



ANSELM:

an open-source framework for
data analytics and
data visualization.

Giuseppe Marziano
Data Scientist and Developer



#OSW2021



RETE ITALIANA
OPEN SOURCE

ANSELM:

an open-source framework for
data analytics and
data visualization.

dat**A**
visualization**N**
analytic**S**
supers**Et**
e**L**asticsearch
dre**M**io

Ingegnerizzato e sviluppato da:

Beatrice Di Pietro

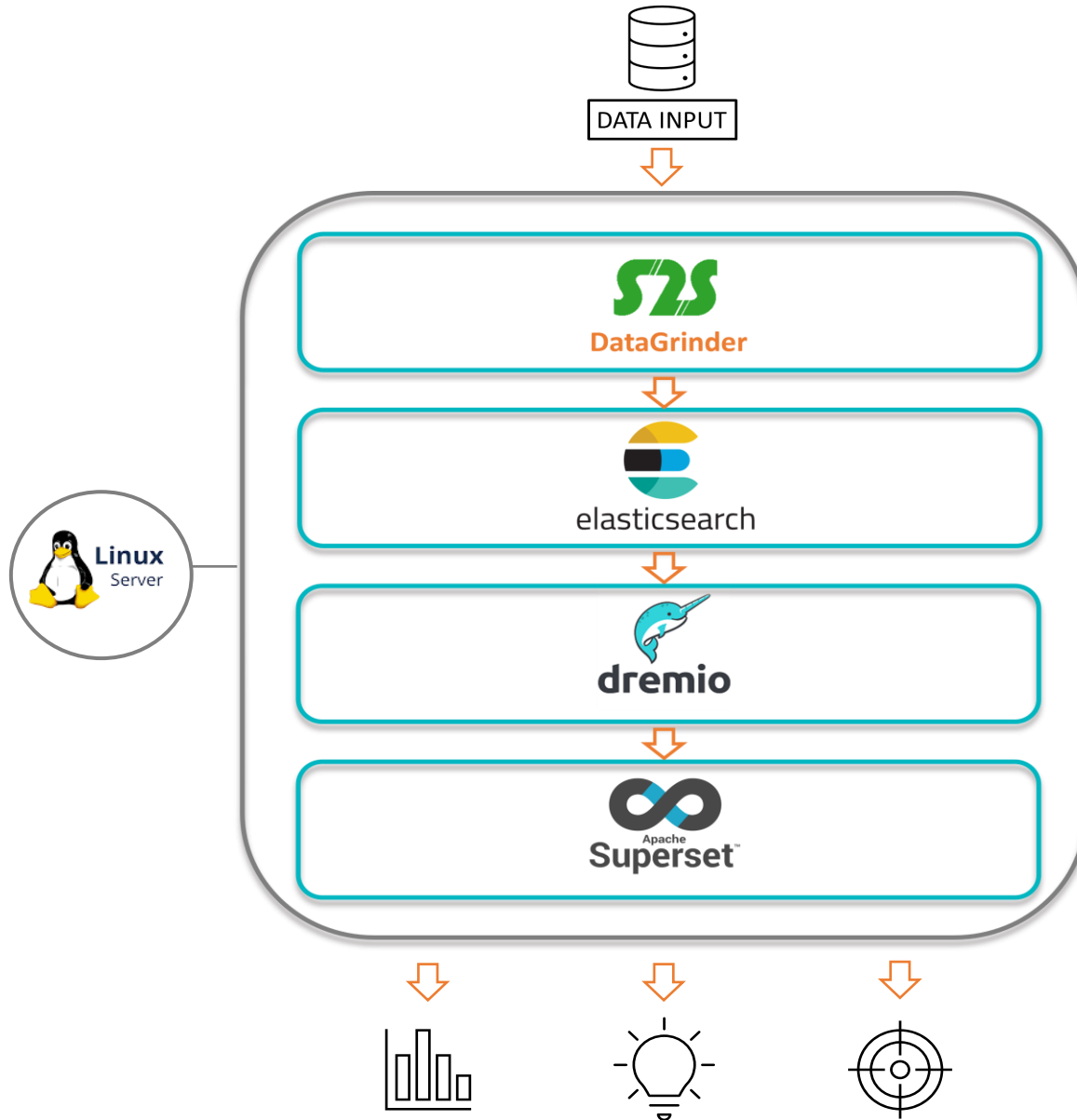
Giuseppe Marziano

Claudia Pozzobon

AGENDA

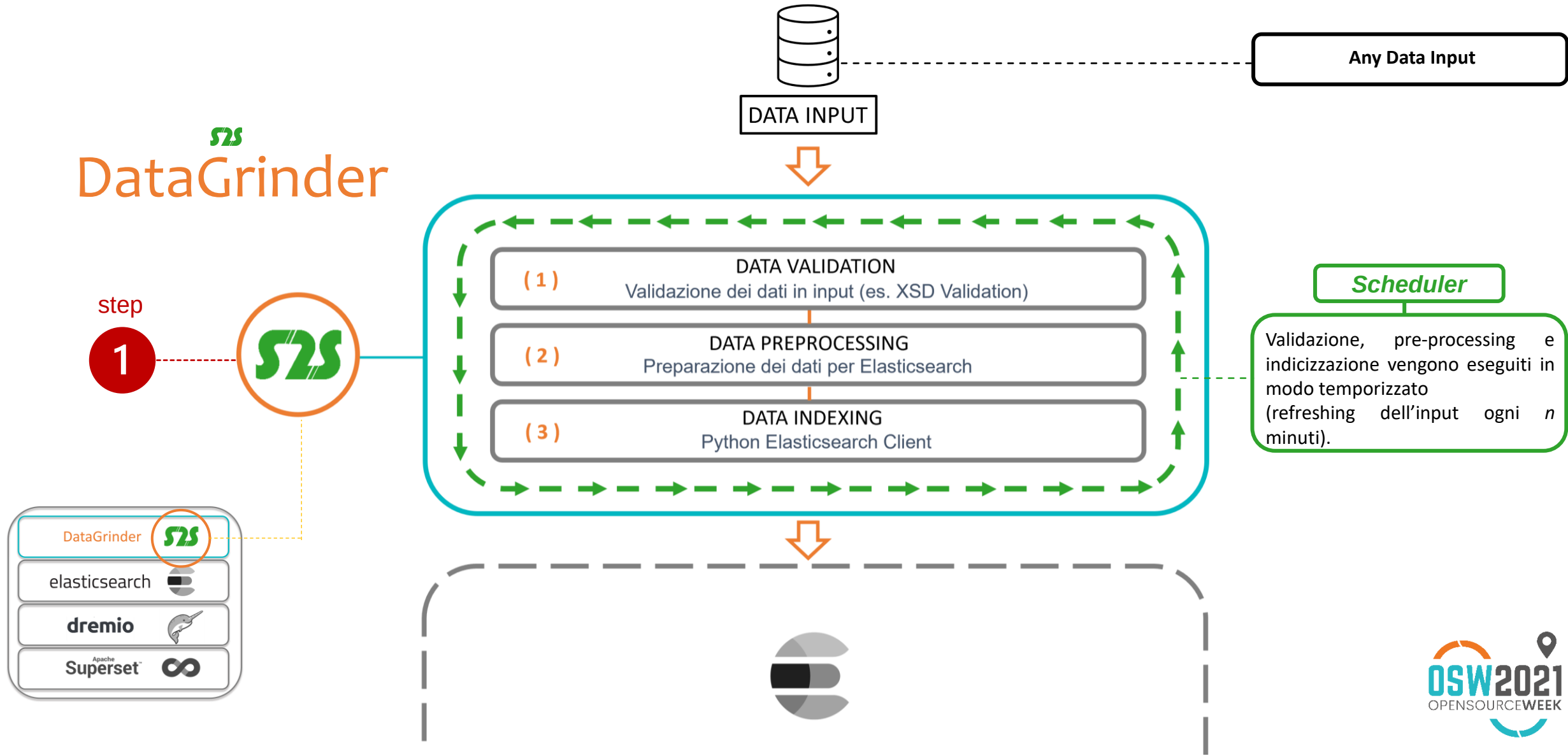
- ANSELM - Introduzione
- ANSELM - Workflow
- SUPERSET - Data Visualization
- Sviluppi In Corso
- Conclusioni

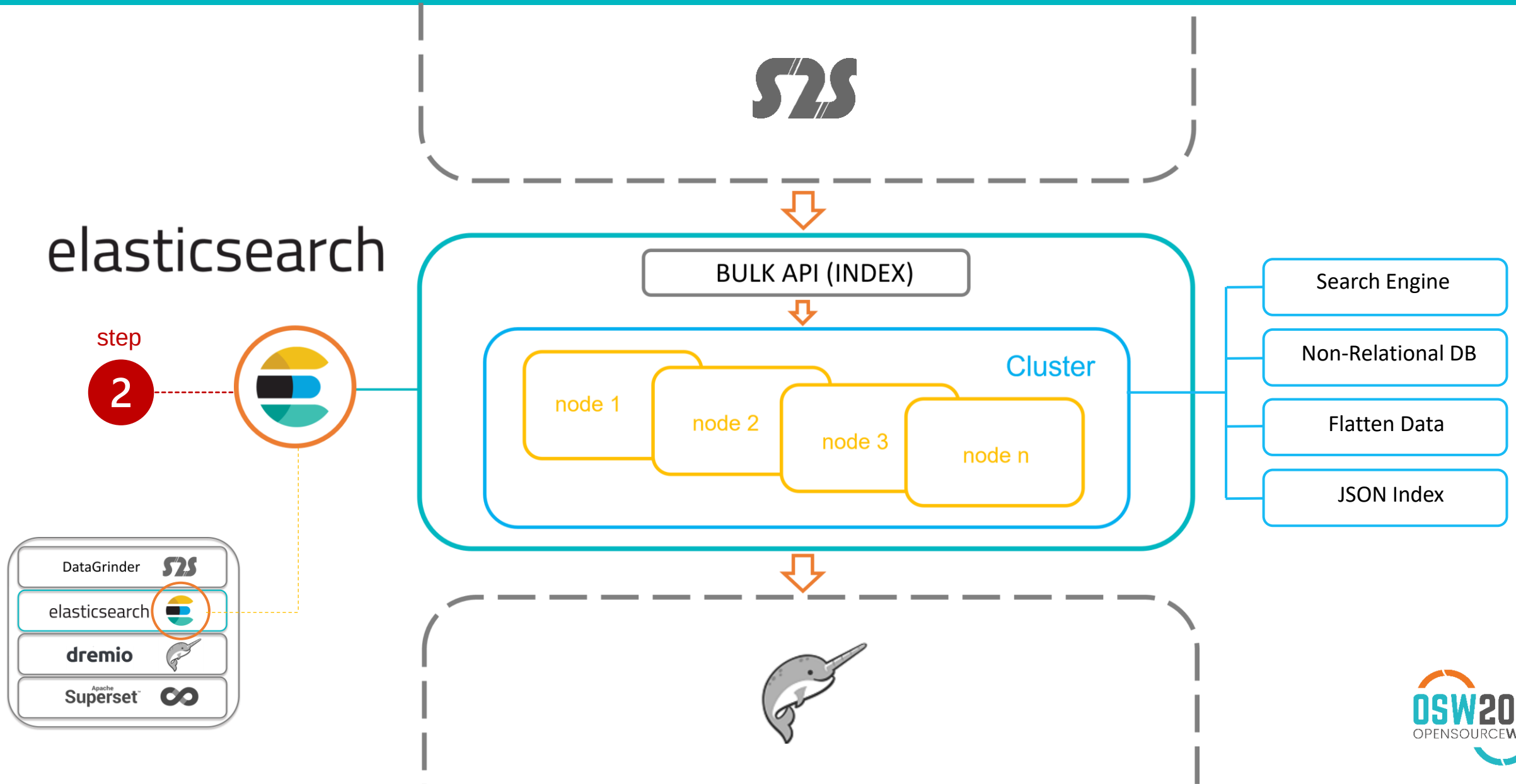
ANSELM - Introduzione



ANSELM

- Gestione Input
- Indicizzazione
- Preparazione dei dati
- Esecuzione delle query
- Analisi descrittiva
- Analisi statistica
- Metriche e benchmarking delle prestazioni
- Visualizzazione dei dati
- Analisi visiva
- Elaborazione di report



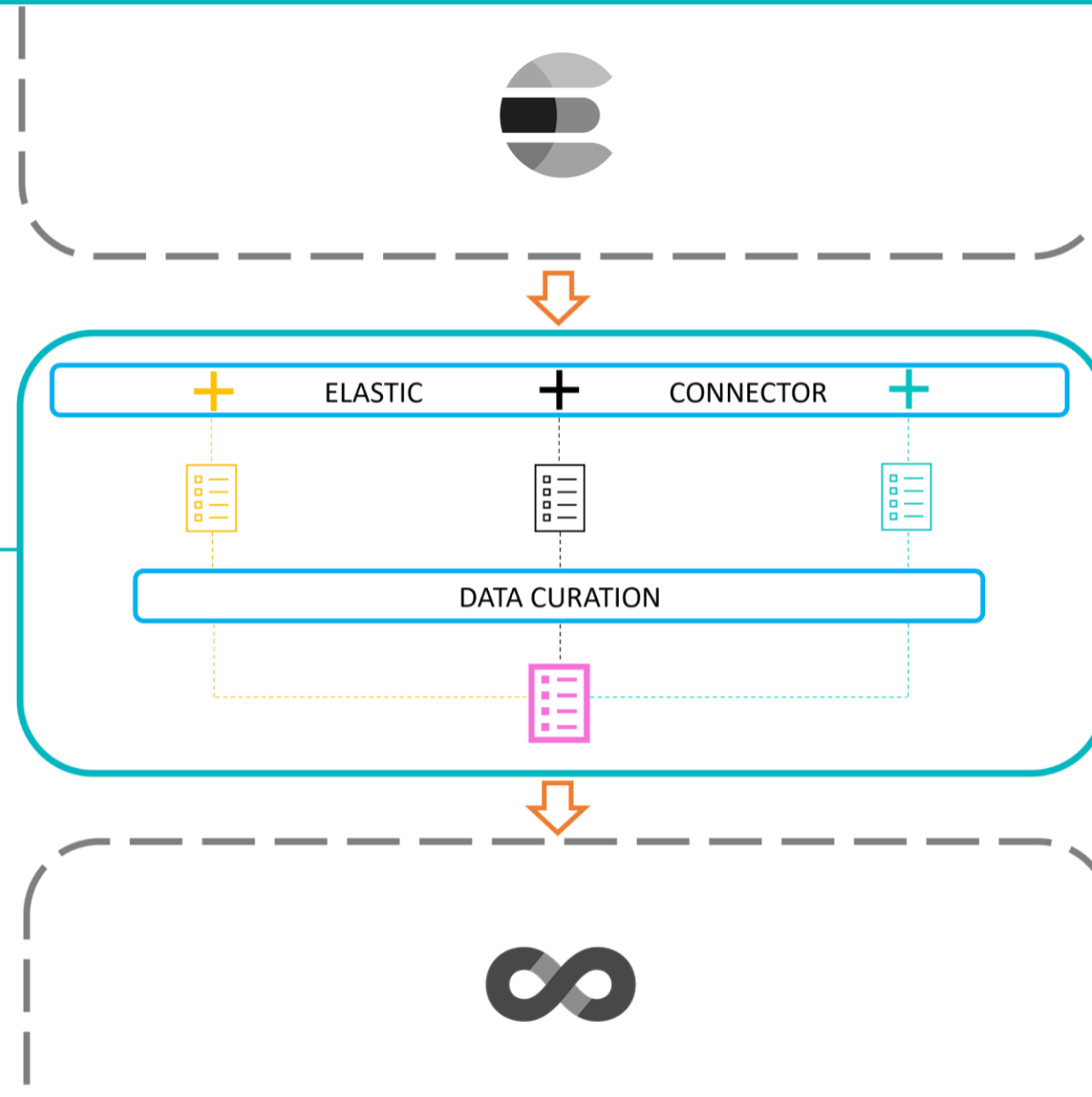


ANSELM – Workflow: Data Curation

dremio

step

3

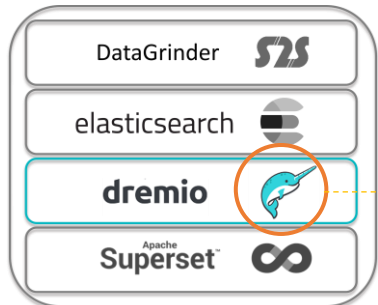


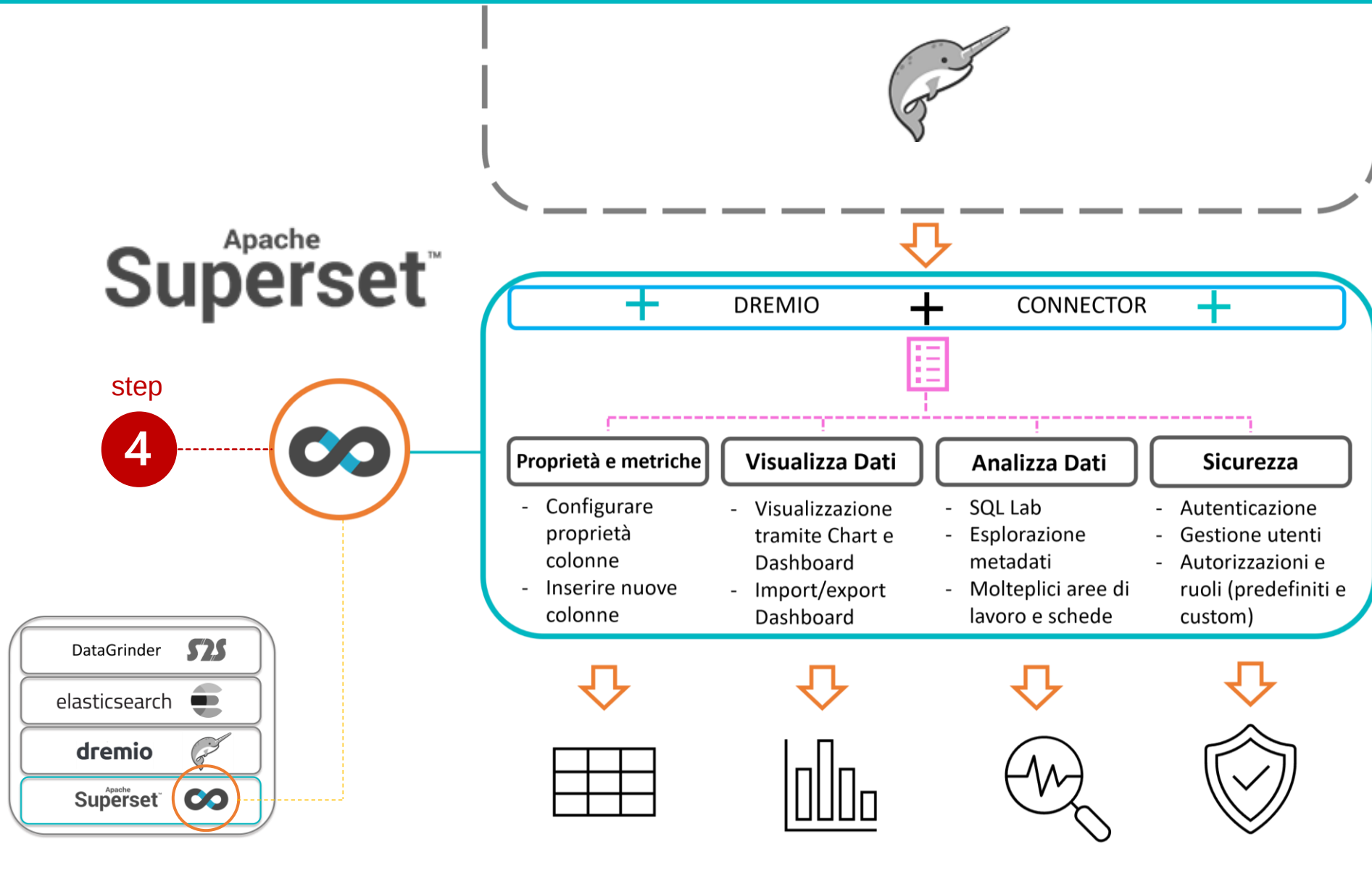
SQL Execution Engine

Possibilità di interrogare i dati con semantica SQL standard.

Data Curation

Combinazione, manipolazione e preparazione dei dati senza creare copie.

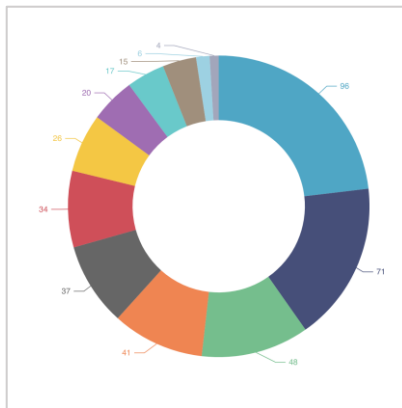




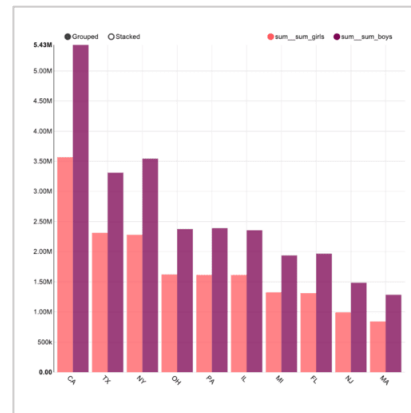
- Analisi dei dati
- Visualizzazioni dei dati
- Grafici
- Report

- **Superset** permette di esplorare i dati mediante diverse tipologie **chart**

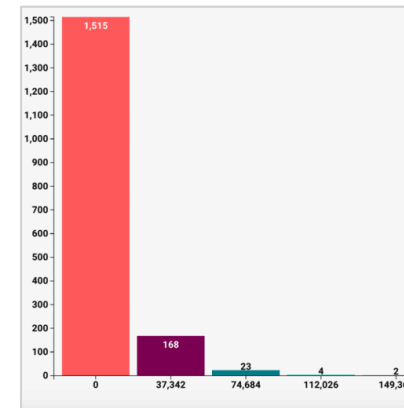
Pie Chart



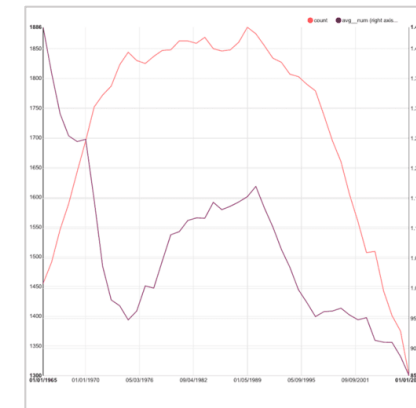
Bar Chart



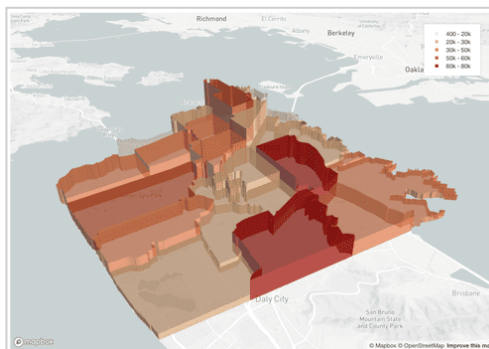
Histogram



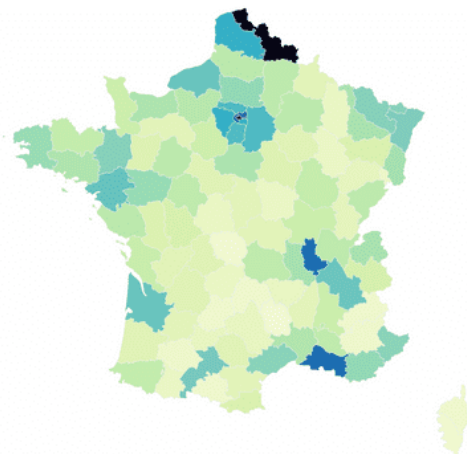
Line Chart



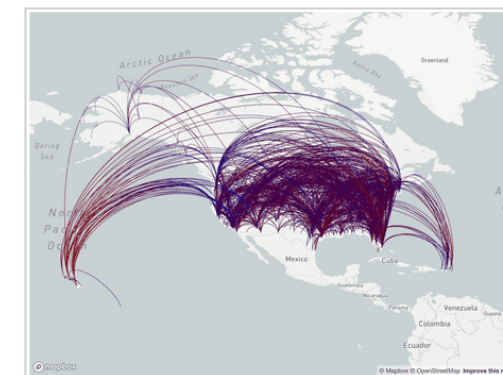
Geospatial Polygon



Country Map



Geospatial Arc



The screenshot shows the Superset SQL Lab interface. At the top, there is a query editor with a SQL query: `SELECT "Sales", "OrderDate", "Status", "Quarter", "Month", "Year", "ProductLine", "MSRP" FROM public.cleaned_sales_data WHERE "OrderDate" > "2020-01-01" AND "OrderDate" < "2021-01-01" LIMIT 5000`. Below the editor, there is a 'RUN' button, a 'LIMIT: 1000' dropdown, and a timer '00:00:00.22'. To the right, there are 'SAVE AS' and 'COPY LINK' buttons. Below the editor, there is a 'RESULTS' tab, a 'QUERY HISTORY' tab, and a 'PREVIEW: 'CLEANED_SALES_DATA'' section. The 'RESULTS' section shows a table with columns: state, started, duration, progress, rows, sql, output, and actions. The first row shows a successful query with 100 rows, and the second row shows a successful query with 0 rows.

state	started	duration	progress	rows	sql	output	actions
success	21:21:12	00:00:00.26		100	<code>SELECT "OrderNumber", "QuantityOrdered", "PriceEach", "OrderLineNumber", "Sales", { ... }</code>	public	🔗 + 📄
success	19:06:28	00:00:00.08		0	<code>SELECT "Date", "Total membership" - la FROM public.superset_stats_jan_21 ORDER BY "Date" DESC</code>	public.daily_new_members	🔗 + 📄

SQL Lab : Editor che consente di scrivere query e visualizzare i metadati:

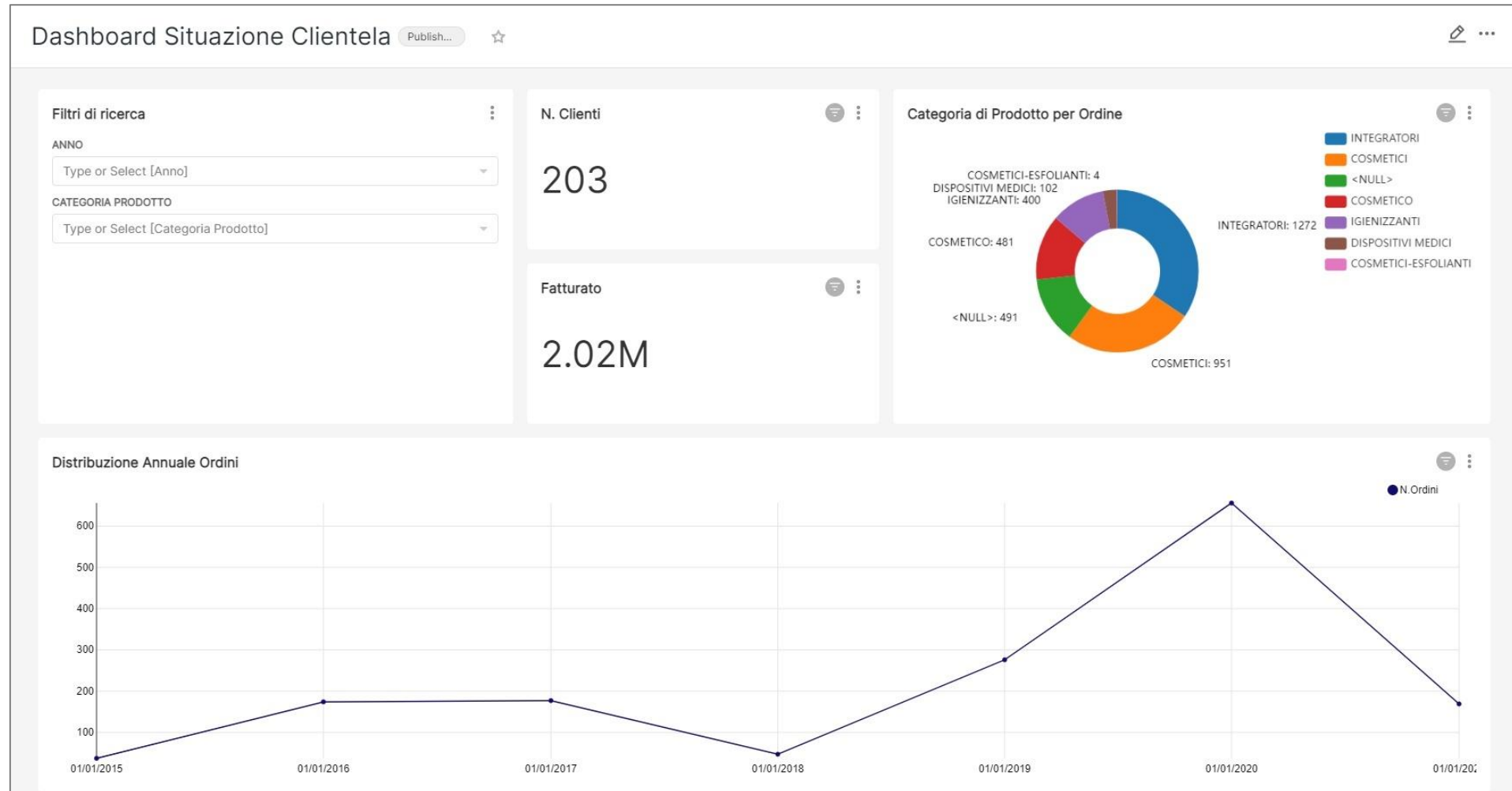
- Molteplici schede come area di lavoro
- Possibilità di creare tabelle e viste sul db



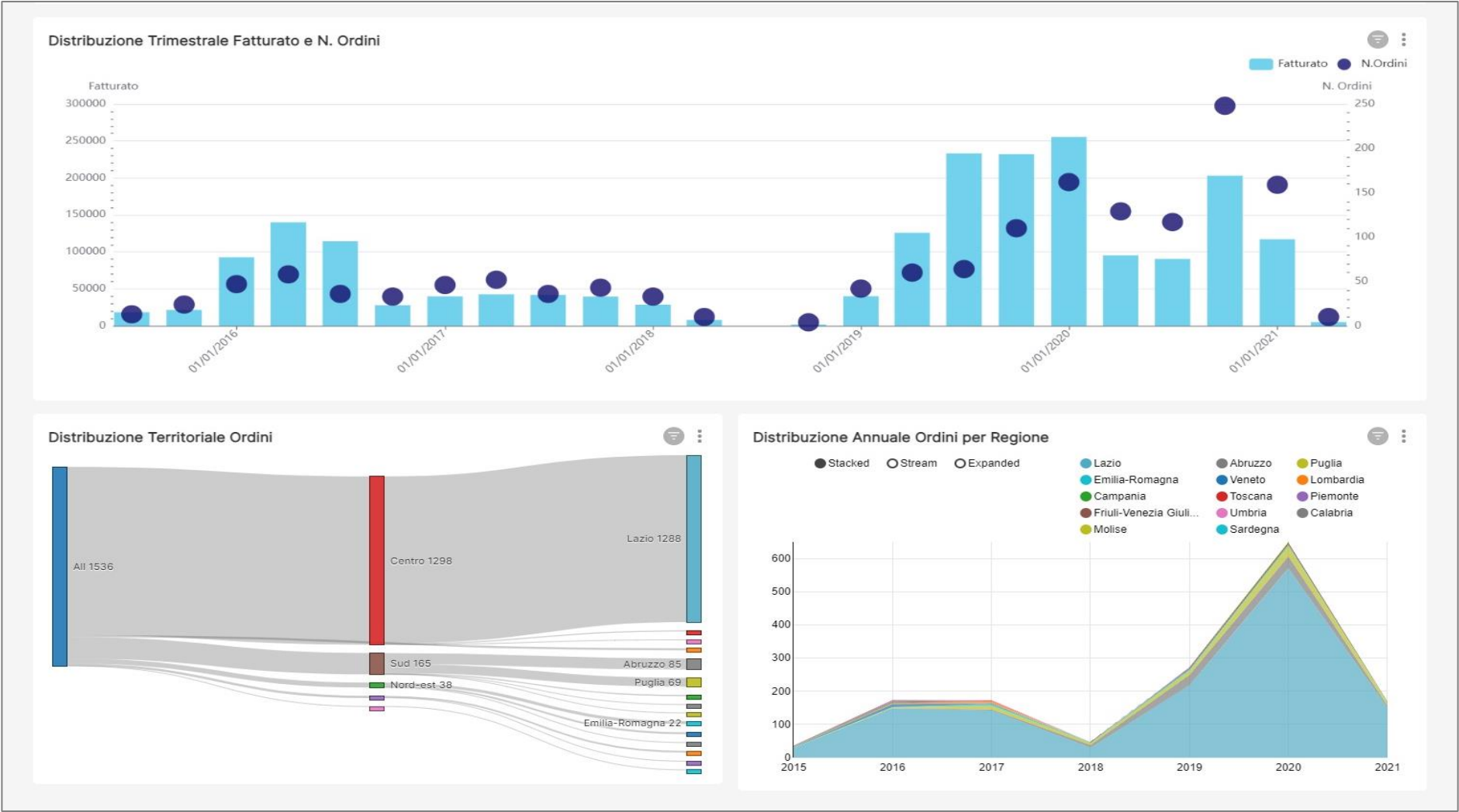
Superset consente di **esportare i Chart Data** in diversi formati:

- Default: **JSON, CSV**
- S2S Export: **XML, XLSX**

SUPERSET - Caso d'uso reale: Dati ERP



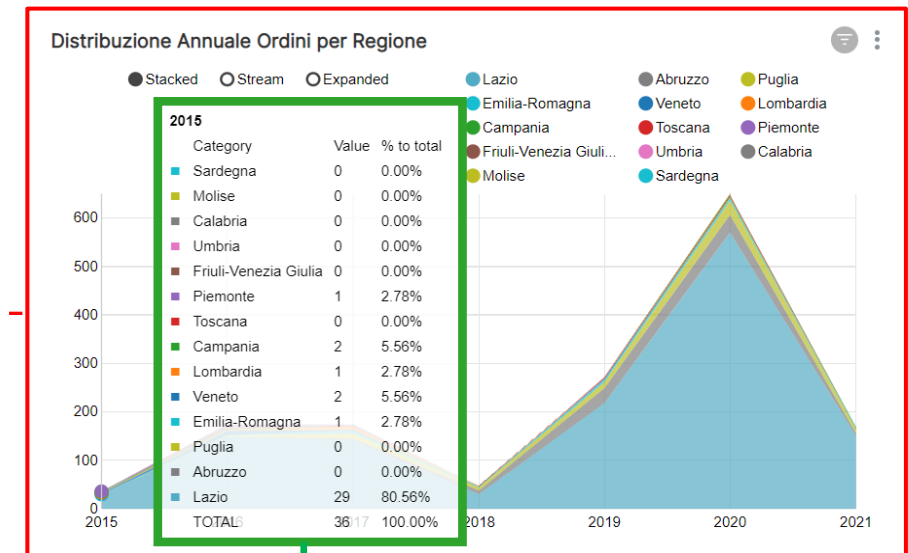
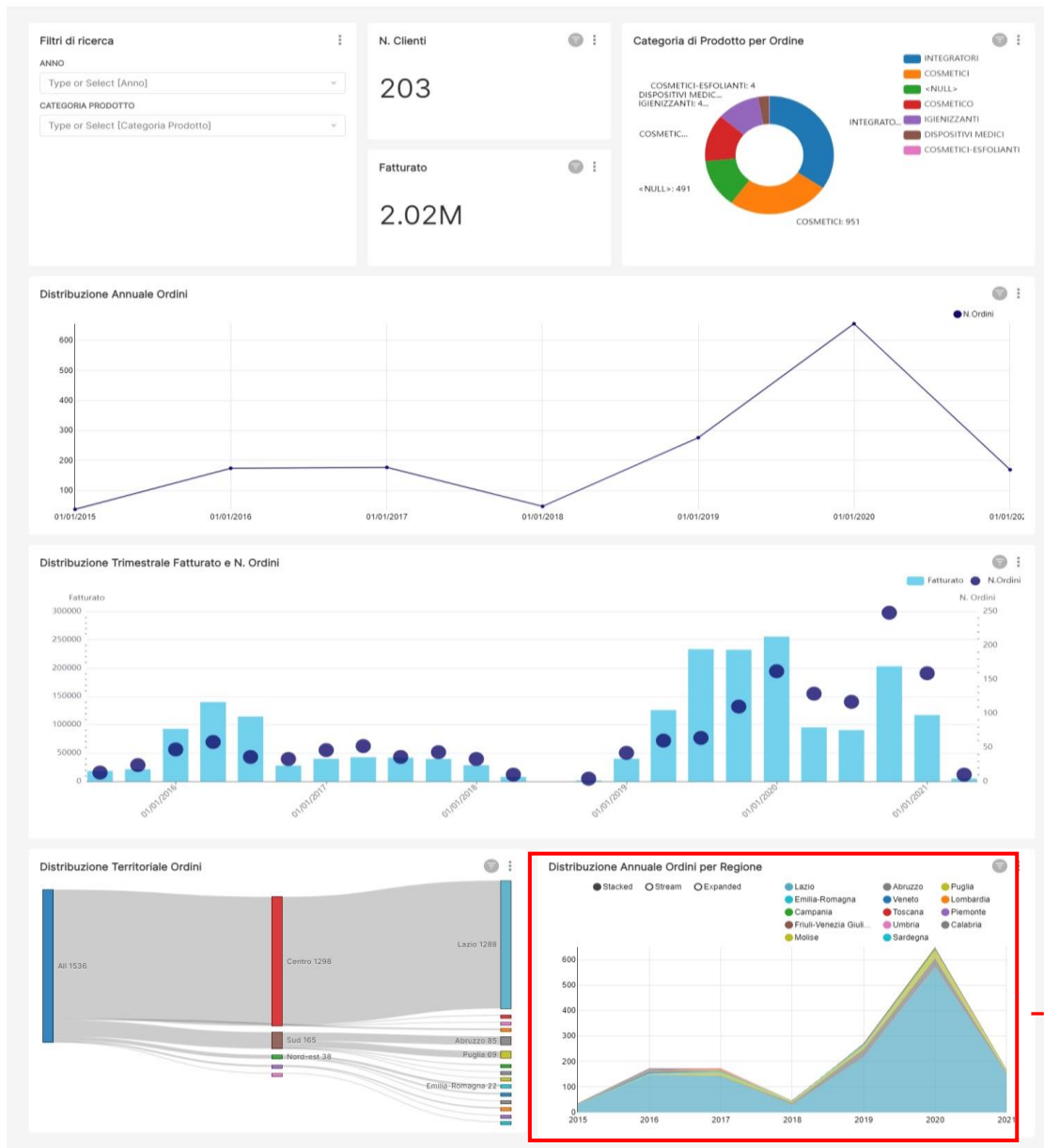
SUPERSET - Un esempio di Dashboard



Sankey Diagram

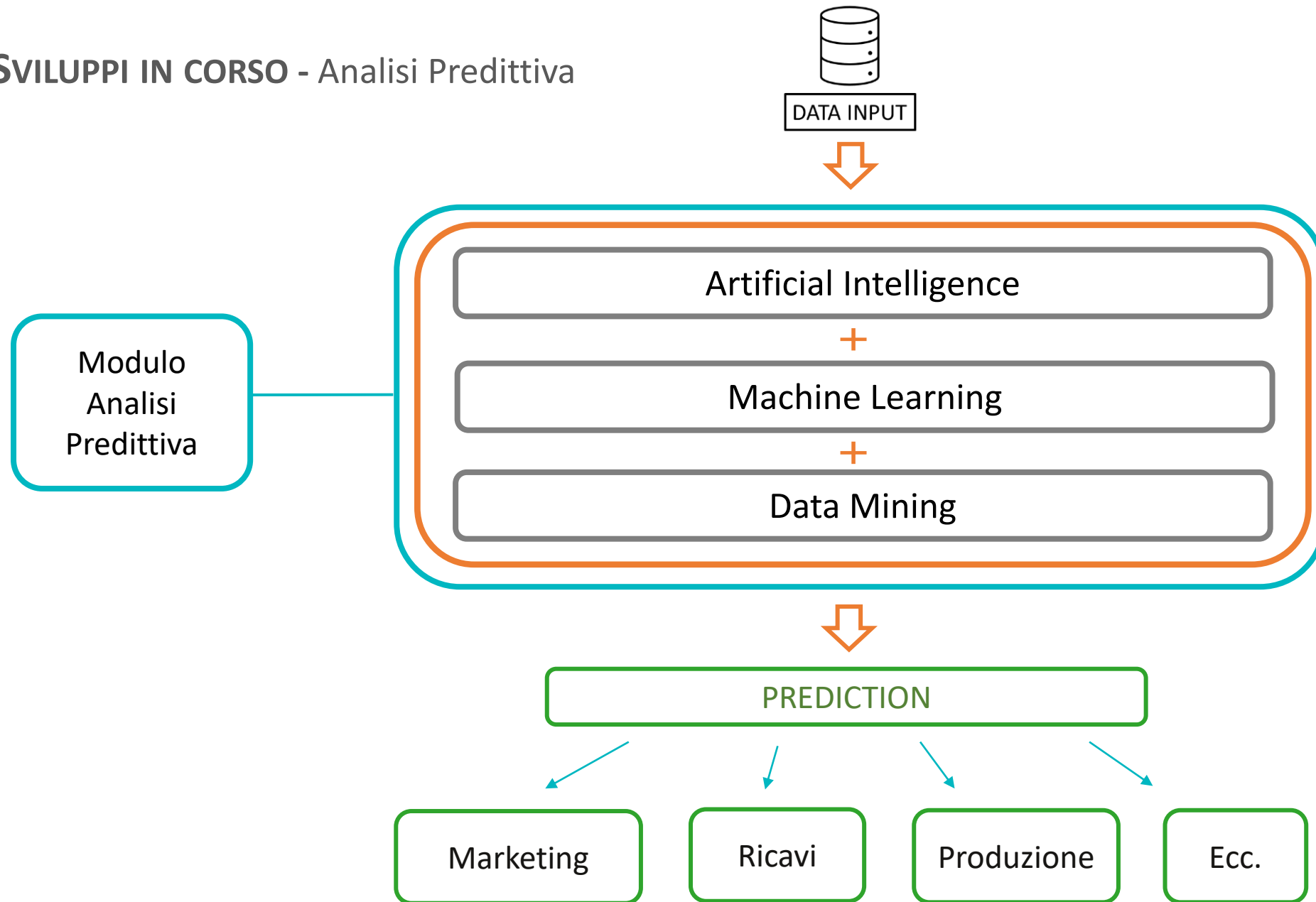
Area Chart

- Superset permette visualizzare i dati attraverso **Dashboard interattive**



Visualizzazione interattiva (Drill-Down)

SVILUPPI IN CORSO - Analisi Predittiva





ANSELM:

an open-source framework for
data analytics and
data visualization.

datA
visualizationN
analyticS
supersEt
eLasticsearch
dreMio

Value Proposition:

- Any Data Input
- Analisi dei Dati Actual
- Data Visualization Self-Service
- Data Outlook
- Stack Open-Source

Grazie per l'attenzione!



Giuseppe Marziano

marziano@s2sprodotti.it

www.s2sprodotti.it