

FELDM

Data Integration and Architecture

Future-proof with a modernized data landscape



An online sports and lifestyle company had grown its user base to more than 42 million people in over 160 countries worldwide since its founding in June 2013. Due to the rapid growth in new users and the associated data streams (events), the data warehouse used to date had reached its limits. The new architecture required had to be designed for the future and equipped to handle further growth in data volumes.

The client and the FELDM team determined that the new system had to be able to process more than 100 events per second. It needed to be scalable in terms of data volume and also allow further development in width (new functions). In addition, a reliable infrastructure had to be introduced to enable IT staff to easily detect and fix bugs by reprocessing without subsequent data loss.

The following main criteria were defined for the new data infrastructure:

- Near real-time data processing
- Designed for more than 100 events per second
- Scalable, reliable and easy to maintain

"The FELDM team was an excellent collaborator who's willingness to go the extra mile was much appreciated. They integrated into our team seamlessly and have strong technical skills. We would work with them again definitely."

THOMAS YOPES
Team Lead Data Engineering
and Analytics at Freeletics



We proposed an array of state-of-the-art technologies, including AWS EC2, DynamoDB, API Gateway, Serverless, Lambda, SQS, S3, Linux, Python and Spark.

Near real-time analyses for better business decisions

With the newly developed architecture, analyses can now be performed in near real time. The success of campaigns can be measured faster. More flexible control options are available. The connection and processing of further data sources is easier, thanks to the forward-looking design of the infrastructure. The client has already carried out such modifications several times. The S3 Data Lake enables the internal Data Analytics Team to perform evaluations of raw data that cannot be covered by the analytics tool Amplitude. We accompanied and supported the Data Analytics Team with first sample analyses in PySpark.

With Amplitude, teams from product development, finance, etc. have the possibility to create generic, but complex, analyses and simple visualizations on their own. These analyses are used today for planning, controlling and product development.