LEVERAGING DATA TO PROMOTE AND SHARE KNOWLEDGE ACROSS THE ENTERPRISE

IMPACT

- Created a single repository for all data in the company, both organized and raw
- Simplified operations, with a strong optimization of transformation jobs
- Reduced data processing times by 400% with monthly data mart uploads reduced from 6 days to 1.5 days
- Enabled knowledge dissemination within the company and a more data-driven culture

INAIL, Italy's National Institute for Insurance against Accidents at Work, is a non-profit public body that manages compulsory insurance against accidents at work and occupational diseases. Founded in 1933, it promotes a culture of prevention in order to reduce the number of accidents and incidence of occupational diseases. In addition to insuring workers who carry out high risk activities, it provides economic and health benefits to reduce and prevent accidents in the workplace.

Around 650,000 accidents at work are reported every year in Italy. With about 10,000 employees distributed throughout the country, INAIL manages a portfolio of about 3.2 million companies, protecting workers and carrying out scientific research aimed at improving safety at work.

Managing data in an efficient and business-friendly way

"Today there are two main challenges impacting the world of data and information management," explains Patrizio Galasso, Head of the Data and Information Asset Management Solutions Office, INAIL. "There is exponential and continuous growth in the amount of information produced every day, the management of which presents significant difficulties. Secondly, business users are increasingly demanding richer and more sophisticated analysis of data. Where previously descriptive analysis was sufficient, now predictive analysis is required, with waiting times becoming shorter and shorter, leading to an increasingly tighter time-to-market."

To speed up processes and increase flexibility, INAIL launched a complete overhaul of the IT subsystem dedicated to business intelligence. The starting point was a traditional data management infrastructure based on a legacy data warehouse equipped with data marts. Over time, the team built various dashboards to answer specific business questions, which had resulted in a fairly typical infrastructure for an organization of this size. However, the INAIL team eventually found itself having to work with an inflexible architecture, which required deep revisions - and the related costs - every time a change was requested. The presence of so many different dashboards also led to a duplication of data, resulting in limited storage space and inconsistencies in information resulting from data readings performed by different departments. This combined with the numerous software licenses required and the operational intricacies of managing such a complex infrastructure further increased costs.

Infrastructure overhaul with Big Data in mind

To solve these critical issues, INAIL partnered with Cloudera to deploy an enterprise data platform. Although the various data sources were not altered, the core systems were completely redesigned, featuring three main components: a data lake that gathers and hosts data from INAIL's sources; a data hub that organizes datasets based on facts and dimensions according to the intended use, and a data lab hosting the tools. The newly deployed system was dubbed "Ianua" ("door" in Latin), as it represents the entrance to INAIL's knowledge. On top of this system, a unified data portal was created to access and consume the data, equipped with tools to perform advanced analysis.

"The strength of the platform is undoubtedly the fact that we created a single point of data storage.

400

Percentage reduction in processing time

"The strength of such a platform is undoubtedly the fact that we have created a single point of data storage. With our Cloudera enterprise data platform, we have a structure to serve the needs of our data scientists, which allows us to use new technologies that were not compatible with the platform we had previously. We have also gained a lot in terms of speed of data preparation."

Patrizio Galasso, Head of the Data and Information Asset Management Solutions Office, INAIL

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

We have created a homogeneous infrastructure that contains both the data in the data hub and the raw data in the data lake, which makes it very easy to integrate, even dynamically," said Galasso. "Today, we have a structure to serve the needs of our data scientists, the data lab, which allows us to use new technologies that were not previously compatible with the platform we had. We have also gained a lot in terms of speed of data preparation."

Concrete and measurable results, both for IT and business users

Through adopting this new data architecture, positive results were achieved across the business. From an IT point of view, the number of transformation jobs has been streamlined, from 2000 down to 800: this was possible because in the new environment it is possible to create big queries and integrate activities that were previously divided among several jobs into a single processing step. Processing times have been reduced: the monthly loading of data marts has improved 400% (from six days to a day and a half). In addition, there has been an improvement in the way information has been structured, with the creation of a single environment where data is stored, netting further benefits in terms of reduction in maintenance time and also in the amount of disk space occupied.

From a business point of view, the elimination of silos has allowed the team to increase both the quality and consistency of the data. Moving to Cloudera's single platform has made it possible to effectively cross-analyze data previously hosted on different systems. For example, Inail can cross-reference the history of accidents at companies that are requesting a reduction of the average rate for prevention, enabling a data-driven decision. With a unified and immediate view of the data available to its users, the enterprise data platform supports better business decisions.

INAIL is currently growing its data analytics capabilities to drive additional value and insights from its datasets. The project guidelines in this area are threefold: enrichment of information assets, strengthening of the infrastructure and improvement of access and knowledge management.

A project that looks to the future

The Ianua platform is gradually being enriched with domains not previously managed by the old data warehouse. Data silos continue to be broken down, reducing the number of dashboards. In parallel, new projects based on machine learning and artificial intelligence are being launched. From an infrastructure perspective, INAIL is migrating to the Cloudera Data Platform Private Cloud Base, which delivers multi-tenant functionalities to meet specific needs and deliver gains in flexibility. The ultimate goal is the improvement of access and knowledge management, with the aim of making life easier for business users who want to interface with the information system without having to deal with too many technicalities.

The end user will have a Google-like screen at their disposal to carry out data analysis using natural language processing (NLP) search. Deploying what are now industry standard technologies, such as NLP, semantic engines and ontology codes, the information requested can be extrapolated: if that information is present on an existing dashboard, the user will be redirected to this application component. Otherwise, the user will be guided in the query they need to make by accessing Ianua directly using an ontological graph, and terminologies and relations known to them.

Knowledge sharing

"From the point of view of the business, we are achieving a number of particularly significant benefits, such as documenting and keeping knowledge up to date within the organization, using bespoke analytics engines and frameworks. We are spreading and uniforming the culture and semantics of data and the meaning of the terms that are used; and of course we are simplifying access to knowledge so that it can actually become the heritage of the entire organization, and not just of IT," concludes Patrizio Galasso.

CLOUDERA