FORGE2KNOW
DATAPLATFORM

Leonardo Secure Data
Management Solution

LEONARDO

Governments, businesses, and organizations recognize data as a precious resource capable of offering insights, guiding decision-making, and enhancing the overall performance of business processes.

Leonardo designed Forge2Know suite to address the challenges of integrated Big Data analytics and Artificial Intelligence governance.

Structured with 2 inter-operating but independent products, Forge2Know integrates modern open-source technologies and proprietary modules, designed to ensure security, scalability and interoperability, environments.

Forge2Know.DataPlatform
Provide tools to speed and simplify data governance, data ingestion, and big data processing and presentation.
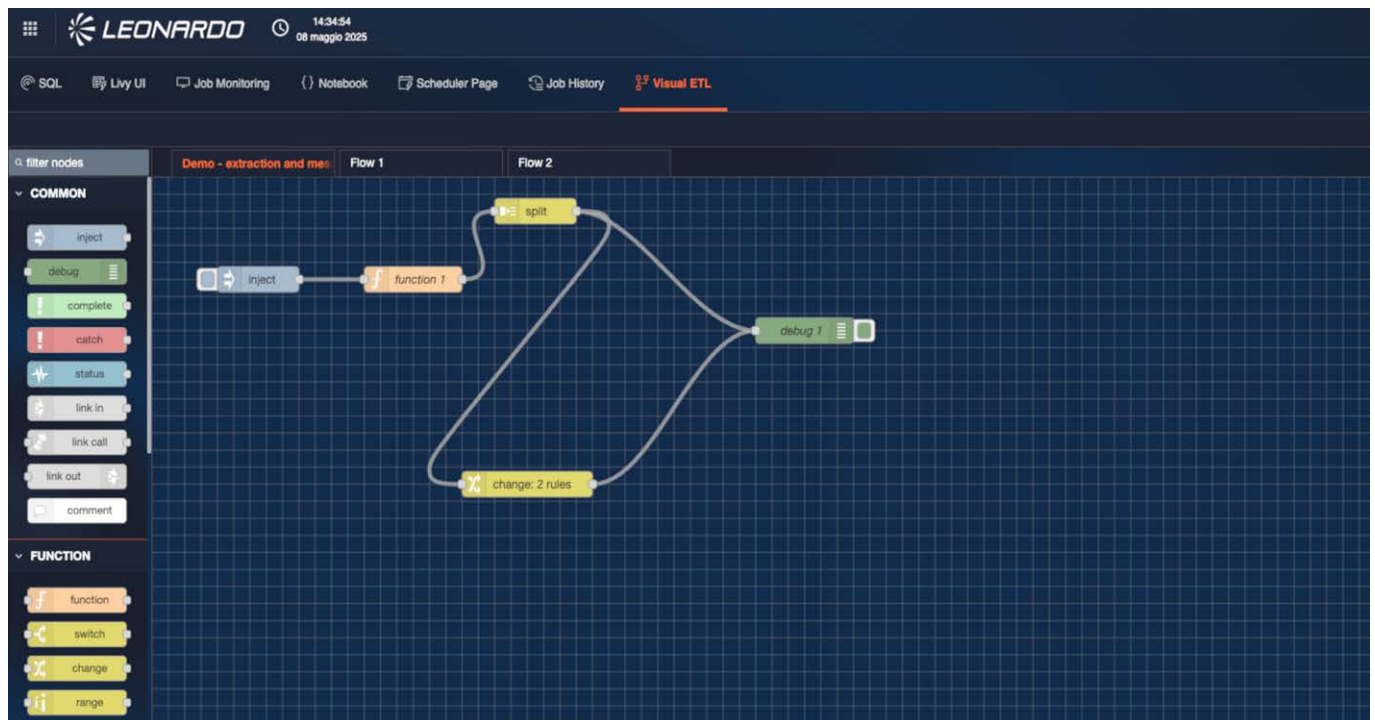
Forge2Know.AIEngine
A platform for the governance of AI models in terms of: Responsible AI, Repository (Model/Dataset), Serving, Active Monitoring, Experience Collector and Benchmark Evaluation.

The two Forge2Know products can be used as stand-alone solutions, each fully operational. By integrating them, you get a complete and scalable platform that enables you to transform data into strategic value, optimizing governance, processing and adoption of artificial intelligence with high standards of security and interoperability.

The combination, with the presence of a Crawling Platform and secure AI Models, however, allows the implementation of a powerful, flexible and optimized solution for data management and valorization that gives rise to the Forge2Know.DataIntelligence product that completes the suite.

A POWERFUL DATA PLATFORM

Forge2Know.DataPlatform is a set of integrated tools for managing, valorizing and enriching large volumes of data, with particular attention to the highest standards of security, scalability and interoperability. The platform integrates advanced open-source technologies related to Big Data, combined with best practices and experience gained in the field by Leonardo in various sectors.

KEY FUNCTIONALITIES

DATA LAKE a ready-to-use multi tenant secure platform for ingesting, storing and retrieving structured, semi-structured and unstructured data of any size.

DATA PROCESSING providing functions for real time and batch elaboration leveraging in-memory execution engine and horizontal scalability, parallel processing and also through vector db. Publish-subscribe capabilities, event management, scheduling and workload management as well as visual ETL (Extract Transform Load) and a Secure Query Engine further empower data processing capabilities.

DATA SECURITY AND GOVERNANCE implementing a single, secure and centralized reference point for data control. By leveraging search and discovery tools and connectors to extract metadata from any data source, it simplifies data lineage, data protection, analysis, and pipeline management, as well as accelerates ETL processes.

MODEL BUILD Set of tools and best practices aimed at developing and maintaining Machine Learning (ML) models reliably and efficiently.

PRESENTATION A module for visualizing and interacting with data through graphical data quality interfaces and business intelligence tools: tables, charts and other types of representations.

In addition Forge2Know.DataPlatform can provide results to higher level Leonardo or third party platform feeding intelligence or analysis processes.

AN INTEGRATED SET OF OPEN TOOLS

Forge2Know.Dataplatform is an end-to-end digital solution, designed to automate and streamline data governance, data ingest and big data processing, interacting seamlessly with third-party applications.

The platform has been developed using the best open-source and commercial components available on the market that represent a technological reference in the Big Data context.

This brings advantages to the platform, making it:

→ Cutting-edge and aligned with best practices thanks to the constant collective innovation carried out by the open source community;

→ Continuously evolving and able to evolve according to the latest technological trends;

→ Characterized by a high degree of interoperability with other systems and software platforms, with benefits in terms of versatility and applicability.

Below is a list of the main software modules offered by the platform:

→ DATA LAKE: Provides a ready-to-use platform with all the necessary features for developers, data scientists, and analysts to easily store data of all sizes, shapes, and speeds.

→ BATCH/REAL TIME PROCESSING: Provides a ready-to-use platform for developing batch and streaming processes based on an in-memory execution engine, horizontal scalability and parallel processing.

→ EVENT MESSAGE: Offers a ready-to-use platform for developing real-time data applications and pipelines, serving as a message broker with publish/subscribe functionalities.

→ DATA GOVERNANCE: Provides a ready-to-use platform that offers a unique, secure, and centralized reference point for data control. Leveraging search and detection tools and connectors, this approach simplifies data protection, analysis, and pipeline management by extracting metadata from any data source and accelerates ETL processes.

→ TEXT SEARCH ENGINE: It provides a distributed search and analysis engine, through which both full-text and semantic searches can be performed, and it also offers a vector store.

→ VISUAL ETL: Offers a graphical (Flow-Based) tool for creating ETL pipelines that integrates processors capable of offloading computation to the Batch and Streaming Processing module. By leveraging tools and connectors provided by open-source community, it simplifies the creation of the ETL pipeline.

→ DAL: Provides a module for exposing a data access service through REST APIs (https) by using SQL queries on a multitude of data sources (internal and external), allowing for granular authorization management.

→ NOTEBOOKS: provides a robust and integrated environment for development activities based on technologies and languages of the Big Data context.

→ MODEL BUILD: is a set of practices aimed at reliably and efficiently developing and maintaining Machine Learning (ML) models.

→ PRESENTATION: set of tools for representing results using graphs and tables.

The product among other things allows to:

→ Be provided both in Cloud and On Premise
→ Guarantee confidentiality and integrity of data
→ Create/define standards and best practices to adopt
→ Guarantee national sovereignty (platform developed by integrating open-source components)
→ Reduce/Cancel third-party product licensing costs
→ Implement the concept of Data Centric Security

A CYBER SECURE PLATFORM

Forge2Know.DataPlatform leverages several security techniques in order to provide a reliable and resilient tool for data management and analysis.

→ Data at rest encryption: Protection of data when stored on disks, achieved through their encryption.

→ Data In Transit Protection: Confidentiality protection of data when in transit, with particular attention to communication via public networks.

→ Inter Modular Communication Protection: Communication between elements composing the Forge2Know.DataPlatform modules through private and secure channels.

→ High Reliability: Inherently achieved (Orchestrator) or individual modules configured in High Availability (HA).

→ Static and Dynamic Code Security: analysis of all software artifacts including open-source modules.

ONE PLATFORM MULTIPLE USERS

Forge2Know.DataPlatform is a useful tool for several class of users dealing with data providing an integrated set of functions that help data professional to perform their work.

Data Engineer may use the platform to perform access and data management, jobs management as well as event visualization and analysis.

Data Ops can take advantage of the function provided to manage and orchestrate data such as monitoring, security, access control, log and audit, management.

Data Steward may benefit of the function provided especially data governance and metadata management, to ensure and monitor the quality of data used by the organization.

Data Security focuses on protecting data by defining proper **access** control policies, preventing unauthorized use, loss, or alteration through technical and organizational measures.

Data Scientist can leverage platform capabilities to manage data and find useful business insight acting with notebooks, IA models and keeping track in a structured way of their experiments and activities.

Data Consumer can take advantage of presentation capabilities as well as query engine and AI models to perform Business Intelligence operations and other data intensive actions.

POSSIBLE USE CASES

Data Lake housing
Forge2Know.DataPlatform can be used for storing and querying large datasets for analytical and reporting purposes. This can be beneficial for businesses that need to analyze large amounts of data to make informed decisions

Data Governance and Security
the platform can be used to implement data governance policies and procedures. This can help businesses ensure the quality, consistency, usability, security, and availability of their data.

Data Integration
data from various sources and provide a unified view of the data. This can be beneficial for businesses that deal with data from multiple sources and need a single platform to manage it all.

Data Analytics
the platform can be used to perform complex data analytics and generate insights from large volumes of data. This can help businesses understand trends, patterns, and derive meaningful insights from their data.

Session Transactions and Data Storage
Forge2Know.DataPlatform can handle session data present in enormous amounts of business settings. It ensures data is maintained up to date and fresh, which is crucial for businesses that need real-time data for decision-making.
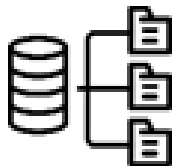
Model Training and Gen AI
The platform can be used as an operating environment where it will be possible to enhance the managed data for the training and fine-tuning of ML models and to support the delivery, through AI Engine, of Gen AI services based on RAG systems.

A CASE STUDY: HISTORICAL DATA ANALYSIS FOR CRITICAL INFRASTRUCTURE AND LOGISTICS

A critical infrastructure can benefit significantly from a platform enabling data valorization for:

→ Operational situational awareness
→ Infrastructure KPI evaluation
→ Early warning management and support for operational and security-related decision-making
→ Predictive and preventive measures for security, energy sustainability, and maintenance
→ Simulation capabilities for planning and optimization

Data from vehicles, sensors, personnel, facilities, systems, and networks is acquired in real time or batch mode—depending on availability—and ingested into a centralized data lake.

The platform's Data Messaging functionality supports integration with live systems such as PSIM, BMS, EDR, NMS, and other complex sources. Security and access policies are governed by "need-to-know" principles, enabling full traceability and auditability of operations performed on the data.

Visual ETL and notebook modules enable the execution of advanced scripts to transform and normalize data for correlation and integration purposes.

Custom dashboards can be built to monitor infrastructure status from multiple perspectives, detect trends, and highlight anomalies. For example, in the context of energy consumption, tailored dashboards can provide a time-based overview of supply conditions, usage patterns, and irregularities.

Infrastructure data can also support predictive maintenance, fault analysis for equipment and systems, and energy modeling based on occupancy levels and weather conditions. Furthermore, dynamic asset data—such as vehicles—can be analyzed to track performance by route, detect anomalies, and evaluate usage-related statistics.

Forge2Know

KEY VALUES

Operational Flexibility

"Turnkey" design (Source integration, Workflow and document access permissions, protocolling, HW management)

High Integration

Big data and AI in a single platform , built-in tools for batch, streaming, real-time, etl processing with the possibility of structured query languages

Data centered security

Data lifecycle easy support for major security and confidentiality policies GDPR, CCPA, HIPA control (data access, data masking, data access)

Deployment everywhere

Ability to operate on prem, on cloud and in hybrid Winfrastructures
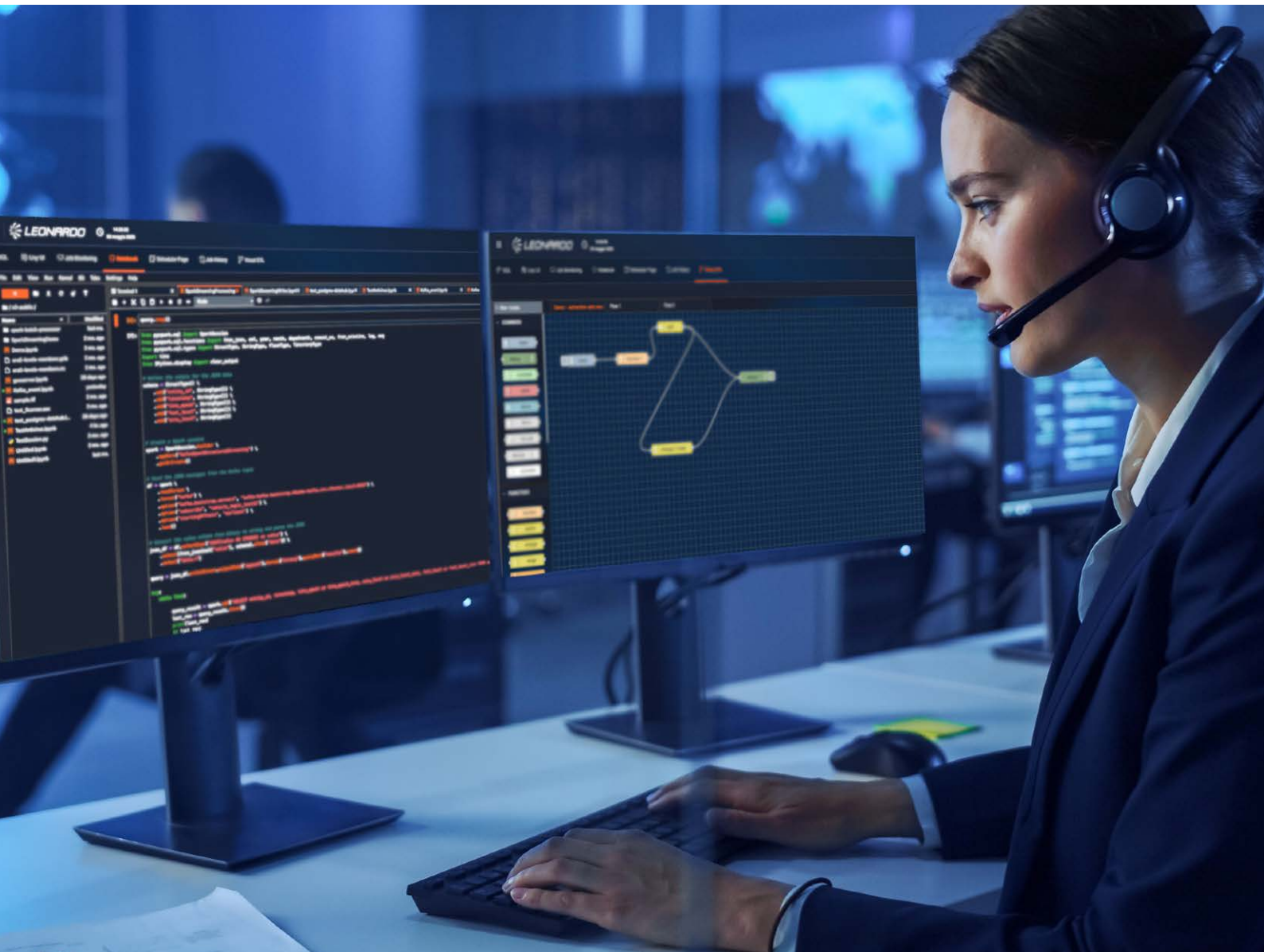
For more information:
cyberandsecurity@leonardo.com

Leonardo Cyber & Security Solutions Division
Via R. Pieragostini, 80 - Genova 16151 - Italy

leonardo.com