USING NEW TECHNOLOGY TO UNLOCK THE VALUE IN YOUR DATA

Sagacity

In today's competitive commercial landscape there is a continuing need to gain advantage by the interpretation and strategic insight into the data we hold and generate. As the volume of data produced within an organisation increases, selecting the right technology to process data efficiently is key.

Our experience has demonstrated that corporate organisations utilise multiple databases, a myriad of spreadsheets and shared drives grown organically over time. How does an organisation make full use of the data it owns?

The traditional relational database management systems (RDBMS) provide insight into core business data; the data is structured, organised, well understood and often hosted in a multi-user environment. However; the proliferation of unstructured data such as social media postings, internet search data and financial transaction information requires a somewhat different approach.

Here at Sagacity we like a good challenge to reset our thinking and try something new, so our technology teams have been actively embracing the use of Hadoop and Spark to harness the power of unstructured data for our clients.

The Theory

Hadoop

Hadoop has evolved as a platform to store, organise and process extremely large and often unstructured data sets using the Hadoop Distributed File System storing data across a multi-server estate, known as a cluster of nodes. This lends itself very well to the cloud-based storage paradigm.

Data can be loaded into the Hadoop framework and stored en masse. This is commonly referred to as a Data Lake. The Hadoop framework makes it easy to store these unconstrained and unrelated data sets . It is then down to the analyst (Data Scientist) to process the data to achieve the analytical goals.

Hadoop uses a MapReduce algorithm to process data, splitting the base data across clusters, performing the processing on each sub dataset, and then assembling the outputs once all processing is complete. The Hadoop platform provides vast scalability to process ever increasing data demands rapidly and efficiently.

Spark

Spark provides an engine that can read files stored in a Hadoop Distributed File System, databases, AWS S3 and many more, in a highly parallelised manner. It applies algorithms to split the task into many small units that can be run in parallel and sent to each worker on a cluster or local machine, allowing vast amounts of data to be processed quickly.

One of the key benefits of using Spark is that data, once processed, can be cached in memory across a cluster allowing multiple reports and data insights to be generated quickly.

Harnessing the power of Hadoop and Spark

At Sagacity we are harnessing the power of Hadoop and Spark to enhance and accelerate our proprietary data matching and data parsing algorithms within our core products. The ability to quickly scan, manipulate and cross match data from multiple data sources has significantly increased the performance and complexity of our Customer Data service offerings through our proprietary software.

Our data credentials in numbers

- -

6 billion data records processed over 3 days

- -

1.7 billion data records output via our Customer Value Model

- -

4 million customer records cleansed for migration

- -

2 million records processed monthly and shared between different client systems

Process changes translated to **c.£12million** OPEX saving

- -

17.5 million records run through 35 data sources to deliver churn insights

- -

Recent advancements on our platform

Hadoop/Spark within our Single Customer View platform:
The implementation of Hadoop/Spark has accelerated the speed and complexity of cleansing of each individual data element of a customer. The core identifiers of name, address, date of birth and contact details are now available using enhanced cleansing algorithms. The distributed processing approach provided with Spark has enabled our matching engine to be used on increasingly larger and more unstructured data sets to bring a further degree of completeness to our customer data.

We have vastly increased the processing capabilities of the SCV service, ensuring our clients gain a complete view of each customer they engage.

The implementation of a multi-dimensional Customer Value Model for pan-European telecommunications corporation:
The Hadoop/Spark architecture allows the platform to handle increased complexity of real life customer events to be incorporated into our model. The inclusion of customer journey events such as active Win-Backs or renewed customer agreements provides further deep insight into the customer base. The CVM also supports the efficient calculation and analysis of forecasting and reporting.

Enhancement of the Sagacity Integrated Data Security System:
The Spark programming architecture has enhanced our ability to collect and interrogate individual user activity across our data landscape. The multiple unstructured user activity files are scanned and analysed to identify any unauthorised access to our systems.

At Sagacity we are always striving to ensure our systems are secure and accessed only by authorised individuals.

We'd love to discuss how we can help you leverage the power of your data using Hadoop and spakr expertise to transform your data to produce efficient and insightful analytics. Get in touch today to find out more.

We'd love to discuss how we can help you leverage the power of your data using Hadoop and spakr expertise to transform your data to produce efficient and insightful analytics. Get in touch today to find out more.