Industrial-Driven Big Data as a Self-Service Solution

Dusan Jakovetic

University of Novi Sad, Faculty of Sciences, Serbia

Virtual BenchLearning Webinar

July 8, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787

Outline

- Brief introduction to I-BiDaaS
- I-BiDaaS pipeline, architecture & technologies
- I-BiDaaS & BDVA reference model
- Benchmarking landscape and opportunities @ I-BiDaaS

Identity card

Project Consortium
13 partners

PROJECT NAME
Industrial-Driven Big Data as a Self-Service Solution

TOTAL BUDGET / TOTAL EC FUNDING
€ 4 997 035

PROJECT TYPE
RIA

START DATE
1 January 2018

DURATION
36 months

http://www.ibidaas.eu/ @Ibidaas https://www.linkedin.com/in/i-bidaas/

Consortium

1. FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS (**FORTH**)
2. BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION (**BSC**)
3. IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD (**IBM**)
4. CENTRO RICERCHE FIAT SCPA (**CRF**)
5. SOFTWARE AG (**SAG**)
6. CAIXABANK, S.A (**CAIXA**)
7. THE UNIVERSITY OF MANCHESTER (**UNIMAN**)
8. ECOLE NATIONALE DES PONTS ET CHAUSSEES (**ENPC**)
9. ATOS SPAIN SA (**ATOS**)
10. AEGIS IT RESEARCH LTD (**AEGIS**)
11. INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP (**ITML**)
12. UNIVERSITY OF NOVI SAD FACULTY OF SCIENCES SERBIA (**UNSPMF**)
13. TELEFONICA INVESTIGACION Y DESARROLLO SA (**TID**)

Key messages

A **complete** and **safe environment** for methodological **big data experimentation**

Tool and services to **increase the quality** of data analytics

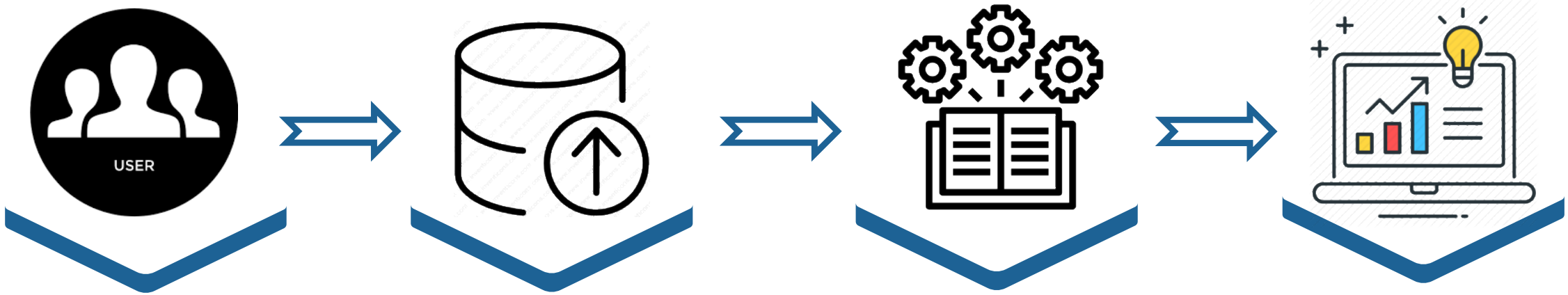
A Big Data as a **Self-Service solution** that helps in **breaking silos** and boosts EU's data-driven economy

Tools and services for **fast ingestion and consolidation** of both realistic and fabricated data

Increases impact in research community and contributes to industrial innovation capacity

Tools and services for the management of **heterogeneous infrastructures**

I-BiDaaS pipeline



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

Results

- Visualize the results
- Share models

Benefits of using I-BiDaaS

Do it yourself In a flexible manner

Break data silos

Safe environment

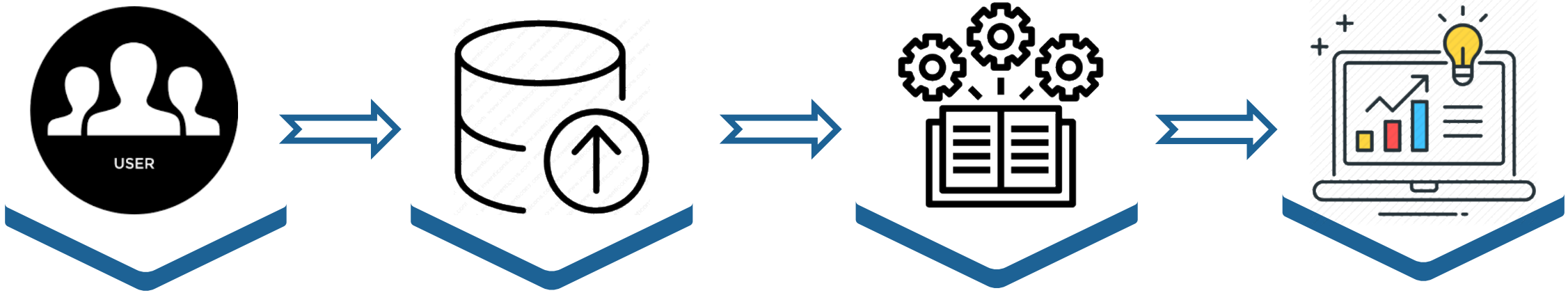
Interact with Big Data technologies

Increase speed of data analysis

Intra- and inter-domain data-flow

Cope with the rate of data asset growth

I-BiDaaS pipeline



Users

- **Expert mode**
- **Self-service mode**
- **Co-develop mode**

Data

- Import your data

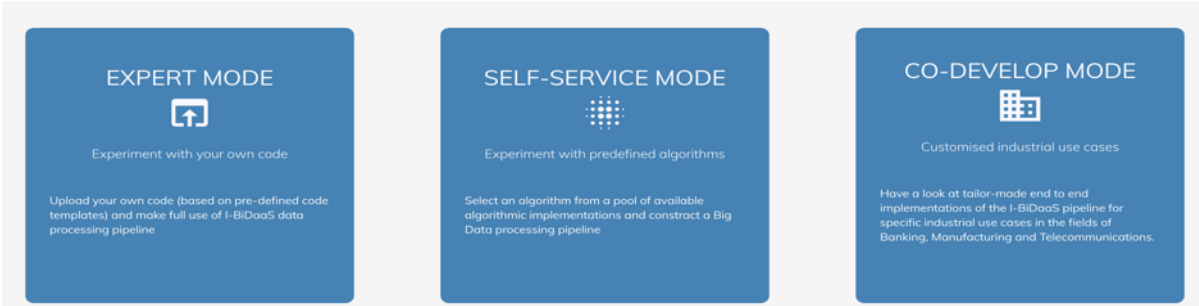
Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code

Results

- Visualize the results
- Share models

Flexible solution



Benefits of using I-BiDaaS

Do it yourself In a flexible manner

Break data silos

Safe environment

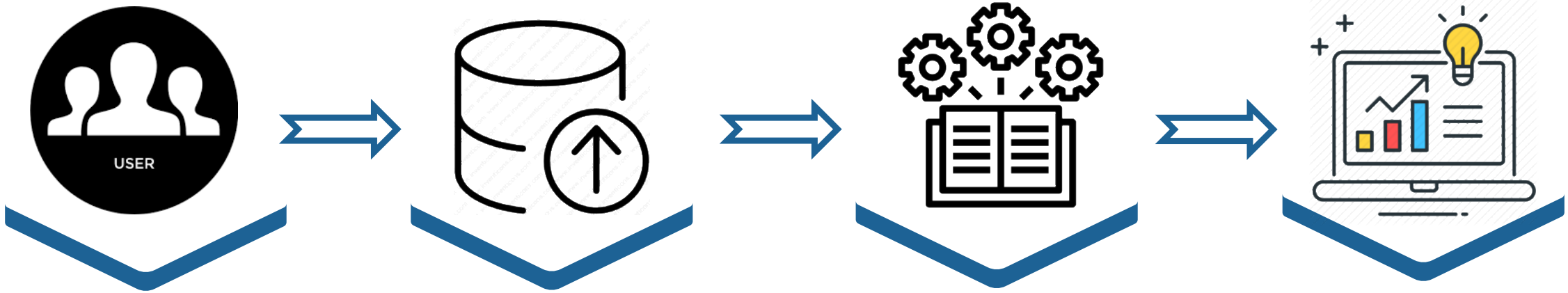
Interact with Big Data technologies

Increase speed of data analysis

Intra- and inter-domain data-flow

Cope with the rate of data asset growth

I-BiDaaS pipeline



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- **Fabricate Data**
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an

Results

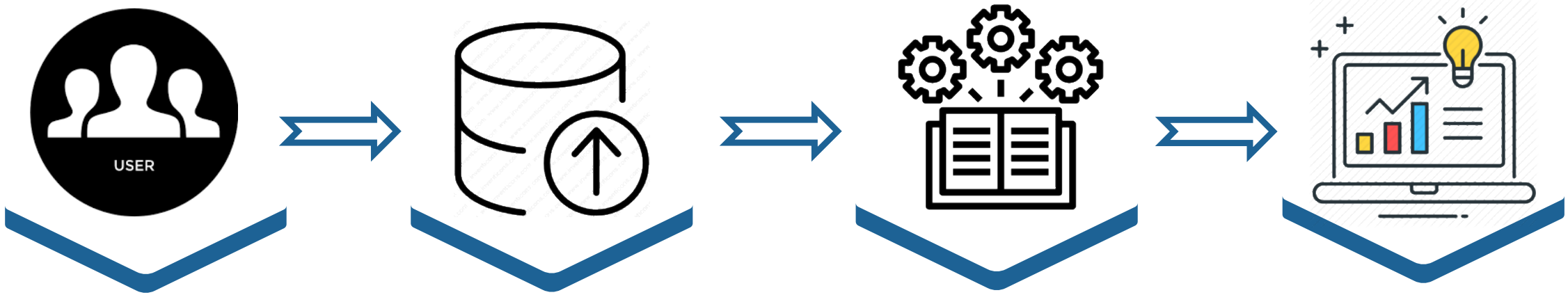
- Visualize the results
- Share models

Data sharing & breaking silos

Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

I-BiDaaS pipeline



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- **Stream & Batch Analytics**
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

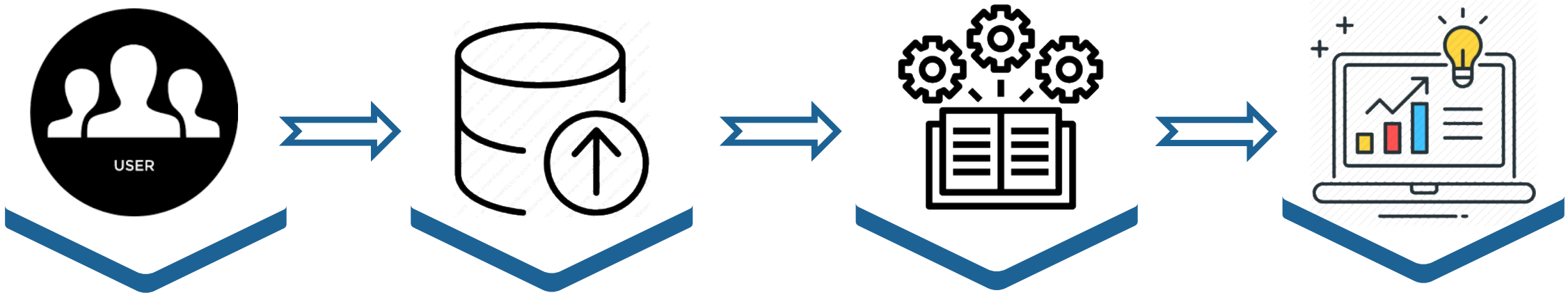
Results

- Visualize the results
- Share models

Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

I-BiDaaS pipeline



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

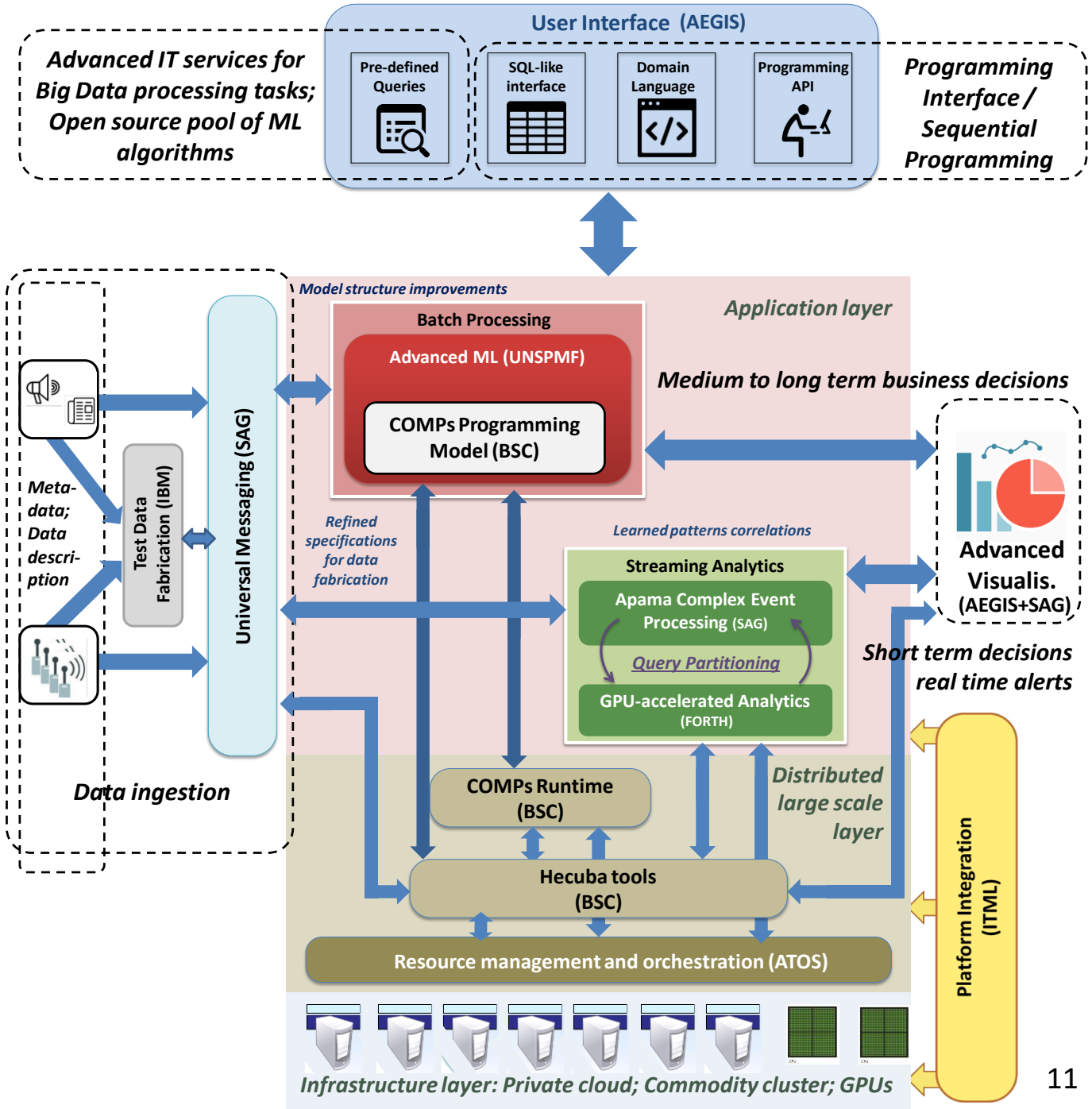
Results

- **Visualize the results**
- **Share models**

Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

The I-BiDaaS solution: Architecture & technologies

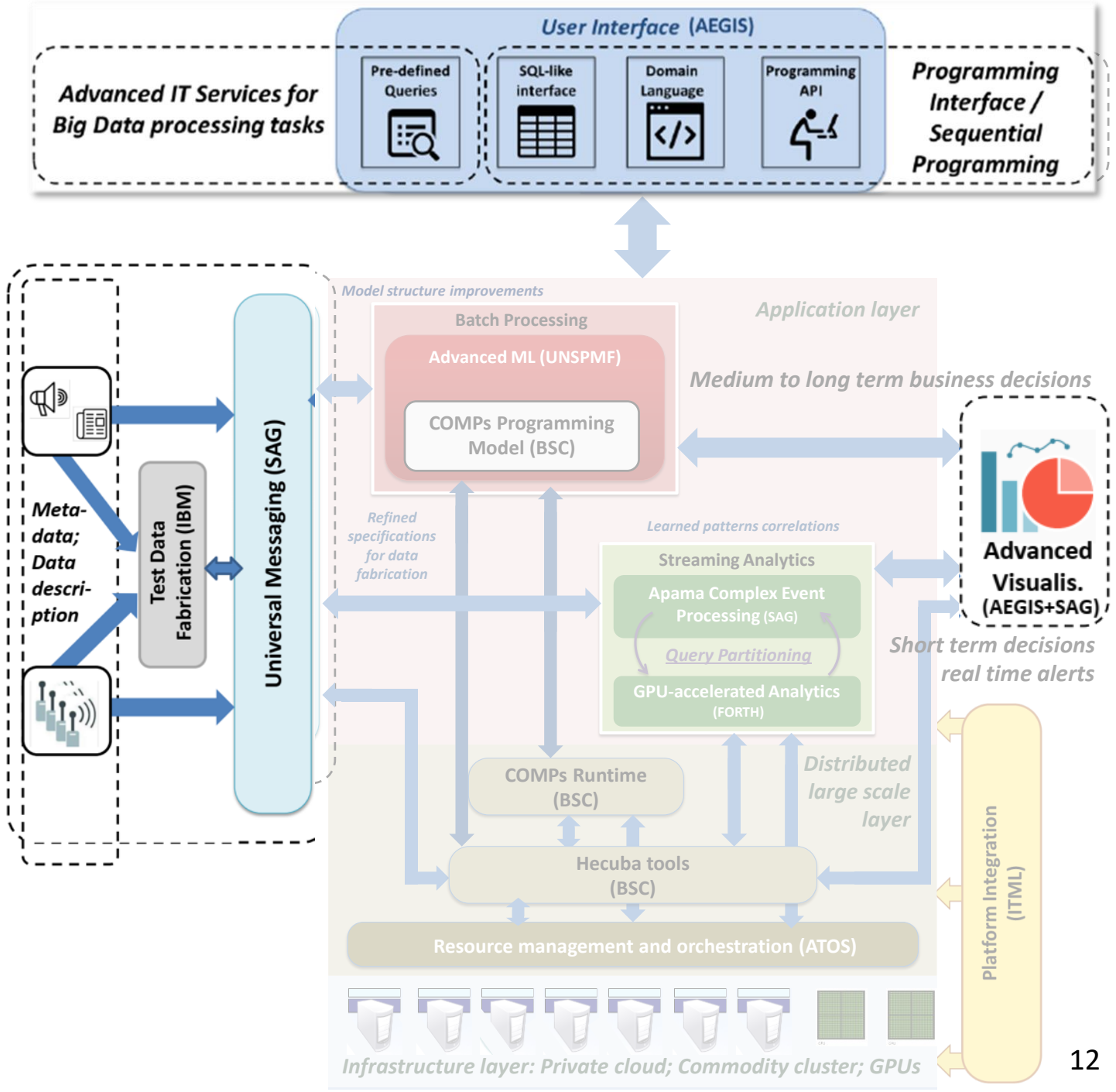


WP2:

Data, user interface, visualization

Technologies:

- IBM TDF
- SAG UM
- AEGIS AVT

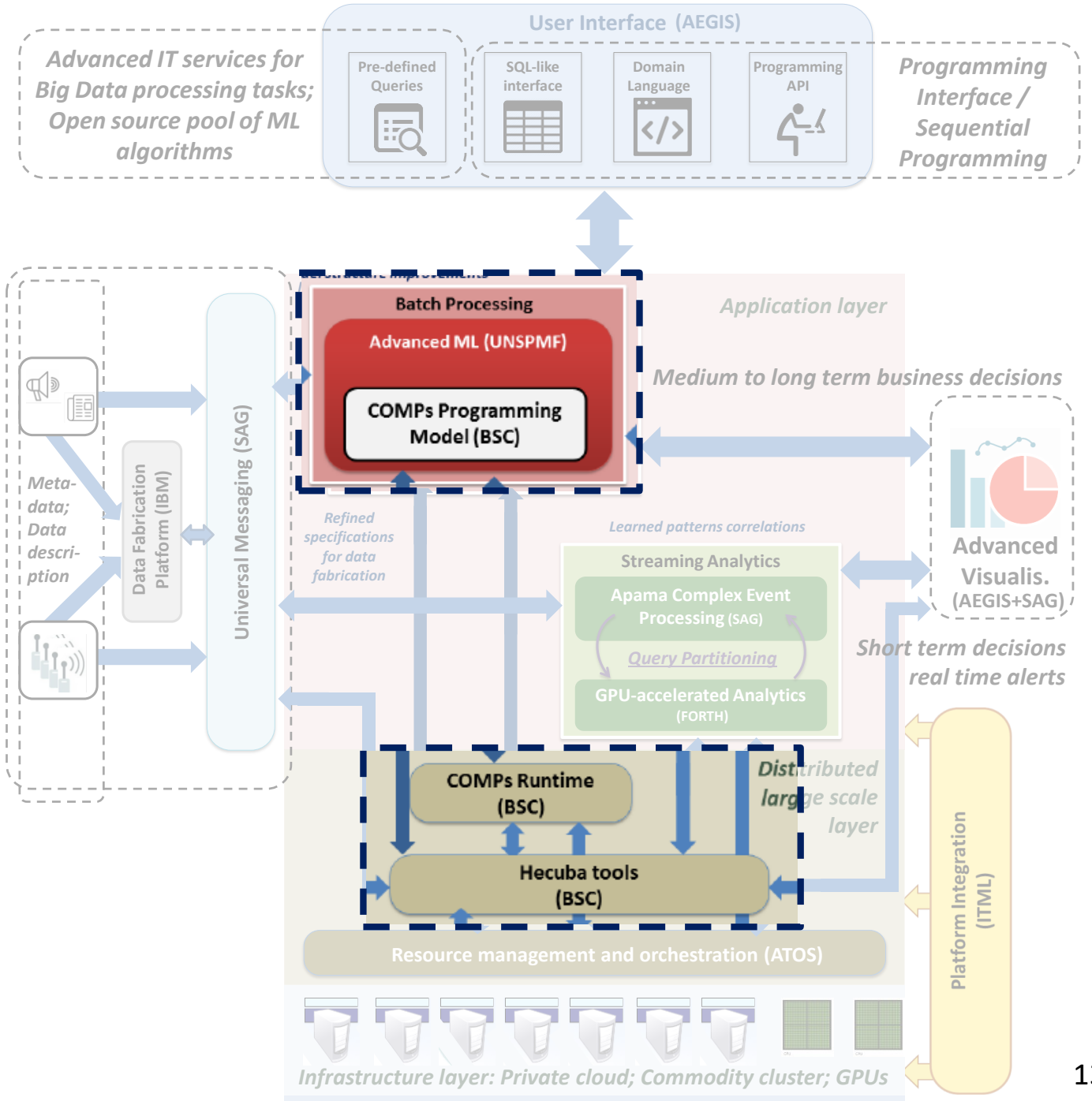


http://ibidaas.eu/tools

WP3:
Batch analytics

Technologies:

- BSC COMPSs
- BSC Hecuba
- BSC Qbeast
- Advanced ML (UNSPMF)

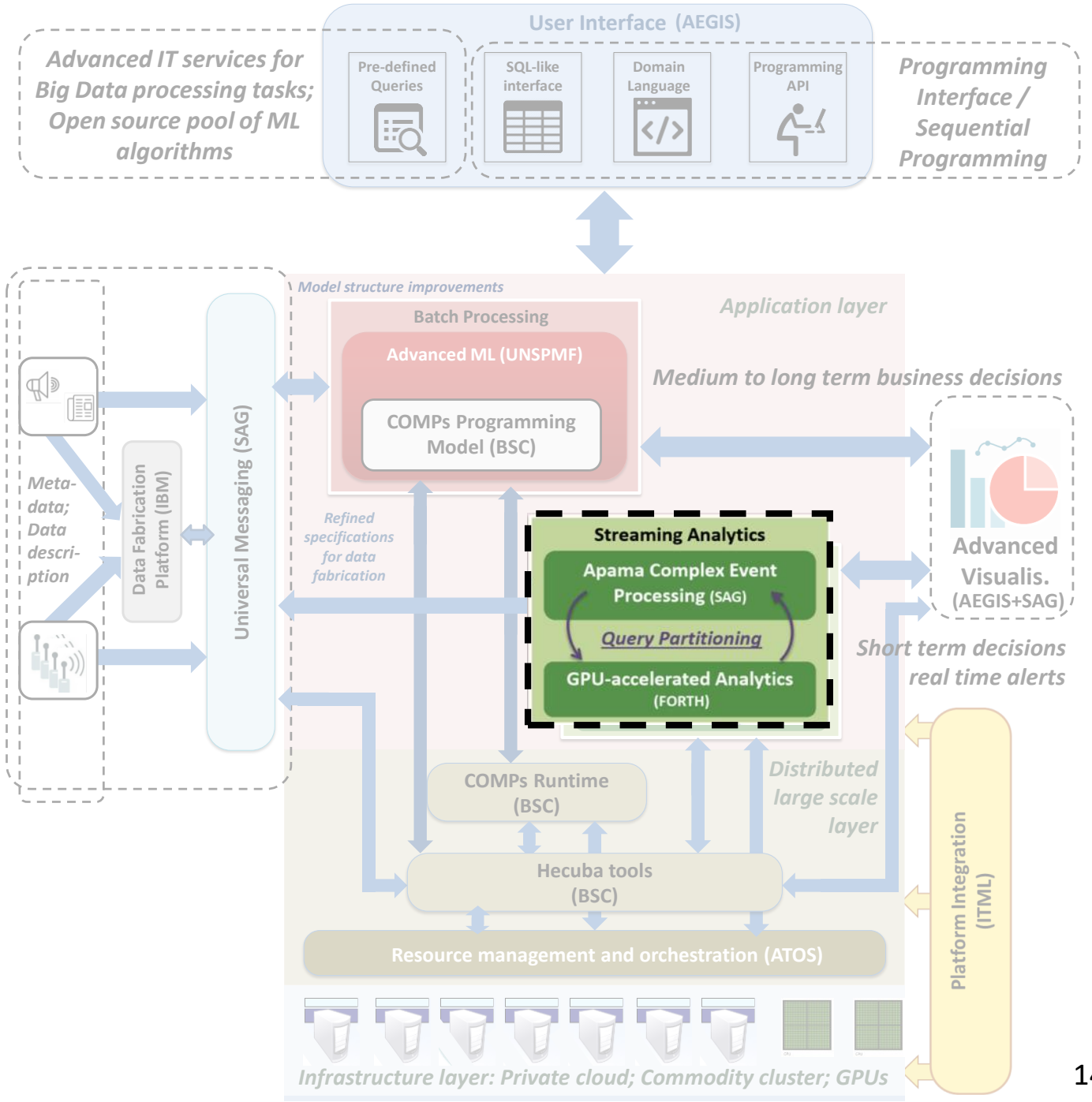


http://ibidaas.eu/tools

WP4:
Streaming analytics

Technologies:

- SAG Apama CEP
- FORTH GPU-accel. analytics

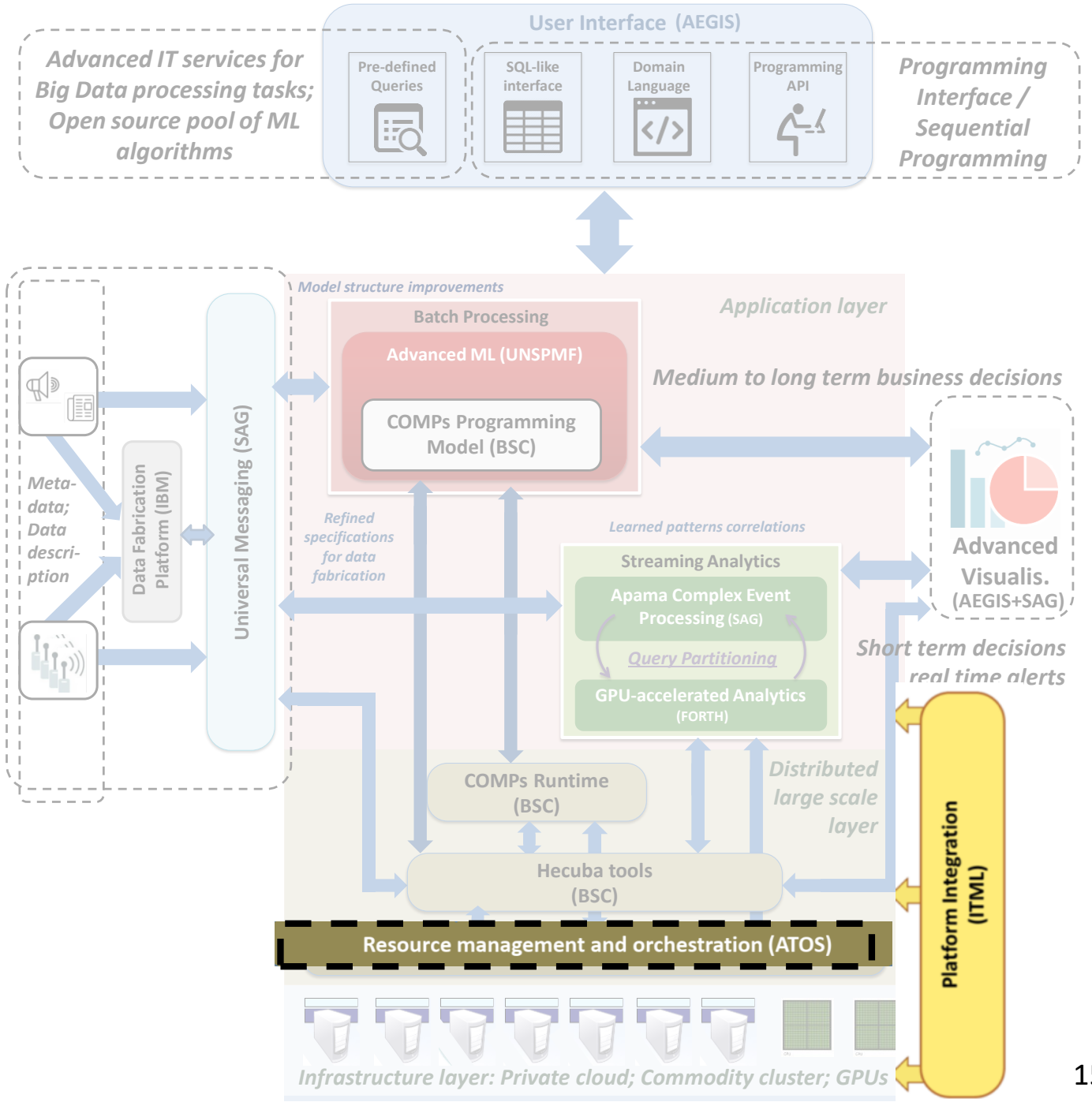


http://ibidaas.eu/tools

WP5:
Resource mgmt & integration

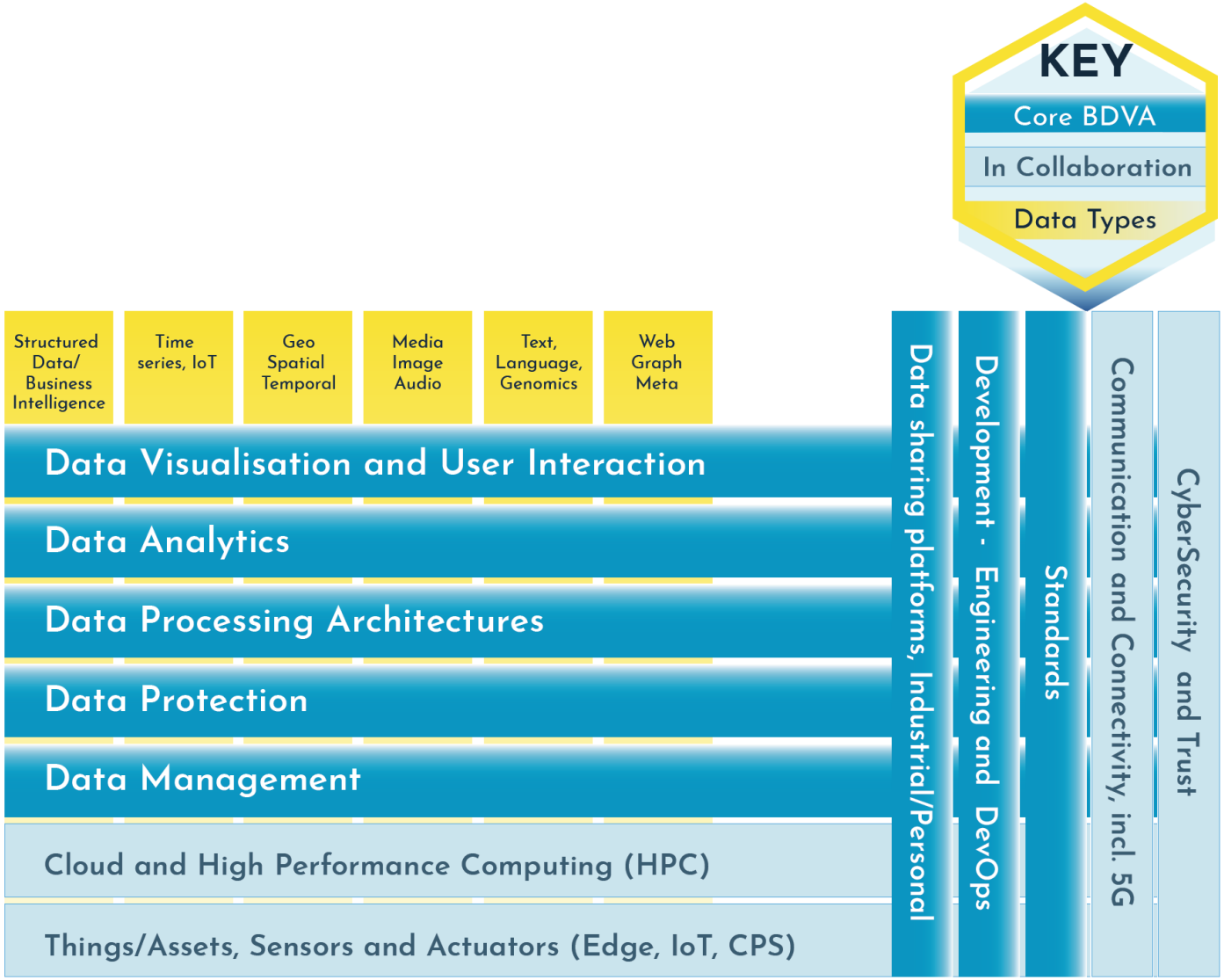
Technologies:

- ATOS Resource mgmt
- ITML integration services



http://ibidaas.eu/tools

BDVA Reference model



BDV SRIA: European Big Data Value Strategic Research and Innovation Agenda

BDVA reference model horizontal concerns & I-BiDaaS

BDV reference model horizontal concern	*I-BiDaaS* module or platform as a whole
Data visualization and user interaction	Advanced visualization module (AEGIS+SAG); User interface (AEGIS)
Data analytics	Batch processing module (UNSPMF+BSC); Streaming analytics module (SAG+FORTH)
Data processing architectures expected advances according to BDVA SRIA	*I-BiDaaS* platform
Data protection	FORTH commodity cluster privacy preservation through commodity hardware (Intel SGX); TDF (IBM) for generation of realistic synthetic data when real data cannot be uploaded to cloud or similar systems
Data management	COMPSs runtime (BSC); ATOS resource management and orchestration module
The Cloud and HPC	(efficient usage of Cloud) ATOS resource management and orchestration module

Outline

- Brief introduction to I-BiDaaS
- I-BiDaaS pipeline, architecture & technologies
- I-BiDaaS & BDVA reference model
- **Benchmarking landscape and opportunities @ I-BiDaaS**

Benchmarking: Technology level

I-BiDaaS partner	Technology name	Big Data pipeline element	Current benchmarks
FORTH	GPU accelerator technology	Data pre-processing, Streaming Analytics	Custom benchmark (throughput, latency)
BSC	COMPSs	Sequential programming model for distributed architectures	Applications (Own use cases)
BSC	Hecuba	Data management framework with easy interface	Applications (Own use cases)
BSC	**Qbeast**	**Multidimensional indexing and storage**	**TCP-H**
IBM	Test Data Fabrication	Synthetic test data fabrication	Several open source + commercial products (e.g., Grid tools of CA) / No known benchmarks yet
SAG	Apama Streaming Analytics Platform	Streaming Analytics	Custom benchmark (throughput)
SAG	Universal Messaging	Message Broker	Custom benchmark (throughput)
SAG	WebMethods Integration Platform	Integration	N/A
SAG	MashZone	Visualization	N/A
AEGIS	Advanced visualization and monitoring	Visualization and interface	N/A
UNSPMF	**Pool of ML algorithms in COMPSs/Python**	**Batch analytics**	**Respective MPI implementation; Sklearn**
ATOS	Resource management and orchestration module	Resource management	N/A

Benchmarking: Business, data & analytics level

	Business Objectives	Data Sets	Data Size	Processing Type	Type of Analysis
Telecoms	- improve and optimize current operations	- Anonymized mobility data (structured) - Anonymized call center data (unstructured)	TB	- batch & streaming	- predictive - descriptive / diagnostic
Finance	- improve decision making - improve efficiency of Big Data solutions	- Tokenized online banking control data (structured) - Tokenized bank transfer data (structured) - Tokenized IP address data (structured)	PB	- batch - batch & streaming	- descriptive / diagnostic
Manufacturing	- improve and optimise current operations - improve the quality of the process and product	- Anonymized SCADA/MES data (structured) - Anonymized Aluminum Die-casting (structured)	GB	- batch - batch & streaming	- predictive - diagnostic

Benchmarking: Business level

I-BiDaaS Partner	Use Case	Most relevant business KPIs
TID	Accurate location prediction with high traffic and visibility	- Acquisition of insights on the dynamics of cellular sectors **- Processing costs (cost reduction)** **- Customer satisfaction**
TID	Optimization of placement of telecommunication equipment	
TID	Quality of service in Call Centers	
CAIXA	Enhanced control on online banking	**- Cost reduction** - Data accessibility **- Time efficiency** - End-to-end execution time (from data request to data provision)
CAIXA	Advanced analysis of bank transfer payment in financial terminal	
CAIXA	Analysis of relationships through IP addresses	
CRF	Production process of aluminium die-casting	- 3 **Product quality** levels (High. Medium, Low) - Overall Equipment Effectiveness (OEE), - Maintenance cost **- Cost reduction**
CRF	Maintenance and monitoring of production assets	