IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"

14-17 February 2022

The data analytics lab – from innovation to products[1]

Mona Amer, Hiren Jani and Mathieu Le Cam,
BIS

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

BIS

The Data Analytics Lab

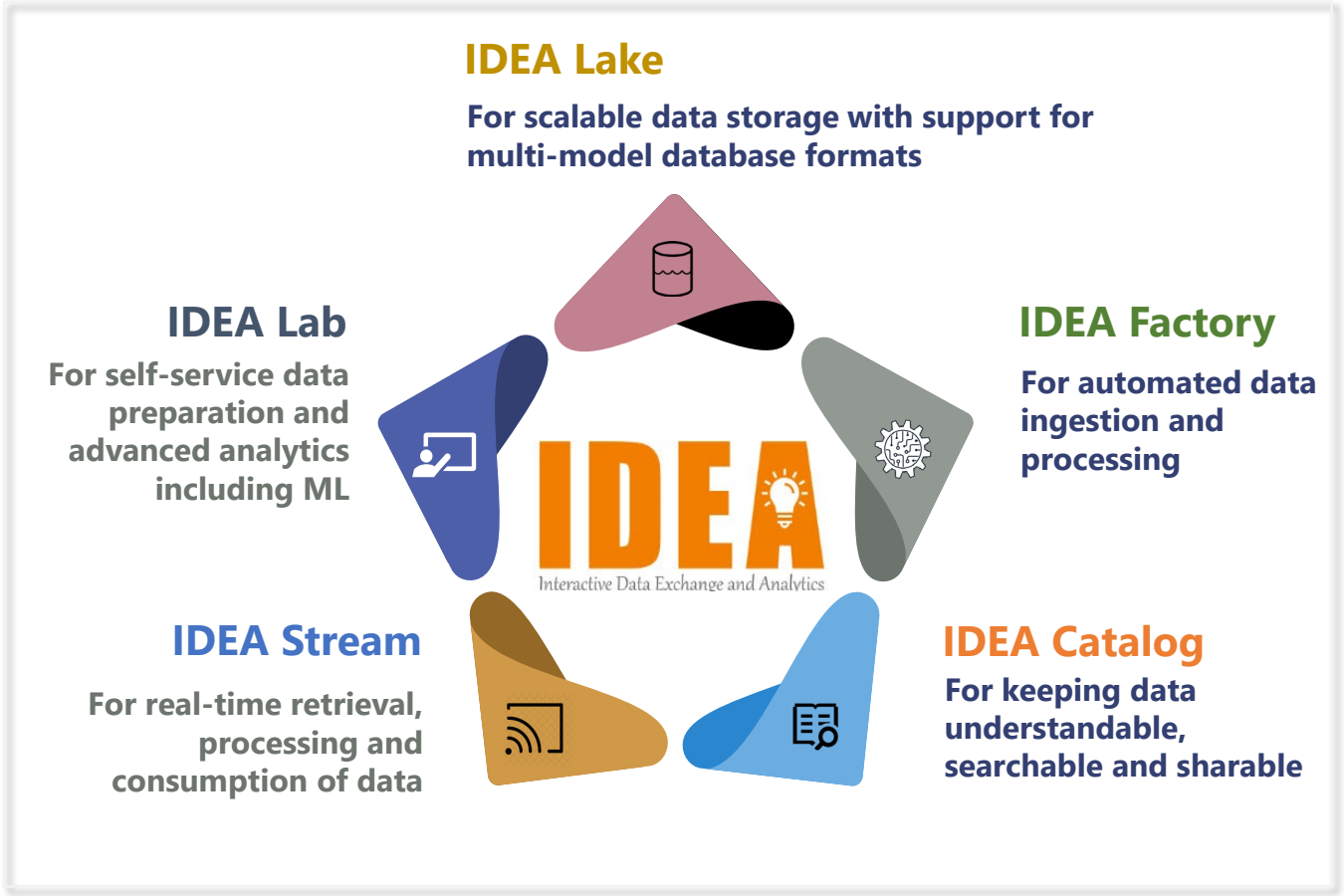
From innovation to products

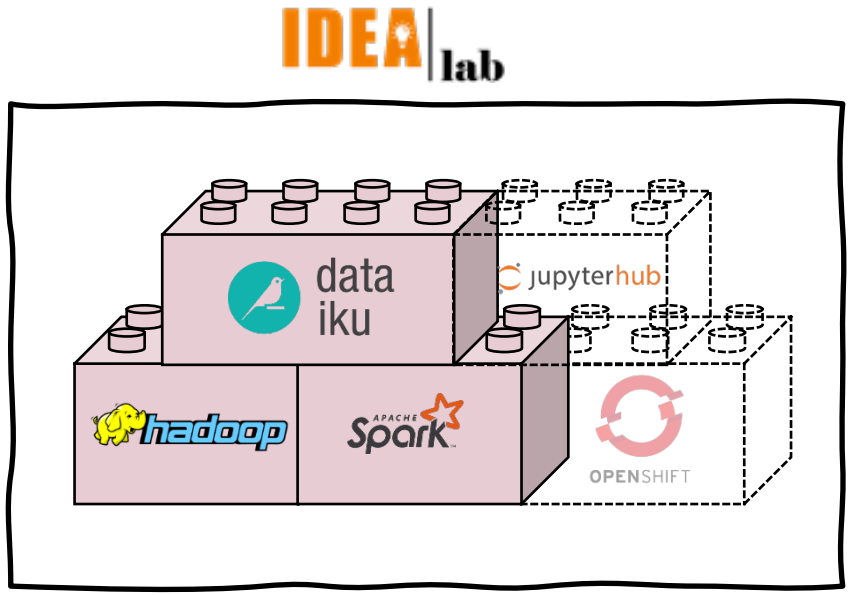
Mission and Vision statements

MISSION: provide **self-service platform** that allows users to use data science, AI and flexible data wrangling capabilities to turn data into insight.

VISION: improve efficiency and collaboration across data analytics community in the Bank by providing data lab solutions that are **easy to use**, **to scale**, **to govern**, **to operationalise** and **to integrate** with existing and new data and analytical systems.

IDEA Lab & the BIS Big Data Platform

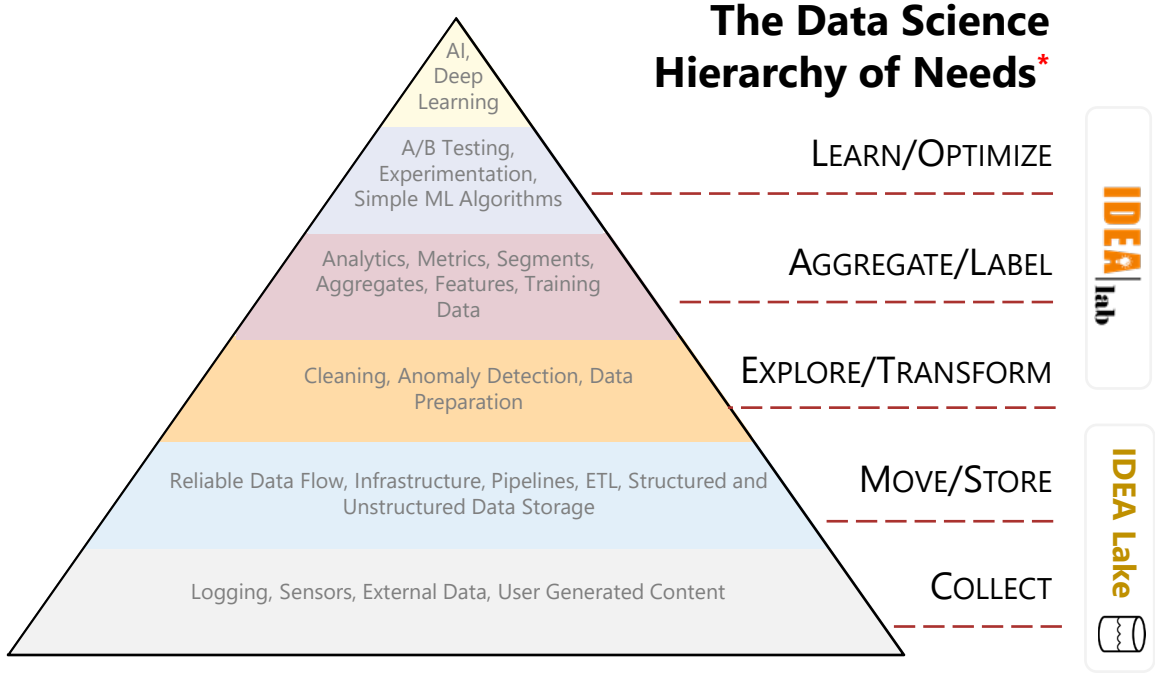




The IDEA Lab design components offer data scientists and analysts greater agility and innovation to maximize value derived from diverse data sources in an efficient, yet, simple manner

How IDEA Lab Delivers Value?

The Data Science Hierarchy of Needs*



IDEA lab

The IDEA Lab goes above and beyond the needs of a successful Data Science and Advanced Analytics Implementation

DRIVING PRINCIPLES

Collaboration & Integration
- Collaborative team workspaces
- Insights sharing

Environment Scalability
- Scale-out storage
- Scale-out/up compute

Versatility
- Wide spectrum of profiles and use cases
- From Low-Code to Full-Code

Reusability/Customization
- Leverage previous work
- Packaging and sharing

Reproducibility/Auditability
- Versioning of artefacts
- Activity logging & monitoring

ML/Ops Enablement
- ML models development support
- Readiness for operations

Path to Operationalization
- User-friendly orchestration
- Packaging and deployment

Community Drive
- Share XP for upskilling and ideation
- Express requirements to shape roadmap

* Based on https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

Spotlight on products

	Financial Projection	D2NLP	TweetWatcher
	• Mid-term **projection of the Bank's key financial measures** • Support discussion regarding dividend policy	• NLP to support **document management** • Help human curate very large document stores	• **Analyze tweets** related to the Central Banking community • Identify trends and topics of interest
	• BIS internal financial information & reporting structure • BIS interest rates projection	• Documents, archives, correspondence • Millions of text documents in various formats (Word, PDF, etc)	• User tweet & mention timelines • User details and metrics • Filtered stream – real time feed of public tweets
	• Financial **data preparation** and consolidation • User-defined projection **scenarios & parameters**	• Apply language models to compute **similarity** scores between documents • Documents **summarization** via BERT embeddings • **Topic modelling** on ad-hoc, user-defined collections	• Data denormalization and **enrichments** • **Hashtags** and **entities** extraction • Language detection and offline neural **machine translation** of non-English tweets • **Sentiment** analysis of user mention timelines
	• Configurable financials projection solution • Automated reporting and dissemination of projection results	• Database of potential duplicate with metadata, summaries and analysis results • Web application for exploration and analysis requests	• Growing database of enriched Twitter datasets • Filtered stream rules management solution • Executive dashboards to deliver key insights

Challenges & Enhancements

Challenges of Establishing and Running a Data Lab

- √ Resistance to change – how to move from "comfort-zone" to "growth-zone"
- √ Keeping up with the needs – to meet swiftly new connectivity requirements, customizations and process capabilities
- √ Making data accessible – set up a process to discover, find and get access to datasets, but also share and ensure quality
- √ Finding the right governance – product criticality varies from initiative to initiative, one size doesn't fit all

Envisaged Enhancements

- √ Community development – increase knowledge sharing sessions, schedule dedicated technical workshops
- √ New compute options – complementary software, scale out/up processes with container orchestration & GPU computing
- √ Cloud-based Innovation Zones – extensions to the IDEA Lab with higher degree of freedom, to facilitate experimentation with non-sensitive data

Lessons Learned

Self-service is (indeed) a catalyst for innovation

Upskilling is critical, requires time and close collaboration between units

Balance between freedom and governance is essential

The frame of a platform constrains but stimulate delivery

Mindset and culture are keys to adoption and take time to change

THANK YOU!