search data
- Usability
- Open architecture



Time to Insight

Report creation from

weeks

down to

hours

Report generation time
decreased by

50-75%

Marketplace analytics
from

1 day

down to

20 minutes

They also wanted a solution that would grow with their business and that they could rely on for the future. The dedication of the Dremio team had already impressed them during the selection process and they were looking forward to collaborating on the implementation of their new query engine.

The Dremio data lake engine has become the key component of eMAG's analytics infrastructure. The solution is implemented on premises and used for querying data from Apache Hadoop/HDFS and Hive, Microsoft SQL Server, MySQL and Amazon S3. It took the eMAG BI team and Dremio consultants only 4 months to set up 200 users. These were onboarded as soon as data was available to be queried.

For Tableau, Dremio serves as the only data source. That way firewall exceptions can be minimized. Data scientists rely on Dremio to quickly answer business questions and for fast prototyping. As prototyping datasets can be quickly performed directly from app databases, a later change of the data source to the data lake is a transparent and straightforward process.

Thanks to Dremio the BI team is now able to:

- Join, aggregate and search datasets from different databases
- Enable ad hoc analysis through live data connections and direct queries for fast BI and dashboard analyses
- Offer analysts and data scientists an easy way to perform data discovery
- Establish a new culture of data democratisation.

The Results

eMAG uses Dremio to quickly provide the different user personas with timely data for a variety of use cases. Technical use cases include automating data queries, cataloguing data and documenting data lineage. Business use cases include:

Reduced time to insight: Report creation from weeks down to hours

A typical use case in finance and marketing is a comparison between third party products and eMAG's own offerings. The creation of a side-by-side report of sales in a category (e.g. coffee and tea) for a certain period including information on campaign, product, seller, product name, price paid, quantity ordered, base price, voucher discounts and shipping fee used to take the BI team days, sometimes even weeks. With Dremio, they can create the report in a few hours. So for a question that is asked in the morning, the data feed will be available by noon.

“How I would describe Dremio to friends and coworkers?
Dremio – that’s where the data is.”

Laurentiu Matei
BI & ERP Director
eMAG

Reduced time to insight: Marketplace KPI analytics from 1 day down to 20 minutes

A MBR Marketplace analysis looks at the impact of inactivated, newly activated and reactivated products from one month to the next based on indicators like conversion rate, ASP, GMV, etc. The data for this analysis had to be exported from several Qlik reports involving five complex QVD files with data of seller products, active products, sales, conversion rate and other details.

All this data had to be assembled in several stages in QlikView. The whole process, from data extraction to processing, took about a day. The BI team has now developed a query that draws this data from Dremio in about 20 minutes.

Reduced time to insight: report generation down by 60% on average

Once the complete marketplace data is replicated in Dremio, the BI team will be able to fully automate the workflows behind the existing reports. This may result in a 50-75% reduction of generation time and will allow the specialists to invest the time saved in the development of additional reports or views.

Completely new insights

Dremio offers analytics that have not been possible before. Data stored in multiple media like data warehouses, the marketplace and big data can now be combined for the first time. It is also expected that data preparation time will come down significantly, as approximately 80% of the total effort for an analysis is spent on assembling and making data analysis-ready.

Freedom and self-service for accelerated decision making

Dremio provides data analysts in all departments with a new capability. They can now create their own analyses and reports by deriving new datasets from existing ones without waiting for the support of the BI team. This makes exploring data and looking for opportunities even faster.

Lightening the load of the BI team and adding value

As users are growing more and independent, the demands on the BI team have decreased. Its members can now spend more time on delivering analytics and adding business value than on moving and assembling data. They also gained the flexibility to outsource urgent requests like the creation of a dashboard to an external consultant by quickly creating a dataset in Dremio and making it available to Tableau



Quantity adds quality

With Dremio, more data than ever before can be queried, joined and aggregated at speed. It is even possible to search for a specific dataset. The more data can be explored the better are the chances that the users can discover hidden relationships they had not been aware of before. They gain a 360° view of the business and get all the information they require to make decisions that directly impact the bottom line.

Business analytics: no programming or IT skills required

The users in the customer care department particularly like the ease with which they can use their data. Dremio provides them with access to different data sources without the need to understand their actual implementation. SQL abstracts all irrelevant details for them and they can focus on extracting data using a unified framework. Available optimizations assure high compatibility with real-time plotting tools such as Tableau.

Future Plans

As the implementation of Dremio at eMAG has been a great success, the BI team aims to onboard the whole company, either directly or through BI apps, planning to reach the 1,000 user mark in a year. They are aware though that switching technologies and democratizing data only work when combined with a change of culture. To facilitate adoption, they intend to use Dremio to make transitions like moving workloads to the cloud or data from data warehouse to data lake transparent to end users.

For their own benefit, the BI experts are determined to make the most of the documentation feature in Dremio and to use the wiki to document certified datasets. They also want to explore Dremio Cloud Services as soon as enough data has been moved to the cloud.

ABOUT DREMIO

Dremio reimagines the cloud data lake to deliver faster time to analytics by eliminating the need for expensive proprietary systems and providing data warehouse functionality on data lake storage. Customers can run mission-critical BI workloads directly on the data lake, without needing to copy and move data into proprietary data warehouses or create cubes/aggregation tables/BI extracts. In addition, Dremio's semantic layer provides easy, self-service access for data consumers, and flexibility and control for data architects. Dremio delivers the world's first no-copy architecture, drastically simplifying the data architecture and enabling data democratization.

Deploy Dremio

[Learn more at dremio.com](https://dremio.com)

CONTACT SALES

contact@dremio.com

Dremio and the Narwhal logo are registered trademarks or trademarks of Dremio, Inc. in the United States and other countries. Other brand names mentioned herein are for identification purposes only and may be trademarks of their respective holder(s). © 2022 Dremio, Inc. All rights reserved.

