Digital transformation starts with data quality

The Hartree® Centre worked with Altus Group to develop an automated data extraction, validation and transformation pipeline to improve data quality and save valuable time and staff effort.

Challenge

In an increasingly data driven market, businesses need confidence in the quality of their data to make effective decisions. New companies may take a clean-sheet approach when it comes to collecting the data they need, but existing businesses often have legacy systems containing data of mixed quality and value, and any low-quality data lurking within them can jeopardise any decisions made on the whole dataset. Altus Group wanted to target and correct issues across their data more effectively – to improve the quality of their existing information and to ensure new data arriving from external sources was of acceptable quality before it was loaded into their systems.

Approach

The Hartree Centre team worked closely with Altus Group to investigate the existing quality of the data systems in scope and understand issues faced, then went on to propose and develop a new data extraction, validation and transformation pipeline. Rather than hard-coding business logic and risk limiting the shelf-life of a potential solution, we created a flexible framework that can be reconfigured according to application. By delivering a prototype quickly and refining through iterative testing, feedback and development cycles, the solution was effectively tailored to address the challenge. Altus Group technical staff were simultaneously trained to develop their own configurations and were comfortable taking ownership of the project outputs going forwards to maintain the system in production.

Benefits

Altus Group are now continually validating their internal data and generating weekly reports highlighting relevant issues to data owners for review. This lightweight system can run at laptop scale and is capable of screening hundreds of thousands of records within minutes, freeing up valuable staff time and effort, with resulting data quality improvements underpinning core business activities. The flexibility and extensibility built in to the project from the outset has enabled the Altus Group team to apply the new system to new datasets. In a single day, the team established a new data ingestion and validation workflow, effectively automating a previously manual data ingestion process.

Image credit: Pexels

> " Altus Group have benefited from immediate efficiency gains using the Hartree Centre's solution, one task that previously took several months can now be completed in hours. "

William Taylor
Altus Group

> **Our experience of working with the Hartree Centre has been outstanding, providing a robust and scalable solution that Altus Group can embed into our data processes delivering fast and more accurate outcomes leading to significant commercial benefits.**

Nick Wright
Altus Group

At a glance

- Flexible system to extract, validate and transform data from several sources into high quality outputs

- Intelligent data extraction, resistant to errors such as typos and synonyms in column labels

- Validation of internal systems data with minimal load to production servers business growth

- Validation of external data upon arrival, to catch issues before they impact automated decision-making

- Built on freely licensed, open-source software – no subscription costs and free performance and stability upgrades from the community

Who we are

- 60+ computational scientists and technologists

- World-leading supercomputing and AI infrastructure

- Bespoke small teams built around your project

- Tailored business development support

- Access to our network of industry, academic and technology partners

What we do

- Boost productivity and enhance innovation for industry

- Big data analytics and artificial intelligence (AI)

- High performance computing and quantum simulation

- Training and skills development

- Insights into future technologies

Our impact on UK industry and society

The Hartree Centre was created by UK Government to transform industry by accelerating the adoption of high performance computing (HPC), big data analytics and artificial intelligence (AI) technologies. We play a key role in realising UK Government's Industrial Strategy by stimulating applied digital research and innovation, creating value for the organisations we work with and generating economic and societal impact for the UK.

The Science and Technology Facilities Council (STFC) Hartree Centre is part of UK Research and Innovation.

hartree.stfc.ac.uk   **@HartreeCentre**   **STFC Hartree Centre**   **hartree@stfc.ac.uk**