

Legacy software modernization for market research company

Domain: **Market Research** Budget: **\$11.000/month** Duration: **since 2017**

Company

Our customer is a global measurement and data analytics company that provides the most complete and trusted view of consumers and markets worldwide.

For more than 90 years it has provided data and analytics based on scientific rigor and innovation, continually developing new ways to answer the most important questions facing the media, advertising, retail, and fast-moving consumer goods industries.

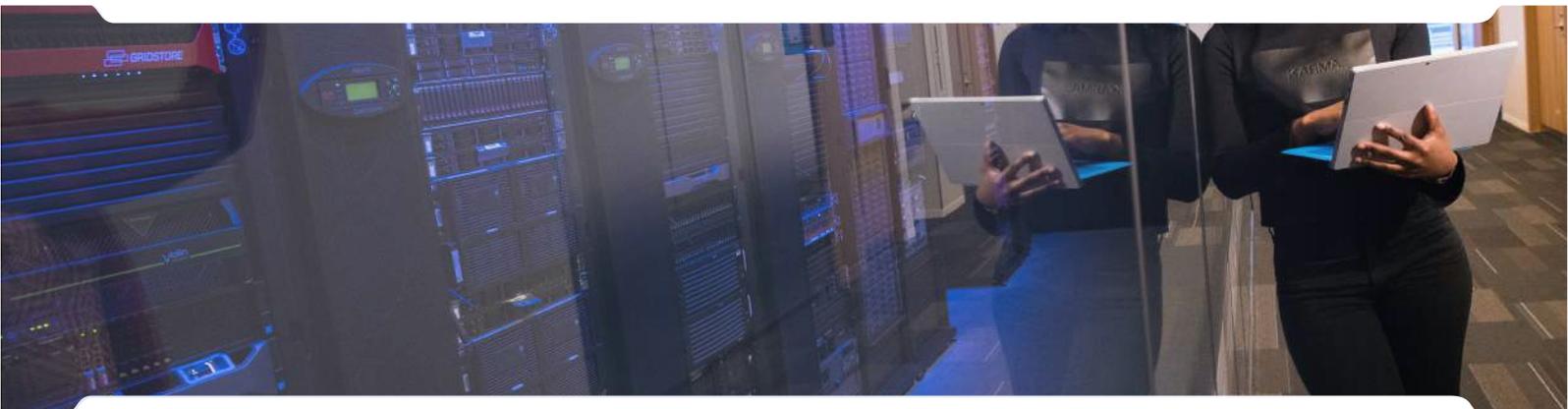
It operates in more than 100 countries and is ranked number 1 among top Market Research companies in the USA.



Challenge

Succeeding in today's market is tough. Innovation - that's where breakthrough opportunities reside. But an innovative approach is impossible without using up-to-date technologies.

Our customer embraces innovation just as we do. To work properly they need to be on the cusp. That's why it has been decided **to migrate from legacy technologies to modern cloud infrastructure.**



Tasks

- **Migration to Spark.** Spark is used at a wide range of organizations for large-scale data processing. It allows running workloads faster, writing applications quickly in Java, Scala, Python, R, and SQL, combining SQL, streaming and complex analytics. Spark runs on Hadoop, Apache Mesos, Kubernetes, standalone, or in the cloud. It can access diverse data sources.
- **Validation of business logic,** which is critical to keep things running smoothly. Publishing data is actually pretty easy. There are lots of ways to do that. While securing and organizing your data is an important bit that requires some thought.
- **Data correctness verification.** Accurate data is the backbone of any database. Rigorous, objective and transparent verification processes are vital to establishing and maintaining high-quality data.
- **Performance improvement** in general and **reducing data processing time in particular,** which, by the way, speak for themselves.

Solution

The solution implies taking the data prepared by the company's different teams from the cloud storage (S3) and executing the project code to message, consolidate and filter data to perform further calculations. Code execution results are recorded in a file that can be used either for the company's other projects or for making reports.

The solution includes:

- Orchestration of all jobs (tasks);
- Migration to AWS cloud services;
- Refactoring of the code to handle increased data volume;
- The possibility of sharing the transformed data with different teams and using it within various company's projects.

Key features



Data verification
(data qualitatively and clearly coincides with what is expected)



Each job (task) gives the expected result



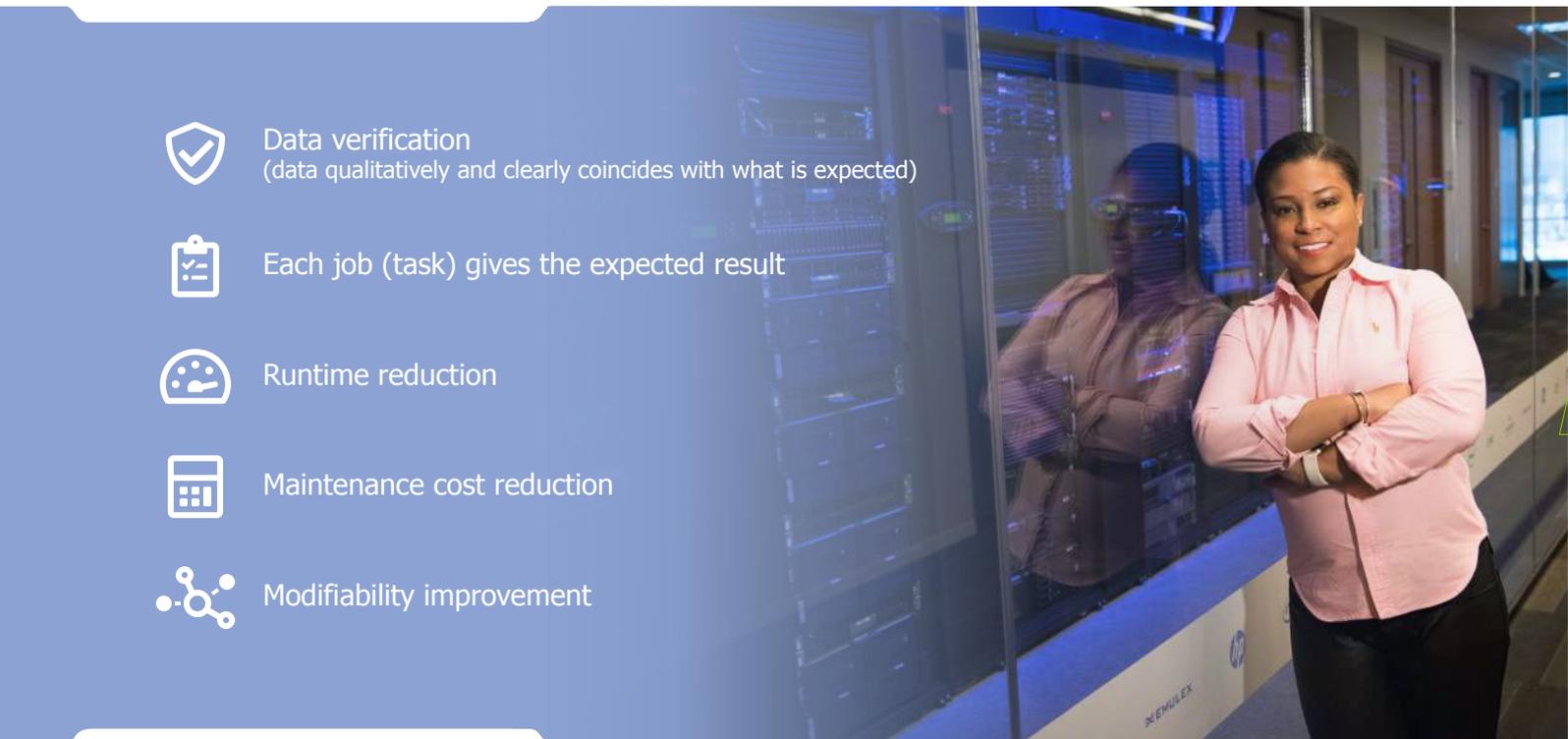
Runtime reduction



Maintenance cost reduction



Modifiability improvement



Value

The implemented technologies provide our customer with qualitative detailed information in a short time and help to deliver expertly crafted reports to the end-users to remain competitive. On the basis of these reports the advertising policy is built, whose success depends critically on data accuracy.

Technologies

- Scala
- Spark
- HDFS
- AWS EMR
- AWS Lambda
- AWS RDS
- AWS SQS
- Grafana
- S3
- Kubernetes
- Airflow

Legacy software modernization

is the key to adapt, grow and remain competitive for any organization. In close cooperation with your company, we will work out the best strategy for this process!

[Learn more](#) 