caser seguros

SUCCESS STORY

PUE

- Caser has opted for a Big Data platform with Cloudera, with the aim of boosting its Data Centric model.
- PUE collaborated by providing technological solutions that allow more efficient data analysis in less time.

About Caser

Caser is a Spanish insurance group with more than 75 years' experience. The insurance company has a wide range of solutions in all branches: home, vehicle, health, and life, among others. Its activity is characterized by clear customer orientation, quality of service, and the professionalism of its distribution lines. Caser also operates in other markets such as the elderly, hospitals, maintenance and assistance services, and financial advice. Caser is part of the Helvetia Group. The group currently has a staff of more than 5,100 people.

The challenge

For Caser, data analytics is essential to make the right decisions. For this reason, it had a Data Centric model based on a Data Warehouse.

However, decisions could not always be made with up-to-date data as this was not available. Processing and query times needed to be reduced in order to have real data available when needed.

It also needed a consolidated, reasoned overview of the data, as well as needing to improve key aspects: scalability and security in massive processes, new capabilities in data governance, traceability, lineage, cataloging, access controls, and auditing for proper data management.

The solution

A 3-year strategy was defined, with a roadmap of first steps that have already been implemented. Firstly, the design and implementation of a Data Lake in Cloudera to form part of the staging area as a central part of the company's Data Warehouse, to unify and considerably optimize the processing and storage of data.

Replacing the relational database not only allows the user to access and exploit the data quickly. As the data is all in one place, it can be used for reporting and analytics, and integrated into production applications.

The next steps within this roadmap involve the consolidation of the Big Data platform as the central pillar for the storage and processing of data for analytics. This will increase the development of use cases and features offered by Cloudera, and boost the use of Machine Learning in the near future.

> "PUE's expertise in Big Data is noteworthy, but also its commitment and collaboration"
>
> **Hugo González**
> Head of the Information Structures and Advanced Analytics area | Caser

Applied technologies

Big Data Platform. Specifically:

- Impala
- Kudu
- Spark

Hugo González
Head of the Information Structures and Advanced Analytics area on Caser

Adrián Buenaventura
Presentador

Watch the full interview here

PUE's services

PUE collaborated by providing technological solutions throughout the project, designing and implementing a data architecture capable of improving the load performance of the Data Lake to enable more efficient data analysis in less time:

- Conversion of traditional programming languages, such as PL/SQL, to Spark processes with Scala. This has reduced data analysis process times from hours to minutes, gaining in speed and resulting in a business diagnosis in almost real time.
- Quick, efficient integration of the ETL tool with the Data Lake.
- Cluster configuration and security policies.
- Integration of PowerCenter processes into the Cloudera Platform.

Overall results

Superior streaming **runtime performance**:

- 85% reduction of runtimes in complete WorkFlows compared to the initial times.
- Improved write times in migrated tables in process (ODBC vs FTP+Spark), with an 80% reduction compared to the initial times.

Improved **data quality**:

- By keeping a detailed record of data input and output in validations, errors in the processes migrated during the project have been corrected.

Perfecting of final **volume management**:

- A detailed record of data input and output, following the complete traceability of the data, has made it possible to validate load volumes with respect to production processes, as incidents can be resolved in the validation cycle.

Review of **PWC developments**:

- As PowerCenter's own developments had to be migrated, the logic applied in conjunction with Caser developers has been improved.

Increased **process scalability**:

- More cluster resources, allowing extra processing power when necessary.

Better **data histories**:

- Implementation of histories designed in the most efficient way for optimal execution of processes and queries.

> "The challenge for us is to start having a platform available that not only has the data and can provide it, but that does so much faster and in a more agile manner, and for the data to be as accessible as possible throughout the company.
>
> Data is a key asset, especially in a financial company like ours.
>
> Facilitating access to data, facilitating the incorporation of new information, which can be exploited in any context in the simplest possible way is a challenge, but that is what we are looking for"
>
> **Hugo González**
> Head of the Information Structures and Advanced Analytics area | Caser