



CYBER & SECURITY SOLUTIONS

FORGE2KNOW DATAPLATFORM

LA SOLUZIONE LEONARDO PER
LA SECURE DATA MANAGEMENT



Governi, aziende e organizzazioni riconoscono i dati come una risorsa preziosa in grado di offrire conoscenza, guidare il processo decisionale e migliorare le prestazioni complessive dei processi aziendali.

Leonardo ha progettato la suite Forge2Know per affrontare le sfide dell'analisi integrata dei Big Data e della governance dell'Intelligenza Artificiale.

Strutturata con 2 prodotti interoperabili ma indipendenti, Forge2Know integra moderne tecnologie open source e moduli proprietari, progettati per garantire sicurezza, scalabilità e interoperabilità in ambienti diversi.

Forge2Know.DataPlatform

Una piattaforma costituita da strumenti per velocizzare e semplificare la governance dei dati, acquisizione, elaborazione e presentazione di big data.

Forge2Know.ALEngine

Una piattaforma per la governance dei modelli di AI in termini di: Responsible AI, Repository (Modello/Dataset), Distribuzione, Monitoraggio Attivo, Raccolta delle esperienze e Valutazione dei Benchmark.

I due prodotti Forge2Know possono essere utilizzati come soluzioni stand-alone, ciascuna pienamente operativa. Integrandoli, si ottiene una piattaforma completa e scalabile che consente di trasformare i dati in valore strategico, ottimizzando la governance, l'elaborazione e l'adozione dell'intelligenza artificiale con elevati standard di sicurezza e interoperabilità.

Entrambi i prodotti sono funzionalmente completi e possono essere utilizzati come strumenti stand-alone. La combinazione, con la presenza di una piattaforma di crawling e di modelli di AI sicuri, consente tuttavia l'implementazione di una soluzione potente, flessibile e ottimizzata per la gestione e la valorizzazione dei dati che dà origine al prodotto Forge2Know.DataIntelligence che completa la suite.



UNA POTENTE PIATTAFORMA DATI

Forge2Know.DataPlatform è un insieme di strumenti integrati per la gestione, la valorizzazione e l'arricchimento di grandi volumi di dati, con particolare attenzione ai più elevati standard di sicurezza, scalabilità e interoperabilità. La piattaforma integra tecnologie open source avanzate legate ai Big Data, combinate con le best practice e l'esperienza maturata sul campo da Leonardo in diversi settori.

FUNZIONALITA' CHIAVE

DATA LAKE acquisizione, archiviazione e recupero di dati strutturati, semi-strutturati e non strutturati di qualsiasi dimensione

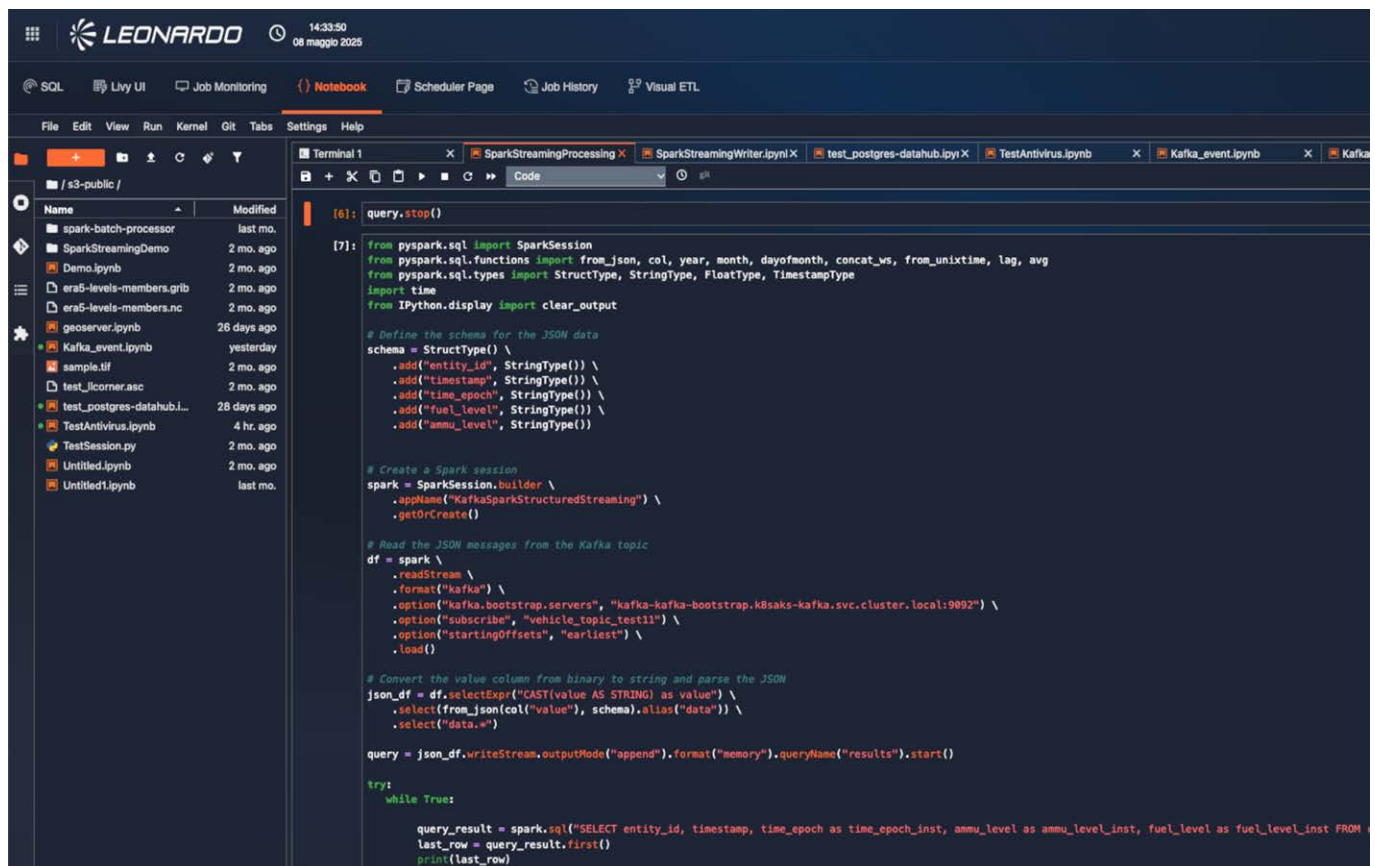
DATA PROCESSING elaborazione in tempo reale e in batch sfruttando il motore di esecuzione in-memory e la scalabilità orizzontale; elaborazione parallela e anche attraverso database vettoriali. Funzionalità di publishsubscribe, gestione eventi, pianificazione e gestione dei carichi di lavoro, nonché ETL visuale (Extract Transform Load) e un motore di query sicuro potenziano le capacità di elaborazione.

DATA SECURITY AND GOVERNANCE punto di riferimento unico, sicuro e centralizzato per il controllo dei dati. Sfruttando strumenti e connettori di ricerca e discovery per estrarre metadati da qualsiasi fonte dati, semplifica la genealogia dei dati, la protezione dei dati, l'analisi e la gestione della pipeline, oltre ad accelerare i processi ETL

MODEL BUILD strumenti e best practice volti a sviluppare e mantenere modelli di Machine Learning (ML) in modo affidabile ed efficiente.

PRESENTATION visualizzare e interagire con i dati attraverso interfacce e strumenti di business intelligence: tabelle, grafici e altri tipi di rappresentazioni.

Inoltre, Forge2Know.DataPlatform può fornire risultati a piattaforme Leonardo o di terze parti di livello superiore che alimentano processi di intelligence o analisi.



The screenshot displays the Leonardo Data Platform interface. At the top, the Leonardo logo and a clock showing 14:33:50 on 08 maggio 2025 are visible. Below the header, there are tabs for SQL, Livy UI, Job Monitoring, Notebook, Scheduler Page, Job History, and Visual ETL. The main workspace is divided into a file explorer on the left and a code editor on the right. The file explorer shows a directory structure with files like spark-batch-processor, SparkStreamingDemo, Demo.ipynb, era5-levels-members.grib, era5-levels-members.nc, geoserver.ipynb, Kafka_event.ipynb, sample.tif, test_lcorner.asc, test_postgres-databu..., TestAntivirus.ipynb, TestSession.py, Untitled.ipynb, and Untitled1.ipynb. The code editor shows a Jupyter Notebook with the following Python code:

```
(6): query.stop()

[7]: from pyspark.sql import SparkSession
from pyspark.sql.functions import from_json, col, year, month, dayofmonth, concat_ws, from_unixtime, lag, avg
from pyspark.sql.types import StructType, StringType, FloatType, TimestampType
import time
from IPython.display import clear_output

# Define the schema for the JSON data
schema = StructType() \
    .add("entity_id", StringType()) \
    .add("timestamp", StringType()) \
    .add("time_epoch", StringType()) \
    .add("fuel_level", StringType()) \
    .add("ammu_level", StringType())

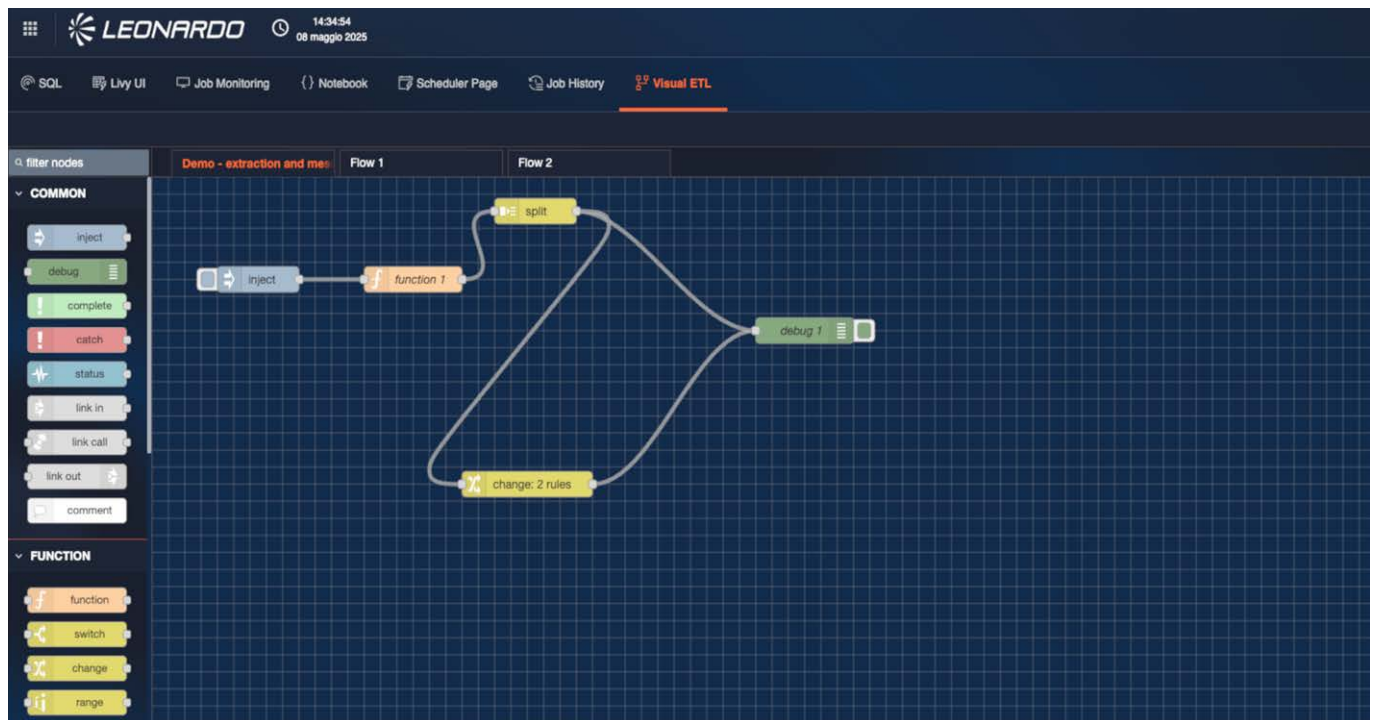
# Create a Spark session
spark = SparkSession.builder \
    .appName("KafkaSparkStructuredStreaming") \
    .getOrCreate()

# Read the JSON messages from the Kafka topic
df = spark \
    .readStream \
    .format("kafka") \
    .option("kafka.bootstrap.servers", "kafka-kafka-bootstrap.k8saks-kafka.svc.cluster.local:9092") \
    .option("subscribe", "vehicle_topic_test1") \
    .option("startingOffsets", "earliest") \
    .load()

# Convert the value column from binary to string and parse the JSON
json_df = df.selectExpr("CAST(value AS STRING) as value") \
    .select(from_json(col("value"), schema).alias("data")) \
    .select("data.*")

query = json_df.writeStream.outputMode("append").format("memory").queryName("results").start()

try:
    while True:
        query_result = spark.sql("SELECT entity_id, timestamp, time_epoch as time_epoch_inst, ammu_level as ammu_level_inst, fuel_level as fuel_level_inst FROM results")
        last_row = query_result.first()
        print(last_row)
        if last_row is None:
            break
```

UN SET INTEGRATO DI OPEN TOOL

Forge2Know.DataPlatform è una soluzione digitale end-to-end, progettata per automatizzare e semplificare la governance dei dati, l'acquisizione dei dati e l'elaborazione dei Big Data, interagendo perfettamente con applicazioni di terze parti.

La piattaforma è stata sviluppata utilizzando i migliori componenti open source e commerciali disponibili sul mercato, che rappresentano un punto di riferimento tecnologico nel contesto dei Big Data.

Questo apporta vantaggi alla piattaforma rendendola:

- all'avanguardia e allineata alle best practice grazie alla costante innovazione collettiva portata avanti dalla community open source;
- in continua evoluzione e in grado di evolversi secondo i più recenti trend tecnologici;
- caratterizzata da un elevato grado di interoperabilità con altri sistemi e piattaforme software, con vantaggi in termini di versatilità e applicabilità.

Di seguito è riportato l'elenco dei principali moduli software offerti dalla piattaforma:

- DATA LAKE: Offre una piattaforma pronta all'uso con tutte le funzionalità necessarie a sviluppatori, data scientist e analisti per archiviare facilmente dati di qualsiasi dimensione, formato e velocità
- BATCH/REAL TIME PROCESSING: Offre una piattaforma pronta all'uso per lo sviluppo di processi batch e streaming basati su un motore di esecuzione in-memory, scalabilità orizzontale ed elaborazione parallela.
- EVENT MESSAGE: Offre una piattaforma pronta all'uso per lo sviluppo di applicazioni e pipeline di dati in tempo reale, fungendo da broker di messaggi con funzionalità di publish/subscribe.
- DATA GOVERNANCE: Fornisce una piattaforma pronta all'uso che offre un punto di riferimento unico, sicuro e centralizzato per il controllo dei dati. Sfruttando strumenti e connettori di ricerca e rilevamento, questo approccio semplifica la protezione dei dati, l'analisi e la gestione delle pipeline estraendo metadati da qualsiasi fonte dati e accelerando i processi ETL.
- TEXT SEARCH ENGINE: Fornisce un motore di ricerca e analisi distribuito, attraverso il quale è possibile eseguire ricerche sia full-text che semantiche, e offre anche un archivio vettoriale.
- VISUAL ETL: Offre uno strumento grafico (FlowBased) per la creazione di pipeline ETL che integra processori in grado di delegare il calcolo al modulo di elaborazione batch e streaming. Sfruttando strumenti e connettori forniti dalla comunità open source, semplifica la creazione della pipeline ETL.
- DAL: Fornisce un modulo per l'esposizione di un servizio di accesso ai dati tramite API REST (https) utilizzando query SQL su una moltitudine di sorgenti dati (interne ed esterne), consentendo una gestione granulare delle autorizzazioni.
- NOTEBOOKS: Fornisce un ambiente robusto e integrato per attività di sviluppo basate su tecnologie e linguaggi del contesto Big Data.

→ **MODEL BUILD**: un insieme di pratiche volte a sviluppare e mantenere in modo affidabile ed efficiente modelli di Machine Learning (ML).

→ **PRESENTATION**: insieme di strumenti per la rappresentazione dei risultati tramite grafici e tabelle.

Il prodotto tra le altre cose consente di:

- essere fornito sia in Cloud che On Premise
- garantire la riservatezza e l'integrità dei dati
- creare/definire standard e best practice da adottare
- garantire la sovranità nazionale (piattaforma sviluppata integrando componenti open source)
- ridurre/annullare i costi di licenza di prodotti di terze parti implementare il concetto di sicurezza Data Centric

UNA PIATTAFORMA, PIÙ UTENTI

Forge2Know.DataPlatform è uno strumento utile per diverse categorie di utenti che gestiscono dati, offrendo un set integrato di funzioni che aiutano i professionisti dei dati a svolgere il loro lavoro.

I **Data Engineer** possono utilizzare la piattaforma per gestire gli accessi e i dati, gestire i processi, nonché visualizzare e analizzare gli eventi.

I **Data Ops** possono sfruttare le funzioni fornite per gestire e orchestrare i dati, come monitoraggio, sicurezza, controllo degli accessi, log e audit, e gestione.

I **Data Steward** possono trarre vantaggio dalle funzioni fornite, in particolare dalla governance dei dati e dalla gestione dei metadati, per garantire e monitorare la qualità dei dati utilizzati dall'organizzazione.

La **Data Security** si concentra sulla protezione dei dati definendo adeguate policy di controllo degli accessi, prevenendo l'uso non autorizzato, la perdita o l'alterazione attraverso misure tecniche e organizzative.

I **Data Scientist** possono sfruttare le funzionalità della piattaforma per gestire i dati e ottenere utili insight aziendali, utilizzando notebook, modelli di AI e tenendo traccia in modo strutturato dei propri esperimenti e attività.

I **Data Consumer** possono sfruttare le funzionalità di presentazione, nonché il motore di query e i modelli di AI per eseguire operazioni di Business Intelligence e altre azioni che richiedono un utilizzo intensivo dei dati.

UNA PIATTAFORMA CYBER SICURA

Forge2Know.DataPlatform sfrutta diverse tecniche di sicurezza per fornire uno strumento affidabile e resiliente per la gestione e l'analisi dei dati.

- Crittografia dei dati a riposo: protezione dei datimarchiviati su disco, ottenuta tramite la crittografia.
- Protezione dei dati in transito: protezione della riservatezza dei dati in transito, con particolare attenzione alle comunicazioni tramite reti pubbliche.
- Protezione delle comunicazioni inter-modulari: comunicazione tra gli elementi che compongono i moduli Forge2Know.DataPlatform tramite canali privati e sicuri.
- Elevata affidabilità: ottenuta intrinsecamente (Orchestratore) o tramite singoli moduli configurati in alta disponibilità (HA).
- Sicurezza del codice statica e dinamica: analisi di tutti gli artefatti software, inclusi i moduli open source.



POSSIBILI CASI D'USO

Data Lake housing

Forge2Know.DataPlatform può essere utilizzato per archiviare e interrogare grandi set di dati a fini analitici e di reporting. Questo può essere utile per le aziende che necessitano di analizzare grandi quantità di dati per prendere decisioni informate.



Data Governance and Security

La piattaforma può essere utilizzata per implementare policy e procedure di governance dei dati. Questo può aiutare le aziende a garantire la qualità, l'coerenza, l'usabilità, la sicurezza e la disponibilità dei propri dati.



Data Integration

Dati provenienti da diverse fonti e fornire una visione unificata dei dati. Questo può essere vantaggioso per le aziende che gestiscono dati provenienti da più fonti e necessitano di un'unica piattaforma per gestirli tutti.



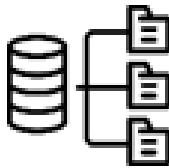
Data Analytics

La piattaforma può essere utilizzata per eseguire analisi complesse dei dati e generare insight da grandi volumi di dati. Questo può aiutare le aziende a comprendere tendenze, modelli e a ricavare insight significativi dai propri dati.



Transazioni di sessione e archiviazione dei dati:

Forge2Know.DataPlatform è in grado di gestire i dati di sessione presenti in un'enorme quantità di contesti aziendali. Garantisce che i dati siano sempre aggiornati e aggiornati, il che è fondamentale per le aziende che necessitano di dati in tempo reale per il processo decisionale.



Model Training e Gen AI

La piattaforma può essere utilizzata come ambiente operativo in cui sarà possibile migliorare i dati gestiti per l'addestramento e la messa a punto dei modelli ML e supportare l'erogazione, tramite AI Engine, di servizi



UN CASO DI STUDIO: ANALISI DI DATI STORICI PER INFRASTRUTTURE CRITICHE E LOGISTICA

Un'infrastruttura critica può trarre significativi vantaggi da una piattaforma che consenta la valorizzazione dei dati per:

- Consapevolezza operativa della situazione
- Valutazione dei KPI dell'infrastruttura
- Gestione degli allarmi precoci e supporto per i processi decisionali operativi e di sicurezza
- Misure predittive e preventive per la sicurezza, la sostenibilità energetica e la manutenzione
- Funzionalità di simulazione per la pianificazione e l'ottimizzazione

I dati provenienti da veicoli, sensori, personale, impianti, sistemi e reti vengono acquisiti in tempo reale o in modalità batch, a seconda della disponibilità, e inseriti in un data lake centralizzato.

La funzionalità di Data Messaging della piattaforma supporta l'integrazione con sistemi live come PSIM, BMS, EDR, NMS e altre fonti complesse. Le policy di sicurezza e accesso sono regolate da principi "need-to-know", consentendo la completa tracciabilità e verificabilità delle operazioni eseguite sui dati.

I moduli ETL visivi e notebook consentono l'esecuzione di script avanzati per trasformare e normalizzare i dati a fini di correlazione e integrazione.

È possibile creare dashboard personalizzate per monitorare lo stato dell'infrastruttura da più prospettive, rilevare tendenze ed evidenziare anomalie.

Ad esempio, nel contesto del consumo energetico, dashboard personalizzate possono fornire una panoramica temporale delle condizioni di fornitura, dei modelli di utilizzo ed delle irregolarità.

I dati infrastrutturali possono anche supportare la manutenzione predittiva, l'analisi dei guasti di apparecchiature e sistemi e la modellazione energetica basata sui livelli di occupazione e sulle condizioni meteorologiche.

Inoltre, i dati dinamici delle risorse, come i veicoli, possono essere analizzati per monitorare le prestazioni in base al percorso, rilevare anomalie e valutare statistiche relative all'utilizzo.

VALORI CHIAVE

Flessibilità operativa

Progettazione “chiavi in mano” (integrazione delle fonti, flussi di lavoro e autorizzazioni di accesso ai documenti, protocolli, gestione hardware)

Elevata integrazione

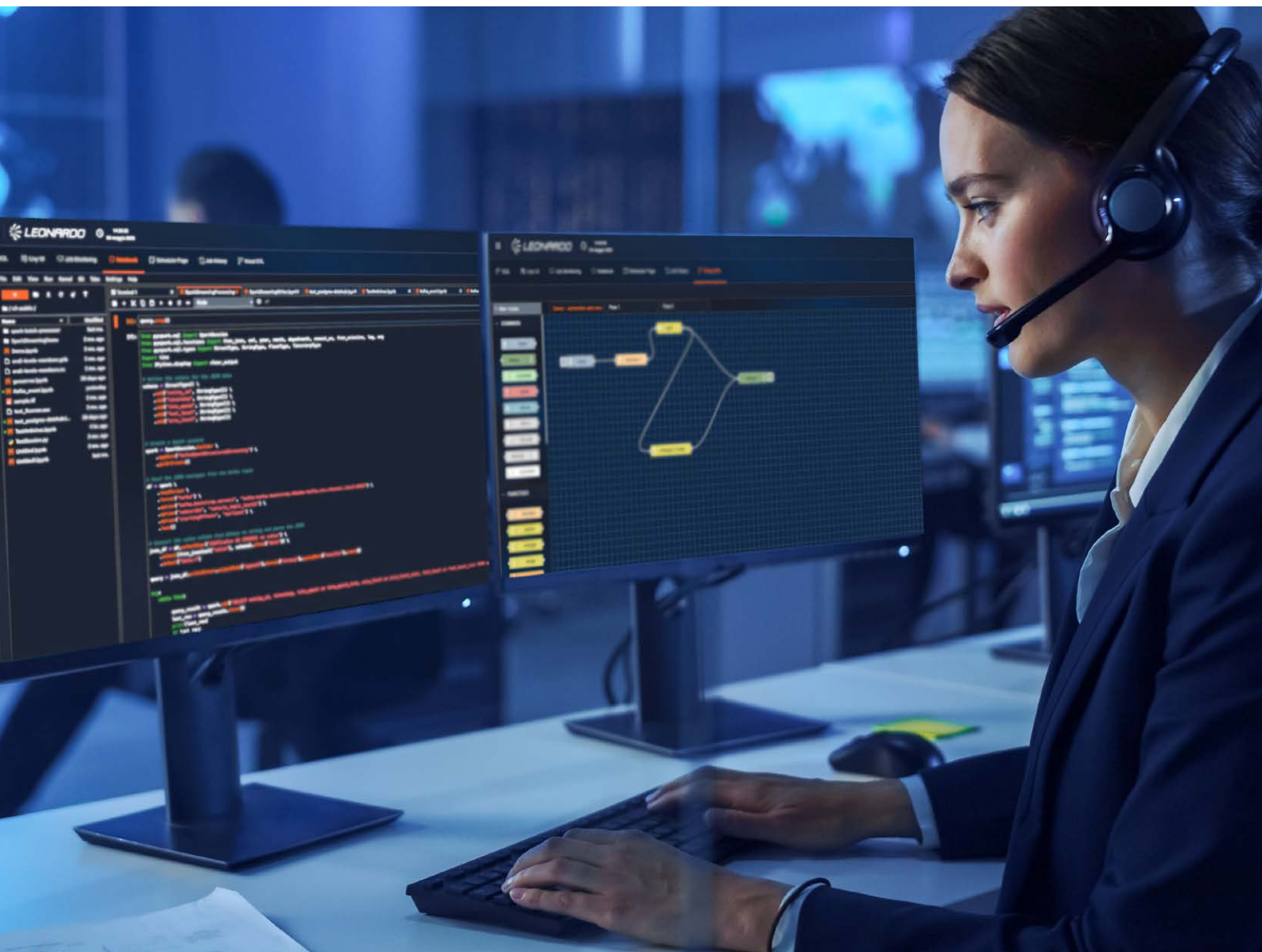
Big Data e AI in un'unica piattaforma, strumenti integrati per l'elaborazione batch, streaming, in tempo reale, ETL con possibilità di linguaggi di query strutturati.

Data centered security

Ciclo di vita dei dati, supporto semplificato per le principali policy di sicurezza e riservatezza e controllo GDPR, CCPA, HIPA (accesso ai dati, mascheramento dei dati, accesso ai dati),

Distribuzione ovunque

Possibilità di operare con infrastrutture ibride.



For more information:
cyberandsecurity@leonardo.com

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

This publication is issued to provide outline information only and is supplied without liability for errors or omissions. No part of it may be reproduced or used unless authorised in writing.
We reserve the right to modify or revise all or part of this document without notice.

LDO_IT25_01487 07-25
JuLY 2025 © Leonardo S.p.A.

