

## Análisis de datos más eficiente en menos tiempo

### Acerca de Caser

Caser es un grupo asegurador español con más de 75 años de experiencia.

La compañía aseguradora cuenta con una amplia oferta de soluciones en todos los ramos: Hogar, Automóvil, Salud o Vida, entre otros. Su actividad se caracteriza por una clara orientación al cliente, la calidad de servicio y la profesionalidad de sus líneas de distribución. Además Caser opera en otros mercados como tercera edad, hospitales, servicios de mantenimiento y asistencia o asesoramiento financiero.

Caser forma parte del Grupo Helvetia. Actualmente el grupo cuenta con un staff de más de 11.600 personas.

**Caser ha apostado por una plataforma de Big Data con Cloudera, con el objetivo de potenciar su modelo Data Centric.**

**PUE ha colaborado aportando soluciones tecnológicas que permiten un análisis de datos más eficiente en menos tiempo.**

“ El reto para nosotros es empezar a tener una plataforma, no solamente que tenga los datos y los pueda proveer, sino que lo hagamos de una manera mucho más rápida, más ágil y que el dato sea lo más accesible posible en toda la compañía.

El dato es un activo clave, y más en una compañía financiera como la nuestra.

Facilitar el acceso al dato, facilitar la incorporación de nueva información, que se pueda explotar en cualquier contexto de la manera más sencilla posible es un reto, pero es lo que vamos buscando. ”

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

### El reto

Para Caser, la analítica de datos es fundamental para la toma de decisiones acertadas. Para ello contaban con un modelo Data Centric basado en un Data Warehouse.

Sin embargo, las decisiones no siempre podían tomarse con datos actualizados, porque no se disponía de ellos. Era necesario reducir los tiempos de procesamiento y consulta para contar con datos reales en el momento oportuno.

También requerían de una visión de los datos consolidada y razonada, así como mejorar aspectos clave: la escalabilidad y la seguridad en procesos masivos, aportar nuevas capacidades en el gobierno del dato, la trazabilidad, el linaje, la catalogación, los controles de acceso y de auditoría para tener una correcta gestión del dato.

### La solución

Se ha definido una estrategia a 3 años vista, con un roadmap cuyos primeros pasos ya se han ejecutado. En primer lugar, diseño e implementación de un Data Lake en Cloudera para formar parte del staging area como pieza central del Data Warehouse de la compañía, que unifica y optimiza considerablemente el procesamiento y almacenamiento de los datos.

Sustituir la base de datos relacional no solo permite que el usuario acceda y explote el dato rápidamente. Al encontrarse los datos en un único sitio, se ha conseguido que puedan utilizarse en reporting y analítica, así como integrarlos en aplicaciones de producción.

Los siguientes pasos dentro de este roadmap están relacionados con consolidar la plataforma de Big Data como pieza central para el almacenamiento y procesamiento de datos para la analítica. De esta manera se incrementará el desarrollo de casos de uso y funcionalidades que ofrece Cloudera, y se impulsará el uso de Machine Learning en un futuro próximo.

### Tecnologías aplicadas

**Big Data Platform** • Impala  
En concreto: • Kudu  
• Spark

### Resultados globales

#### Rendimiento superior de tiempos de ejecución en flujo

- Reducción de los tiempos de ejecución en WorkFlows completos al 85% sobre los tiempos iniciales:
- Tiempos de escritura sobre las tablas migradas en proceso (ODBC vs FTP+Spark) mejorados, con una reducción al 80% sobre los tiempos iniciales.

#### Mejoría en calidad del dato

- Al llevar un control detallado de la entrada y salida del dato en validaciones, se han corregido errores en los procesos migrados durante el proyecto.

#### Perfeccionamiento del control de volumen final

- Un control en detalle de la entrada y salida de datos, siguiendo la trazabilidad del dato completa, ha permitido validar volúmenes de carga con respecto a procesos en producción, pudiendo solventar incidencias en el ciclo de validación.

#### Revisión de desarrollos en PWC

- Al tener que migrar desarrollos propios de PowerCenter se ha mejorado la lógica aplicada en conjunto con desarrolladores de Caser.

#### Aumento de la escalabilidad del proceso

- Más recursos del clúster, consiguiendo ampliar la potencia de procesamiento en caso de ser necesario.

#### Mejor historificación del dato

- Más recursos del clúster, consiguiendo ampliar la potencia de procesamiento en caso de ser necesario.

“ Destaco la experiencia de PUE en Big Data, pero también el compromiso y la colaboración ”

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



[Ver la entrevista](#)

### Servicios de PUE

PUE ha colaborado aportando soluciones tecnológicas durante todo el proyecto, diseñando e implementando una arquitectura de datos capaz de mejorar el rendimiento de carga del Data Lake para permitir un análisis de datos más eficiente en menos tiempo:

- Conversión de lenguajes de programación tradicionales, como PL/SQL, a procesos Spark con Scala. Esto ha supuesto la reducción del tiempo de los procesos de análisis de datos de horas a minutos, ganando en agilidad y disponiendo de un diagnóstico de negocio casi en tiempo real.
- Integración de la herramienta de ETL con el Data Lake de manera rápida y eficiente.
- Configuración de clústeres y políticas de seguridad.
- Integración de procesos de PowerCenter a Cloudera Platform.

#### MADRID

Arregui y Aruej, 25-27 | 08007 Madrid  
T. 91 443 51 21

#### BARCELONA

Diagonal, 98-100 | 08019 Barcelona  
T. 93 206 02 49

**Any data problem?**

Queremos ser tu partner de confianza en tus proyectos de datos

Contacta con nosotros en [sales@pue.es](mailto:sales@pue.es)