bdva.eu
bigdatastack.eu
ibidaas.eu
trackandknowproject.eu
policycloud.eu

@BDVA_PPP
@BigDataStackEU
@Ibidaas
@Track&Know
@PolicyCloudEU

Big Data Pilot Demo Days
I-BiDaaS Application to the Telecomunications Sector

Despina Kopanaki (FORTH) dkopanaki@ics.forth.gr

Marieke Willems (Trust-IT) m.willems@trust-itservices.com

Andrea Schillaci (Trust-IT) a.schillaci@trust-itservices.com

Diego Simoni (Trust-IT) d.simoni@trust-itservices.com

Dusan Jakovetic (UNSPMF) dusan.jakovetic@dmi.uns.ac.rs

Ioannis Arapakis (TID) ioannis.arapakis@telefonica.com

The below-mentioned projects have been funded by the European Commission Horizon 2020
BigDataStack: grant agreement No 779747
I-BiDaaS: grant agreement No 780787
Track and Know: grant agreement No 780754
Policy Cloud: grant agreement No 870675

Webinar Speakers

Assistant Professor at the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Serbia

I-BiDaaS Scientific & Technical Manager.

Dr. Dušan Jakovetić
University of Novi Sad, Serbia

Researcher at Telefonica, Spain

Dr. Ioannis Arapakis
Telefonica

Agenda

- Big Data Pilot Demo Days. A joint effort by BigDataStack, I-BiDaaS, Track & Know & Policy Cloud – Marieke Willems (Trust-IT, Policy Cloud)
- I-BiDaaS overview – Dusan Jakovetic (UNSPMF)
- TID's Pitches: Setting the requirements – Ioannis Arapakis (TID)
- I-BiDaaS architecture: Scientific & Technical view; how it addresses the requirements set by TID – Dusan Jakovetic (UNSPMF)
- Step by Step demonstration of I-BiDaaS solution and its application to telecommunications sector – Ioannis Arapakis (TID)
- Questions & Answers

Questions

- **To which of our stakeholder types do you belong?**
  (Big Data Provider, Big Data Technology Provider, Telecommunications, Research & Academia, Policy Maker, Standardisation Body, other)

- **Are you working with Big Data?**
  (Yes, No)

- **Are you interested in Big Data Technologies to optimize your customer experience?**
  (Yes, No, Maybe)

- **What is the main barrier or risk preventing you from implementing Big Data analytical solutions in your organization?**
  (Costs, Lack of expertise, Uncertain Value (ROI))

BDV PPP Summit 2020 went virtual

Why Big Data Pilot Demo Days?

- The new data-driven industrial revolution highlights the need for big data technologies to unlock the potential in various application domains.
- BDV PPP projects I-BiDaaS, BigDataStack and Track & Know and Policy Cloud deliver innovative technologies to address the emerging needs of data operations and applications.
- To fully exploit the sustainability of the developed technologies, the projects onboarded pilots that exhibit their applicability in a wide variety of sectors.
- In their third and final year, the projects are ready to demonstrate the developed and implemented technologies to interested end-users from industry as well as technology providers, for further adoption.
- The recently started Policy Cloud project will highlight the adoption of technologies developed by the more mature BDV PPP project BigDataStack, showcasing its application for the policy making sector

BDV PPP Projects Join Forces

Holistic stack for big data applications and operations

Industrial-Driven Big Data as a Self-Service Solution

Big Data for Mobility Tracking Knowledge Extraction in Urban Areas

Cloud for Data-Driven Policy Management

BDV PPP projects joining forces to showcase application of innovative technologies in a variety of domains, fostering futher adoption, contributing to Europe's digital future.

bdva.eu
bigdatastack.eu
Ibidaas.eu
trackandknowproject.eu
policycloud.eu

@BDVA_PPP
@BigDataStackEU
@Ibidaas
@Track&Know
@PolicyCloudEu

Big Data Pilot Demo Days - A Series of Webinars

1. I-BiDaaS Application to the Financial Sector
2. BigDataStack Connected Consumer demo
3. A BigDataStack Seafarer's Tale on Real Time Shipping
4. I-BiDaaS Application to the Telecommunication Sector
5. Track & Know applied to the fleet management sector
6. Policy Cloud: Data Driven Policies against Radicalisation
7. I-BiDaaS Application to the manufacturing sector
8. Track & Know applied to the automotive insurance sector
9. BigDataStack Smart Insurance
10. Track & Know applied to patient mobility in the healthcare sector

BIG DATA VALUE PPP

BigDataStack
Holistic stack for big data applications and operations

I-BiDaaS
Industrial-Driven Big Data as a Self-Service Solution

Policy Cloud
Cloud for Data-Driven Policy Management

Track & Know

bdva.eu
bigdatastack.eu
ibidaas.eu
trackandknowproject.eu
policycloud.eu

@BDVA_PPP
@BigDataStackEU
@Ibidaas
@Track&Know
@PolicyCloudEU

I-BiDaaS Overview

Dusan Jakovetic

Ass. Professor, University of Novi Sad, Faculty of Sciences, Serbia;
I-BiDaaS Scientific & Technical Manager

I-BiDaaS Application to the Telecomunications Sector

Thursday, June 25, 2020 - 14:00-15:00 CEST

The below-mentioned projects have been funded by the European Commission Horizon 2020
BigDataStack: grant agreement No 779747
I-BiDaaS: grant agreement No 780787
Track and Know: grant agreement No 780754
Policy Cloud: grant agreement No 870675

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

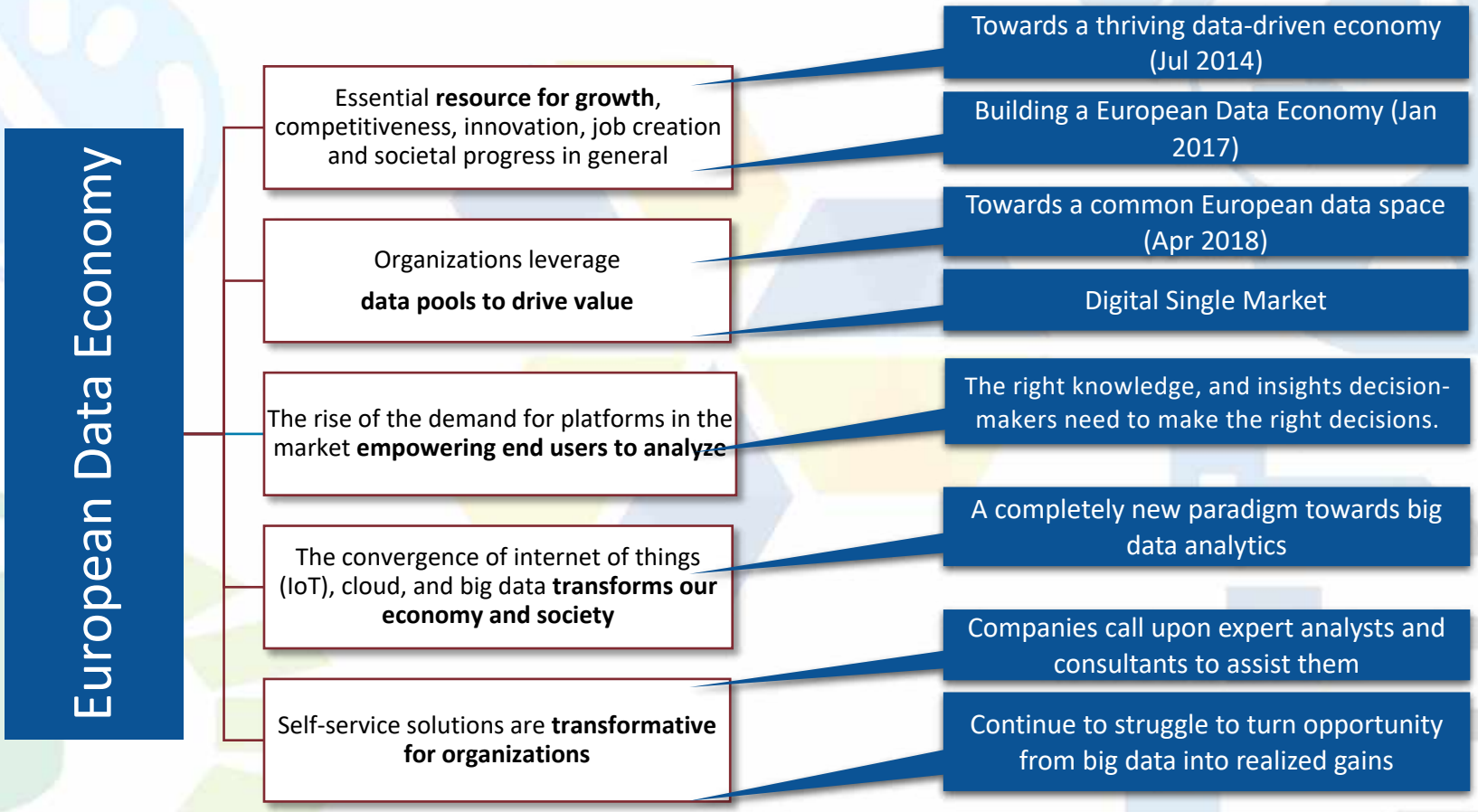
DURATION
36 months

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

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**)

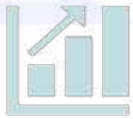
Motivation


European Data Economy
Essential **resource for growth**, competitiveness, innovation, job creation and societal progress in general
Towards a thriving data-driven economy (Jul 2014)
Building a European Data Economy (Jan 2017)
Organizations leverage **data pools to drive value**
Towards a common European data space (Apr 2018)
Digital Single Market
The rise of the demand for platforms in the market **empowering end users to analyze**
The right knowledge, and insights decision-makers need to make the right decisions.
The convergence of internet of things (IoT), cloud, and big data **transforms our economy and society**
A completely new paradigm towards big data analytics
Self-service solutions are **transformative for organizations**
Companies call upon expert analysts and consultants to assist them
Continue to struggle to turn opportunity from big data into realized gains

Our Vision

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 breaking industrial data 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** including elasticity

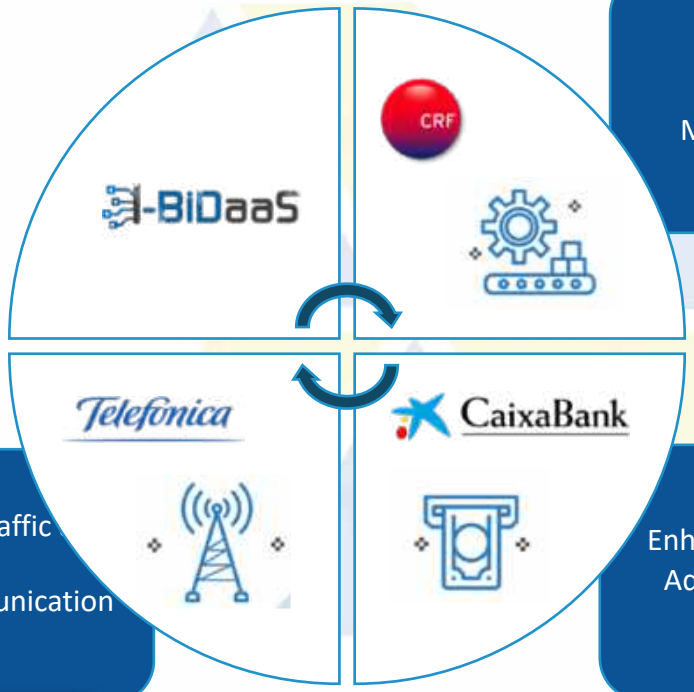
Project Statement

I-BiDaaS *aims to* ***empower*** *users to easily* ***utilize and interact*** *with* ***big data technologies****, by designing, building, and demonstrating, a* ***unified framework*** *that:*

significantly increases the ***speed of data analysis*** *while coping with the rate of* ***data asset growth****, and facilitates* ***cross-domain data-flow*** *towards a thriving* ***data-driven EU economy****.*

I-BiDaaS will be tangibly validated by three real-world, ***industry-lead*** *experiments.*

Application / Experimentation



CRF

Production process of aluminium casting

Maintenance and monitoring of production assets

Manufacturing Industry

TID

Accurate location prediction with high traffic visibility

Optimization of placement of telecommunication equipment

Quality of Service in Call Centers

Telecommunications Industry

CAIXA

Enhance control of customers to online banking

Advanced analysis of bank transfer payment in financial terminal

Analysis of relationships through IP address

Banking/Finance Industry

I-BiDaaS application domains

Big Data Pilot Demo Days

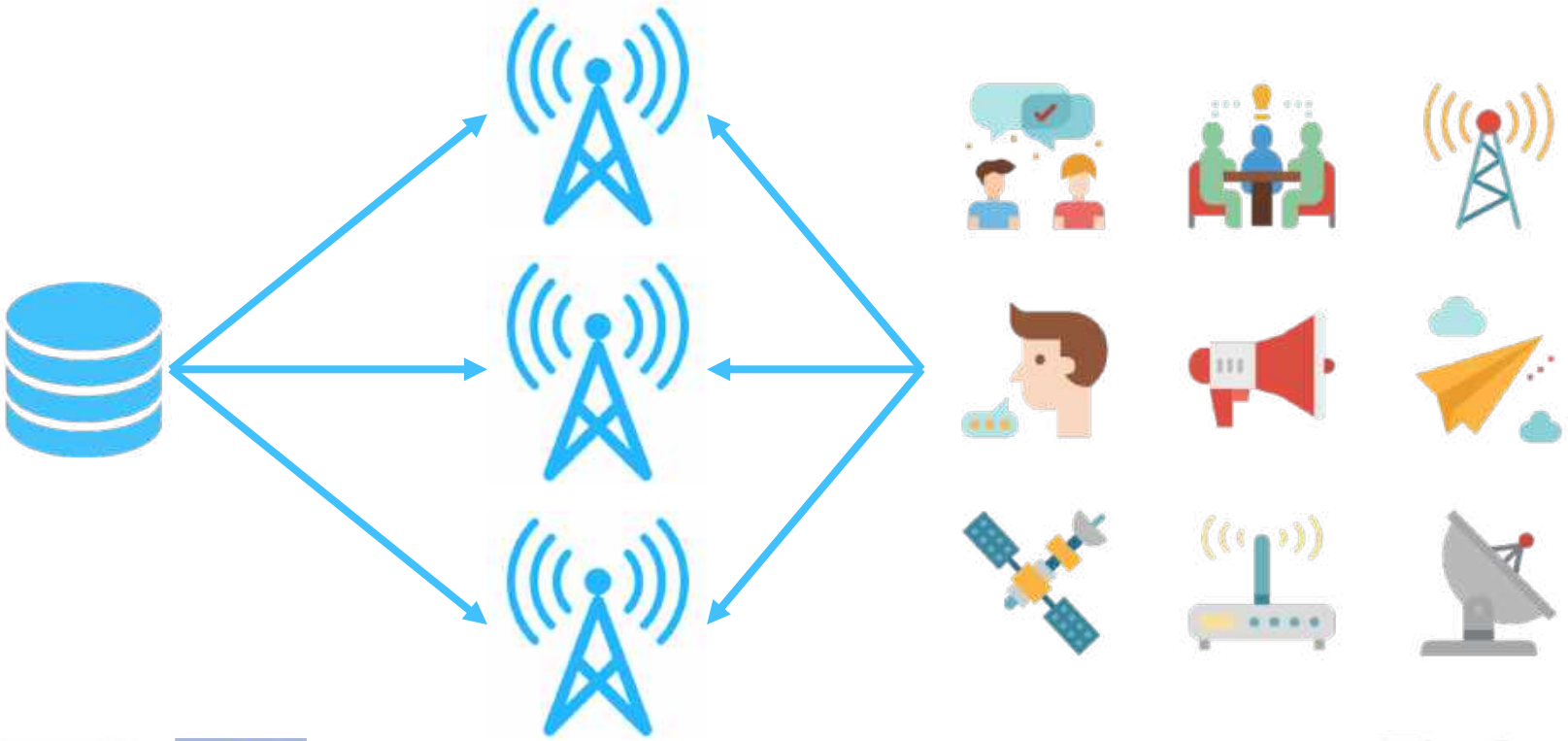
I-BiDaaS Application to Telecommunications Sector

Arapakis, Ioannis (Telefonica Research)

Telefonica in Numbers

- Telefonica is one of the largest telecommunications companies in the world by market capitalization and number of customers
- It provides quality of connectivity that is delivered over world class fixed, mobile and broadband networks

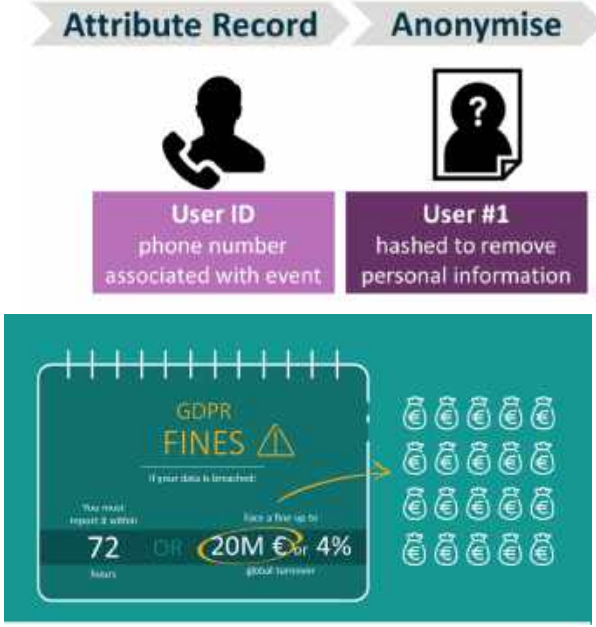
Data Landscape



Industrial Challenges

	Record Event	Anonymise	Parse	Transform	Aggregate
Event Data					
Example	User - +447872020200 Cell tower– 10532 Area location – 9823 Time – 09:01:43	User – cg5DrF872Ca Cell tower– 10532 Area location – 9823 Time – 09:01:43	User – 1147562 SS Cell ID– 583013 Time – 09:01:43	User – 1147562 Location – 10010034 Start Time – 8:15 End Time – 12:30	Client – Company A Site – Stockholm Date – 08/07/16 Hour – 11:00 – 12:00 Visitor count – 3410

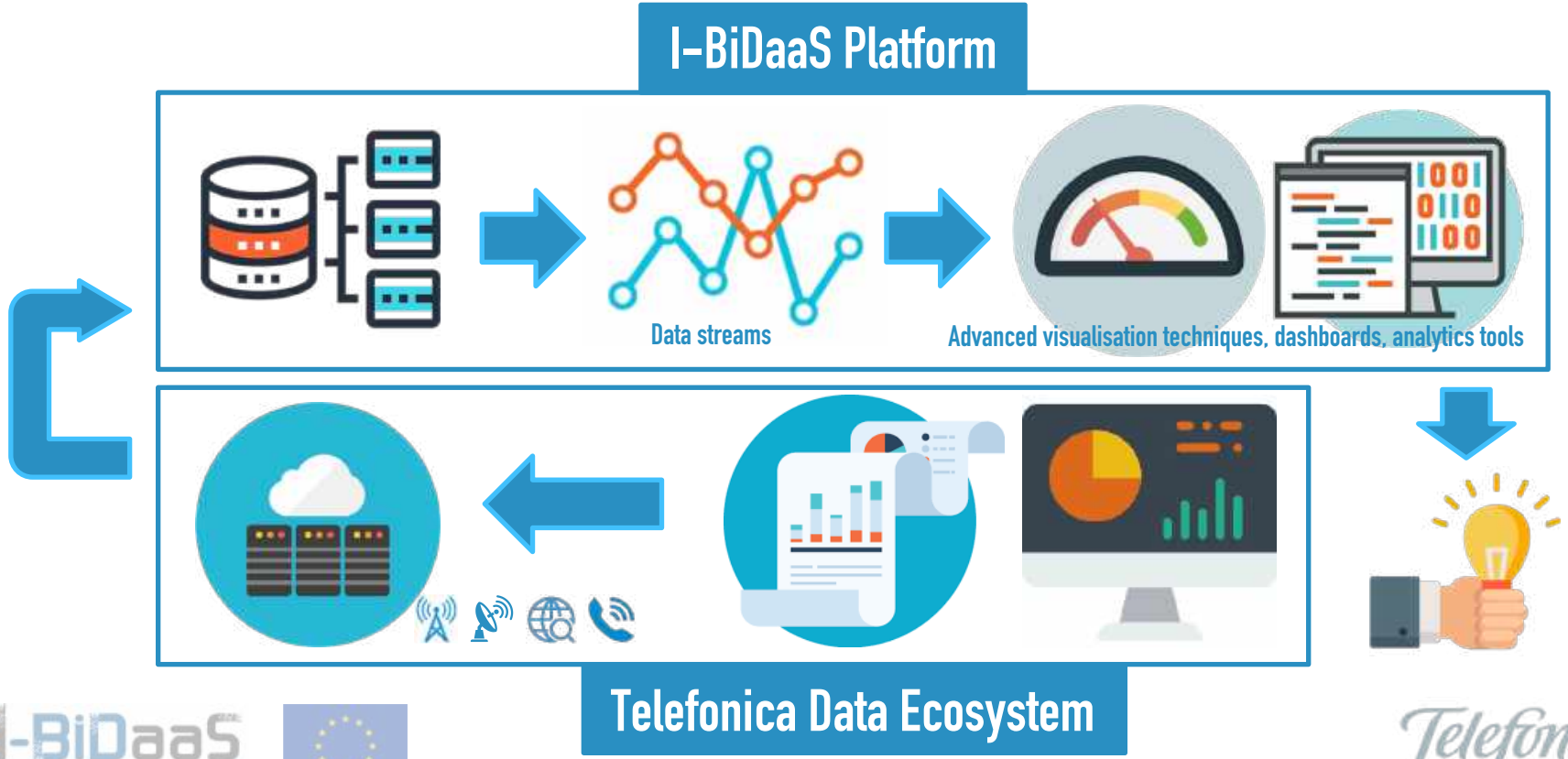
Industrial Challenges



Attribute Record	Anonymise	Parse	Privacy
User ID phone number associated with event	**User #1** hashed to remove personal information	**User #2** new identifier unrelated to previous	**Privacy Module** Remove completely/ Remove attributes

Protect from Re-identification:	**Categorised** Age rounded to bands	**Truncated** (postcodes etc.) Protect from re-identification
Reduce confidence in re-identification:	**Noise Addition** 10% +/- in 1 category	**Policy & Code of Conduct** TEF policy, logging etc

The I-BiDaaS Solution



I-BiDaaS: Advantages of (Synthetic & Real) Data

- The sensitive nature of real-life industrial datasets introduces substantial **privacy** and **security risks**
- TID assisted in the **synthetization of** realistic data that:
  - Mimic a real dataset
  - Facilitate the early exploration and development phases in I-BiDaaS

- To break the **inter-** and **intra-sectorial data-silos**, and support **data sharing**, **exchange**, and **interoperability**, TID is also providing in-house access to real-life datasets
- Involve different business units and external companies, for interfacing and exploring novel data analytic technologies

Expected Outcomes

- **Embed** I-BiDaaS innovative research into relevant use cases and products, and **improve** on the corresponding **KPIs** (e.g., better quality of service, optimized antenna placement, etc.)
- Motivate secure **data sharing** both **internally** and **externally**, and break silos by interfacing with other companies
- Develop technologies and tools that support a **highly-dynamic** and **complex** data analysis and empower both **expert** and **non-expert** big data practitioners

I-BiDaaS
Industrial-Driven Big Data as a Self-Service Solution

I-BiDaaS Application to the Telecommunications Sector

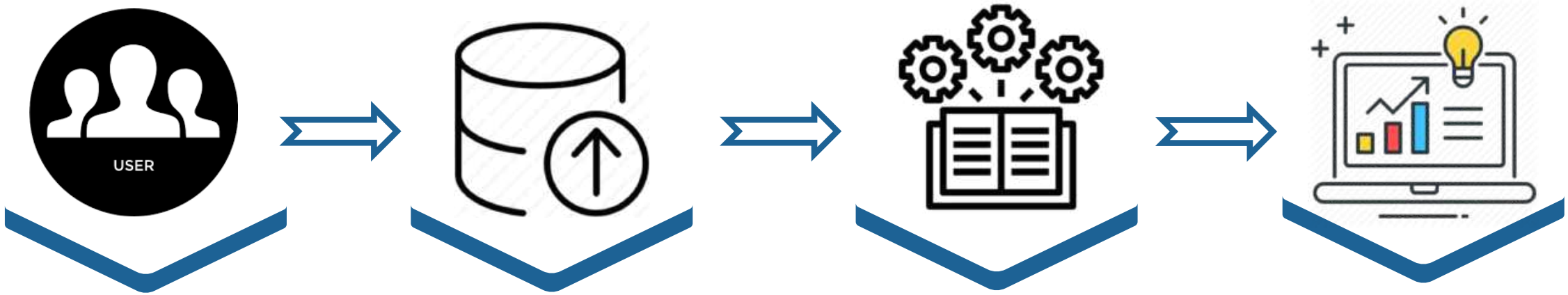
Big Data architecture

Dusan Jakovetic

University of Novi Sad, Faculty of Sciences

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

The I-BiDaaS solution: Front-end



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

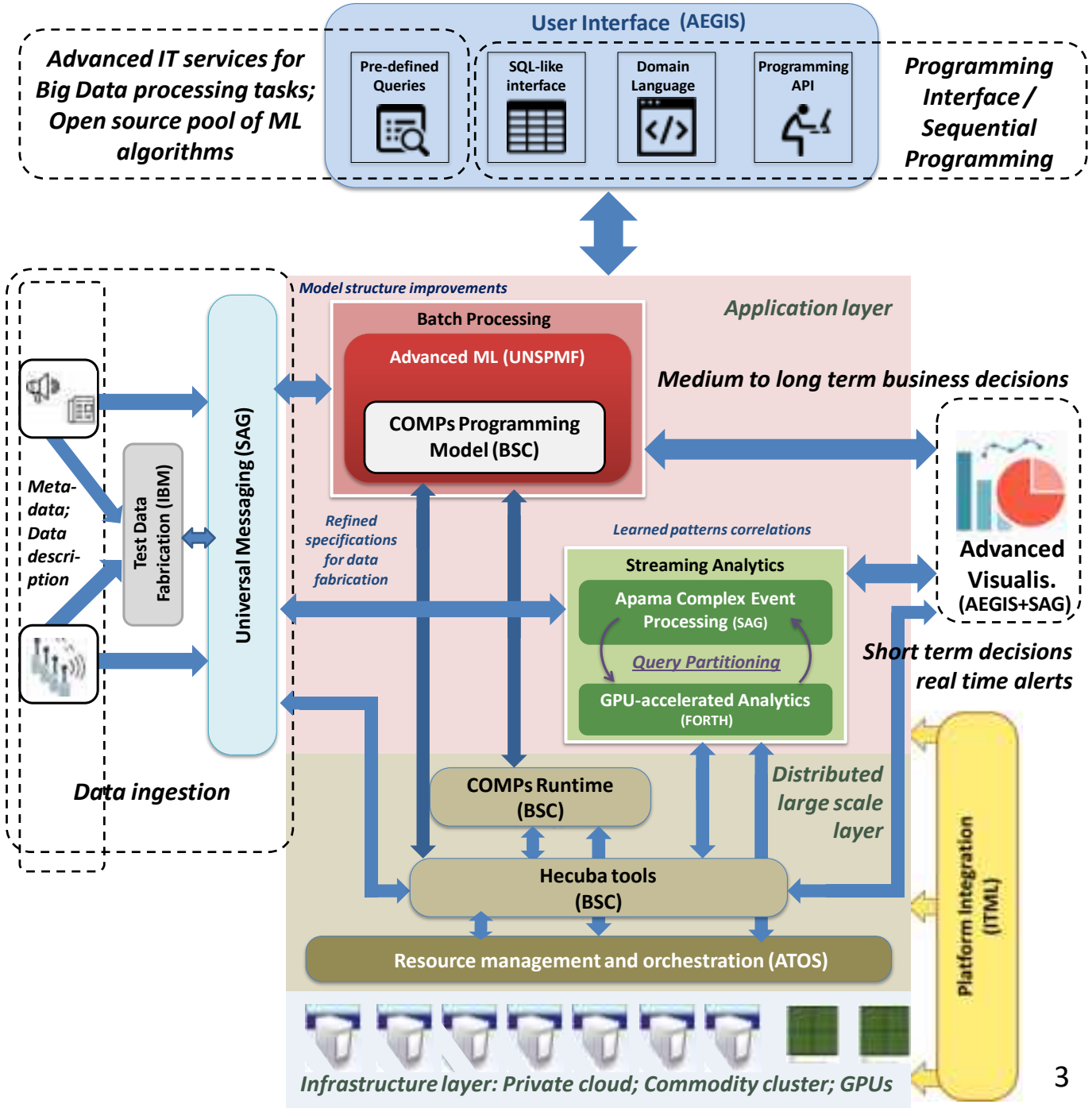
Interact with Big Data technologies

Increase speed of data analysis

Intra- and inter-domain data-flow

Cope with the rate of data asset growth

The I-BiDaaS solution: Architecture/back-end

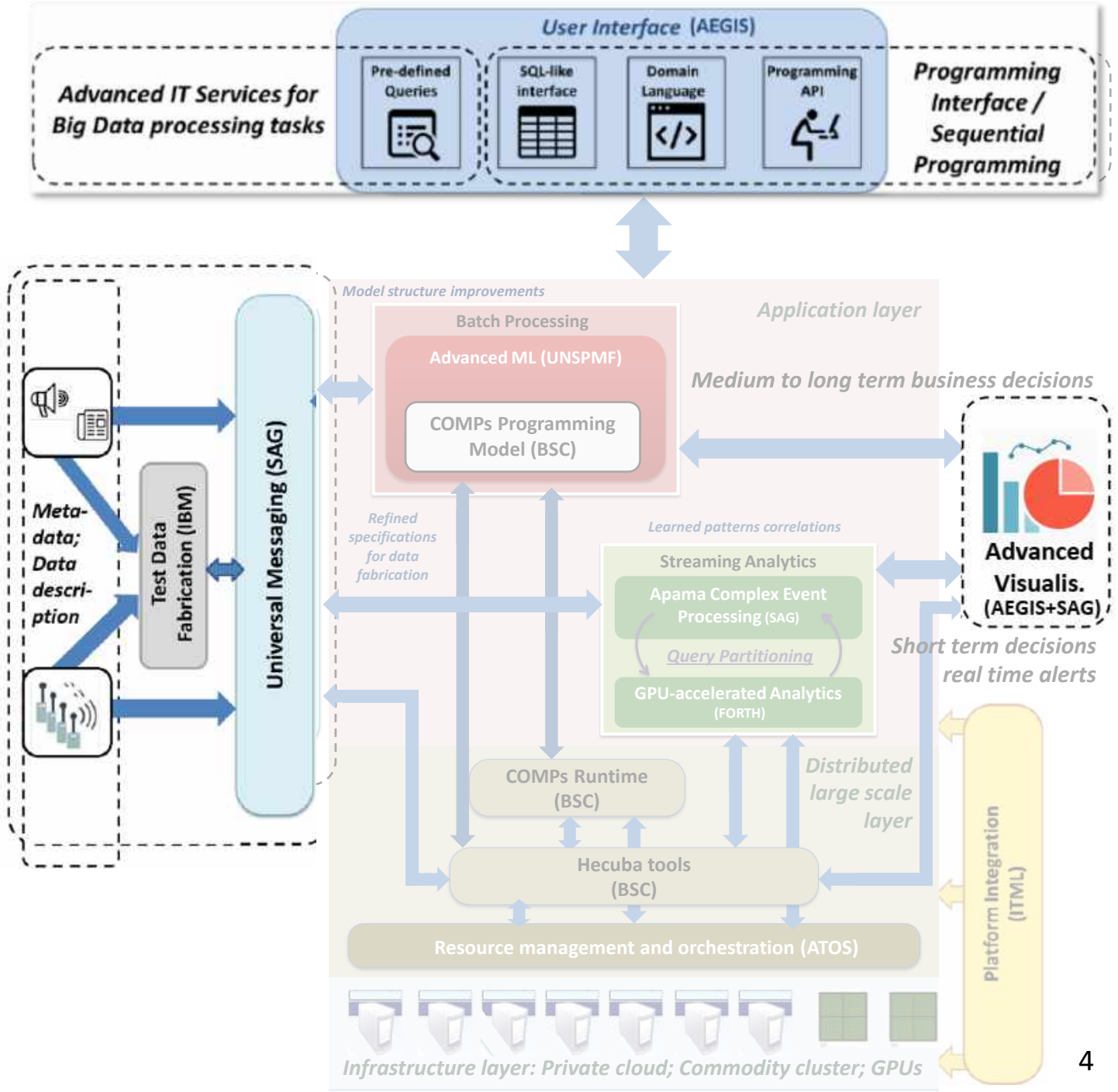


WP2:

Data, user interface, visualization

Technologies:

- IBM TDF
- SAG UM
- AEGIS AVT

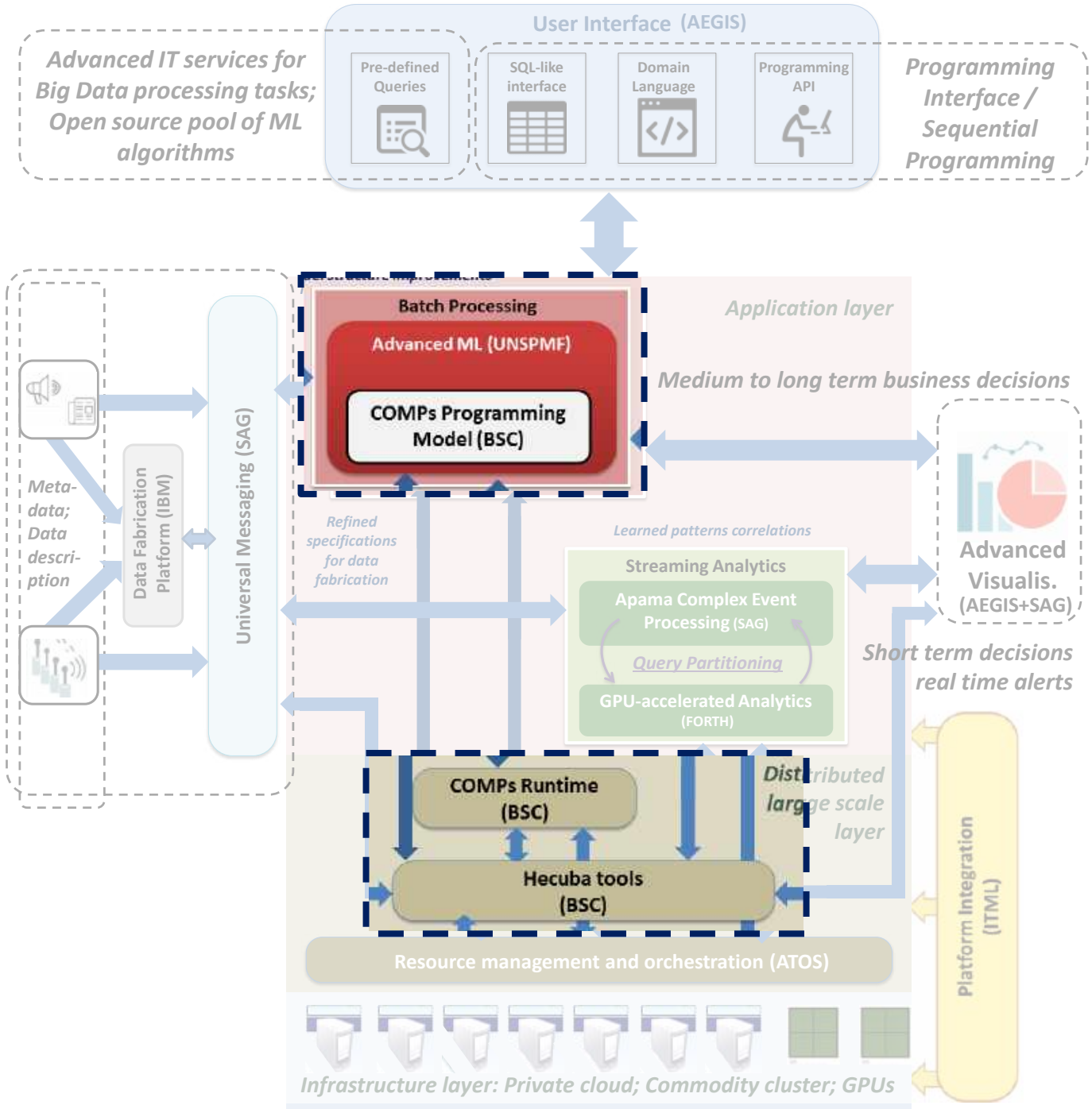


http://ibidaas.eu/tools

WP3:
Batch analytics

Technologies:

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

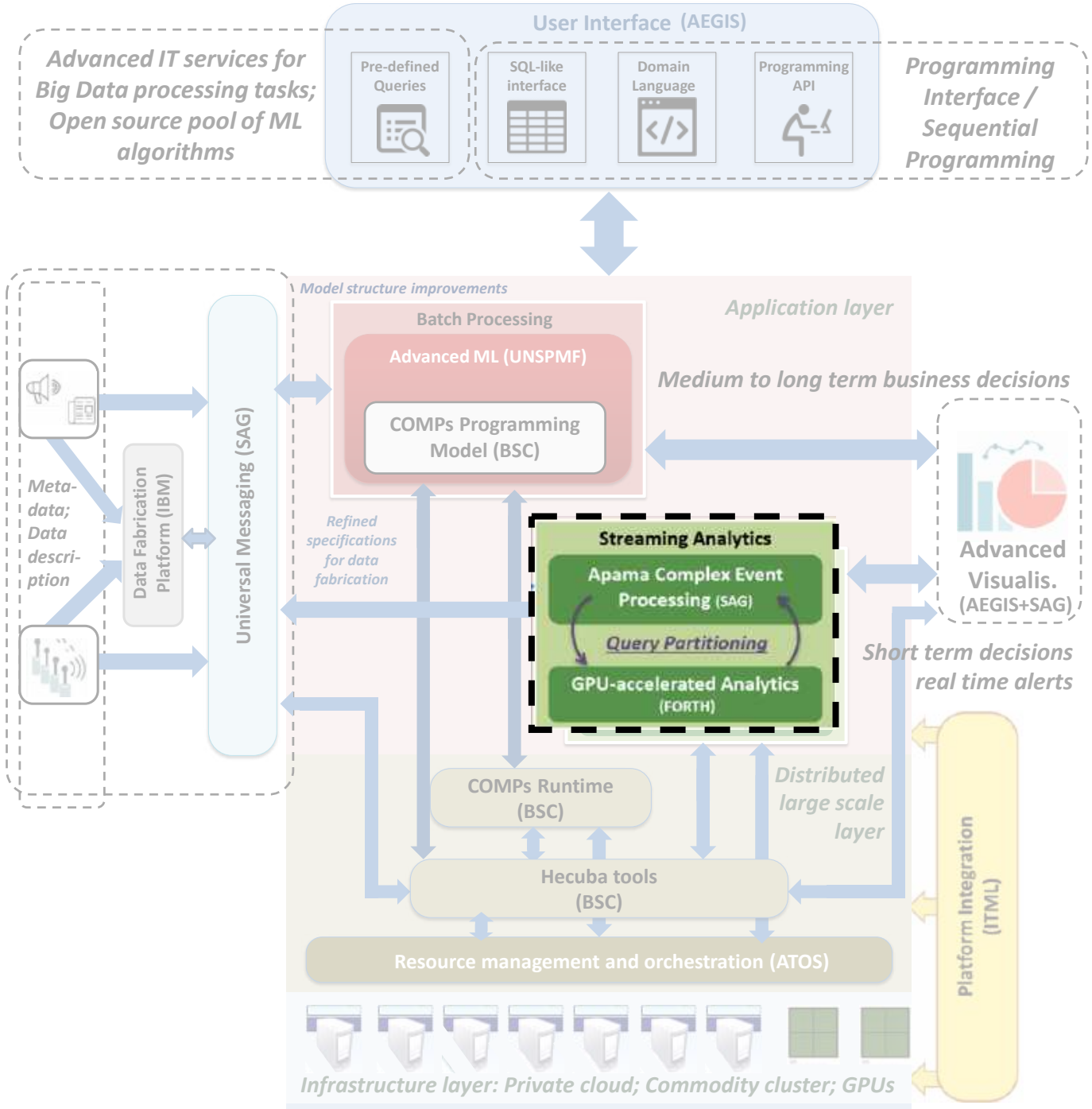


http://ibidaas.eu/tools

WP4:
Streaming analytics

Technologies:

- SAG Apama CEP
- FORTH GPU-accel. analytics



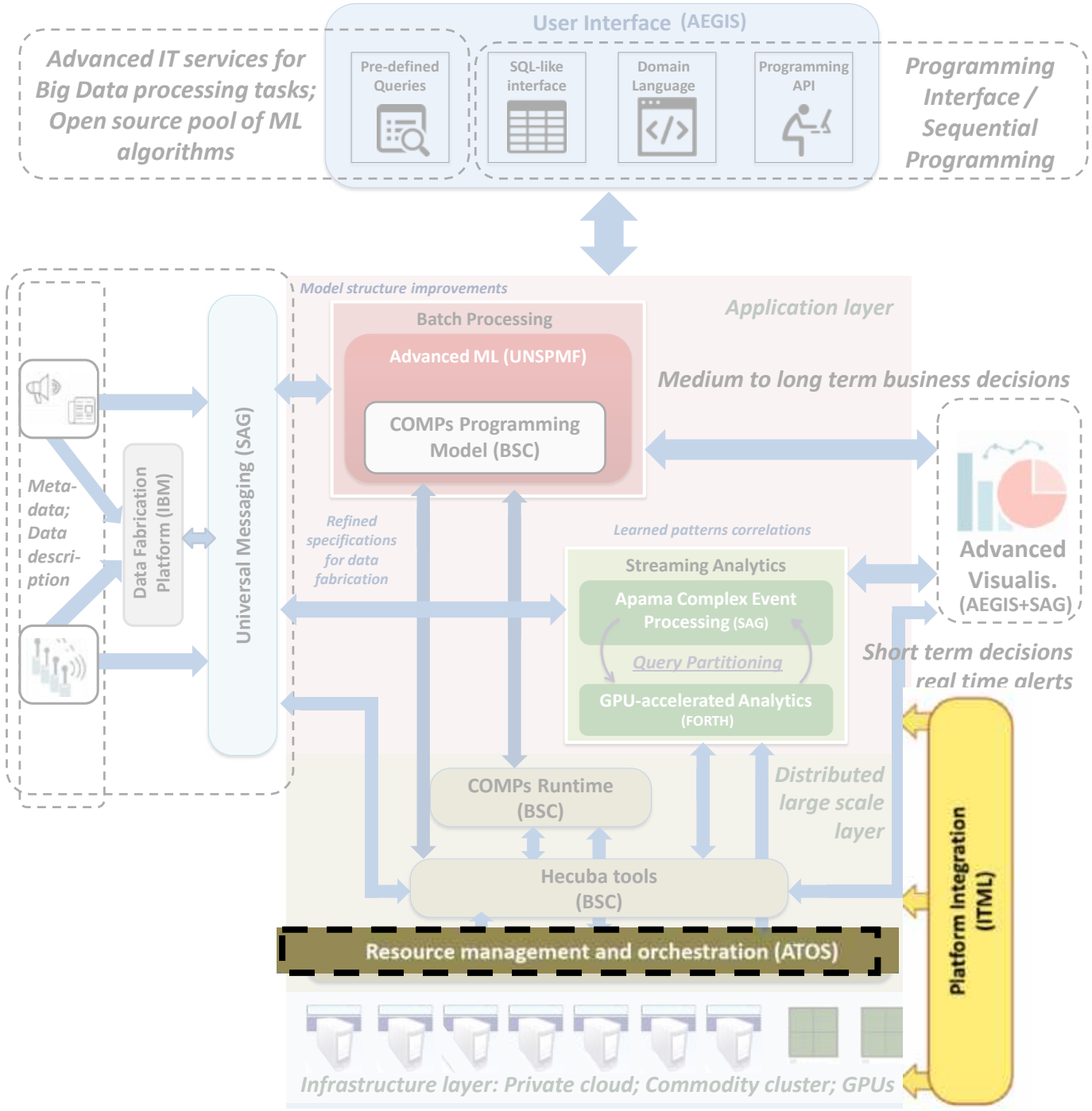
http://ibidaas.eu/tools

WP5:

Resource mgmt & integration

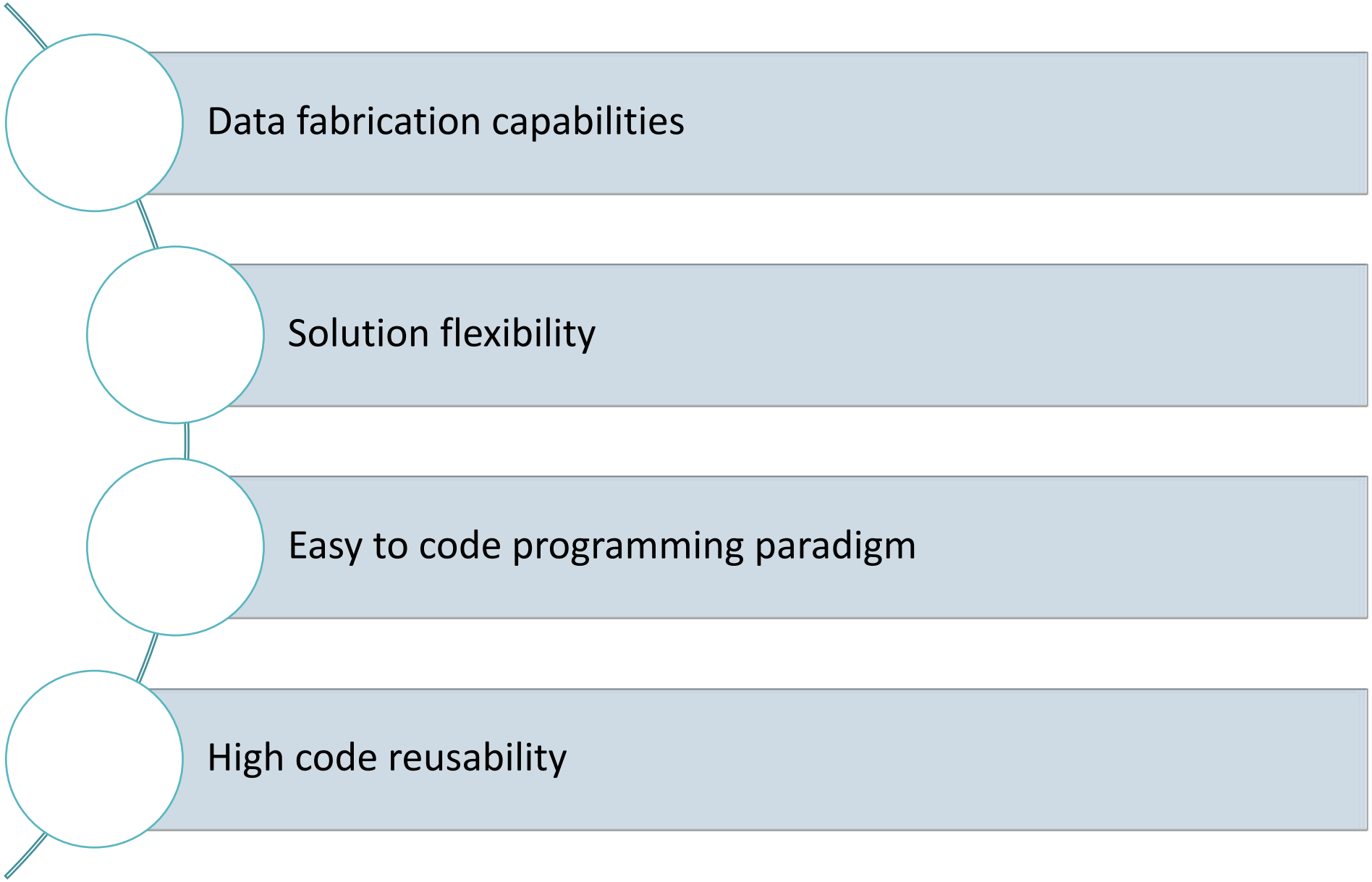
Technologies:

- ATOS Resource mgmt
- ITML integration services



http://ibidaas.eu/tools

Key features & innovations



Key features & innovations (Cont'd)

- GPU-accelerated analytics; Synergy of CEP and GPU-accelerated analytics for streaming data
- Feedback from analytics to data fabrication
- Feedback from analytics to problem modelling
- Demonstrated on use cases across 3 different data providers and 3 different industries

I-BiDaaS applied to TID use cases

- **Antenna KPIs**
  - Operators need info which sectors underperform at any given time
  - Can we predict with accuracy which antennas will become the next “hot spots”?

Antenna KPIs

- **Challenges**
  - Over 40 thousands of antennas; 16 KPIs per hour during one day
  - **Highly imbalanced dataset** – 0.012% of positive class
  - “Data leakage” when splitting data using standard train/test split

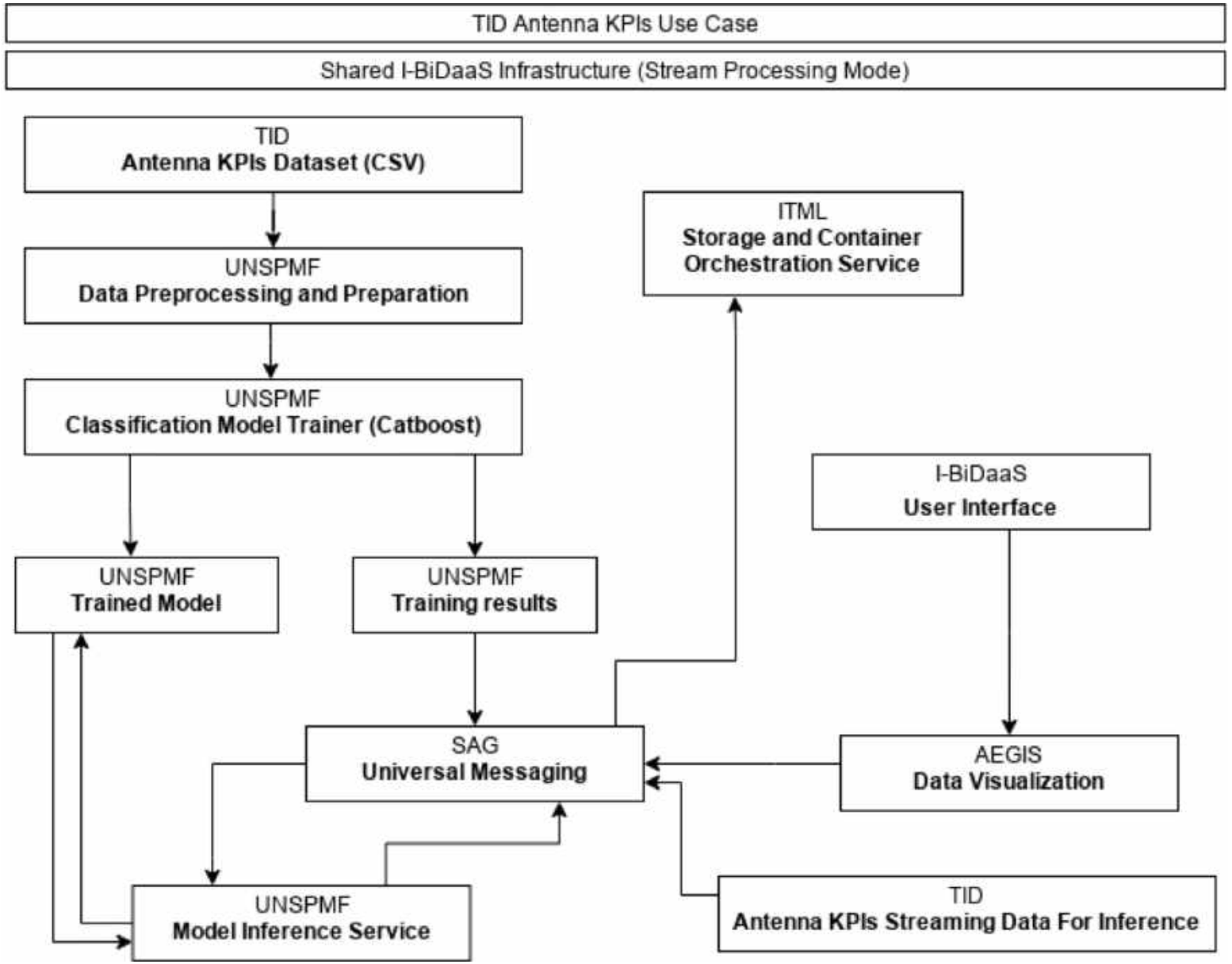
- **I-BiDaaS approach**
  - State-of-the art fast & accurate parallel classifiers
  - Random Forest: **COMPSs;** Scikit-Learn
  - Gradient boosting: XGBoost & CatBoost libraries

Antenna KPIs: I-BiDaaS results

- **High accuracy, High efficiency**

Algorithm	Accuracy	Precision	Recall
XGBoost	0.999	0.998	0.998
CatBoost	0.999	0.977	0.961
Scikit-Learn Random Forest	0.999	0.998	0.975
PyCOMPSs Random Forest	0.999	0.995	0.997

I-BiDaaS solution specialized to Antenna KPIs



I-BiDaaS applied to TID use cases

- **User Mobility:**
  - Users travel around the city and possibly create congestions
  - Can we predict when and where the congestions will appear?
  - Can we predict movements at scale?

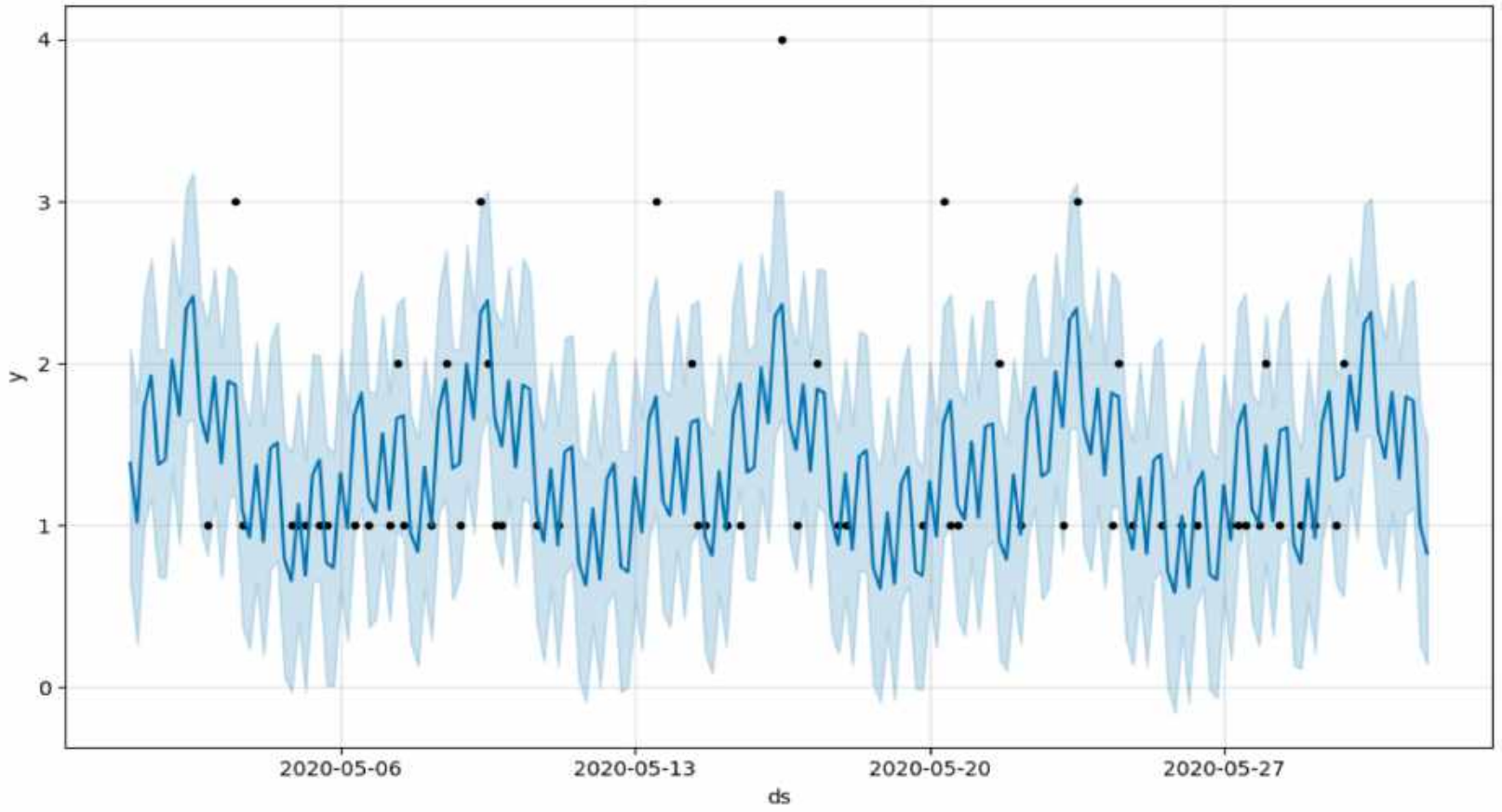
User mobility

- **Challenges**
  - High volume dataset
  - **A lot of missing data**: 85% of data is missing

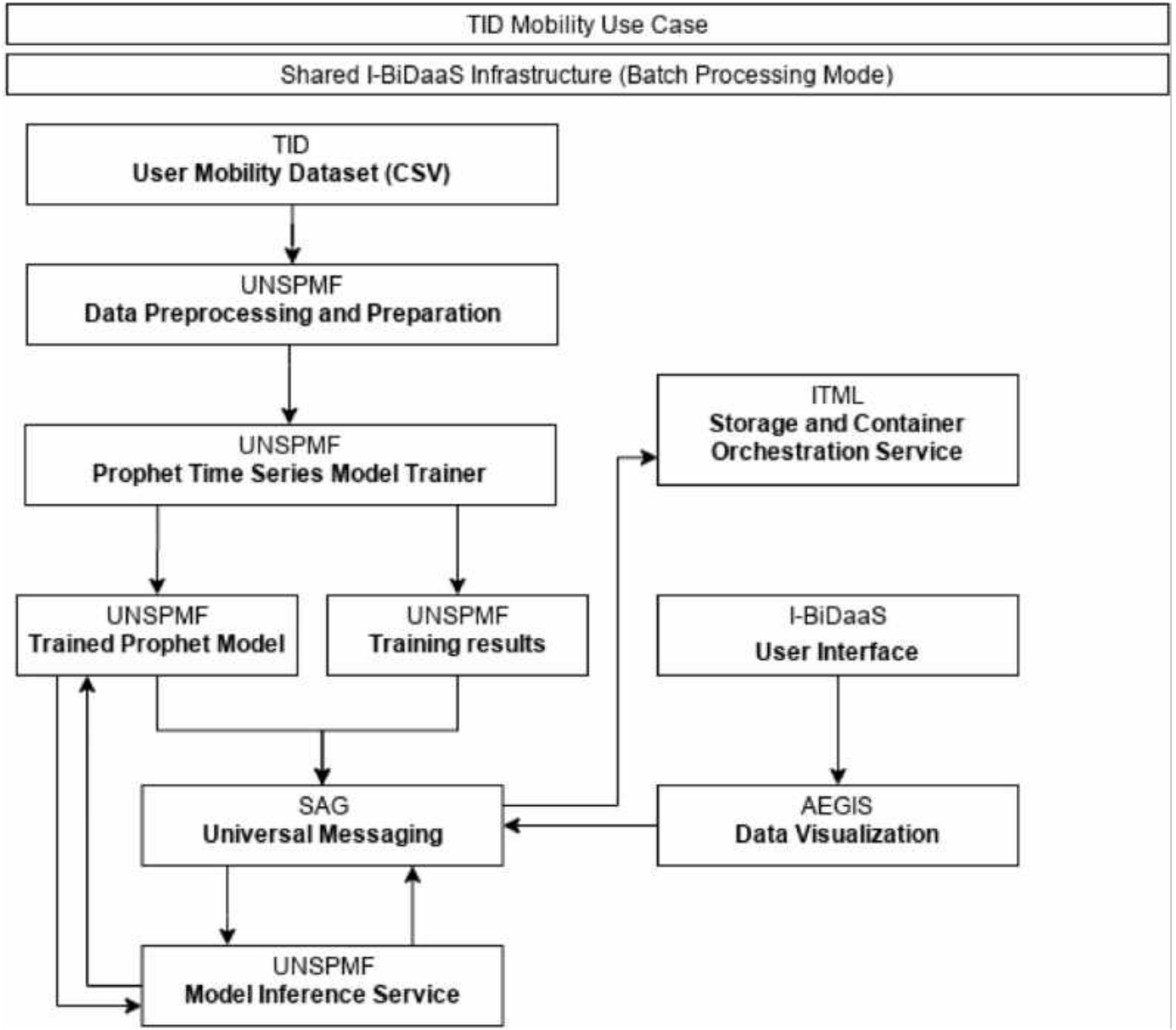
- **I-BiDaaS approach**
  - Pandas library for preprocessing
  - Time-series approach to modeling (Facebook Prophet)
  - Parallelization of training across antennas (Joblib)

User mobility: I-BiDaaS results

- **Good accuracy:** Average MAE on top 1000 models was 1.2565
- **High efficiency:** Trained 100K time series within 3h on 64 cores CPU



I-BiDaaS solution specialized to user mobility



I-BiDaaS applied to TID use cases

- **Call centers**
  - Users calling to complaint, to follow-up with problem, to provide feedback
  - Quickly get familiar and understand customer's perspective
  - Shorten call duration and waiting time

Call centers

- **Challenges**
  - Provide real-time insights in a semantics sense by automatic information retrieval from speech interactions in phone calls
  - Processing of large amount of natural language data
  - Streaming analytics

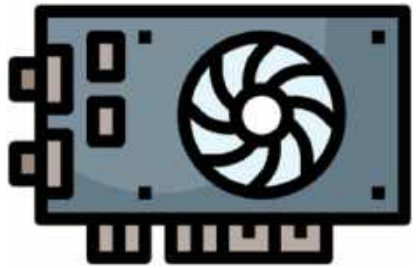
- **I-BiDaaS approach**
  - GPU-accelerated text matching
  - Sentiment score estimation per word/call as it correlates with Customer Satisfaction Index[1]
  - Top-K frequent words or 2-grams as it provides a quick overview of the CC current scenario and operative
  - Matching the keywords in the incoming data stream

[1] Luque, J., Segura, C., Sánchez, A., Umbert, M., & Galindo, L. A. (2017). The Role of Linguistic and Prosodic Cues on the Prediction of Self-Reported Satisfaction in Contact Centre Phone Calls. In *INTERSPEECH* (pp. 2346-2350)

I-BiDaaS solution to real-time calls processing

transcript.txt:

```
...
Script01_mono 10.15 0.76 pentagons 0.77
Script01_mono 11.21 0.20 betrayals 0.78
Script01_mono 12.11 0.12 wreath 0.95
Script01_mono 13.31 0.30 thrower's 0.61
Script01_mono 14.30 0.51 specialties 0.84
Script01_mono 15.32 0.32 racegoers 0.84
Script01_mono 16.50 0.58 donnybrooks 0.59
Script01_mono 17.22 0.51 shadowy 0.97
Script01_mono 17.96 0.61 Jacksonville 0.76
Script01_mono 18.64 0.21 candle 0.83
Script01_mono 19.32 0.16 Somalia's 0.98
Script01_mono 20.47 0.70 protestation 0.85
...
```



top-5000_2000s.txt:

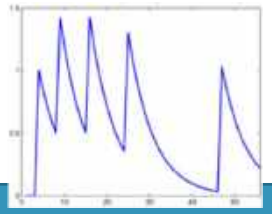
...		
ugly	-3.90	1.16
painful	-3.69	1.53
terrible	-3.38	1.55
...		
delicate	2.72	0.93
beautiful	2.73	0.69
wonderful	2.76	0.71

freqs:

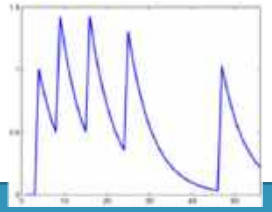
...	
'ugly'	3
'march'	2
'failed'	5
'weak'	2
...	

...
'ugly' found
'march' found
'failed' found
'weak' found
...

$f(t) = 1 + e^{-\lambda\Delta t} * f(t-1)$



$f(t) = 1 + e^{-\lambda\Delta t} * f(t-1)$



Top-K:

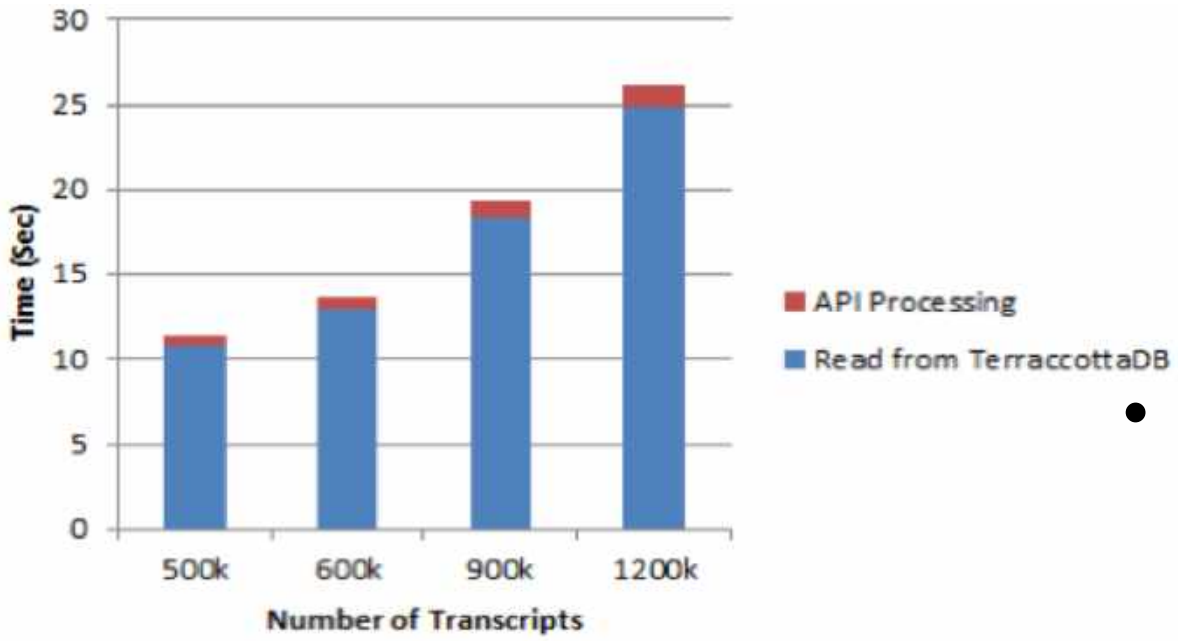
Last min: "weak", "fault", ..
Last hour: "affect", "tired", ...
Last day: "nice", "perfect", ...
Last week: "wondeful", ...

Sentiment Score:

Last min: 79%
Last hour: 82%
Last day: 85%
Last week: 90%

Performance

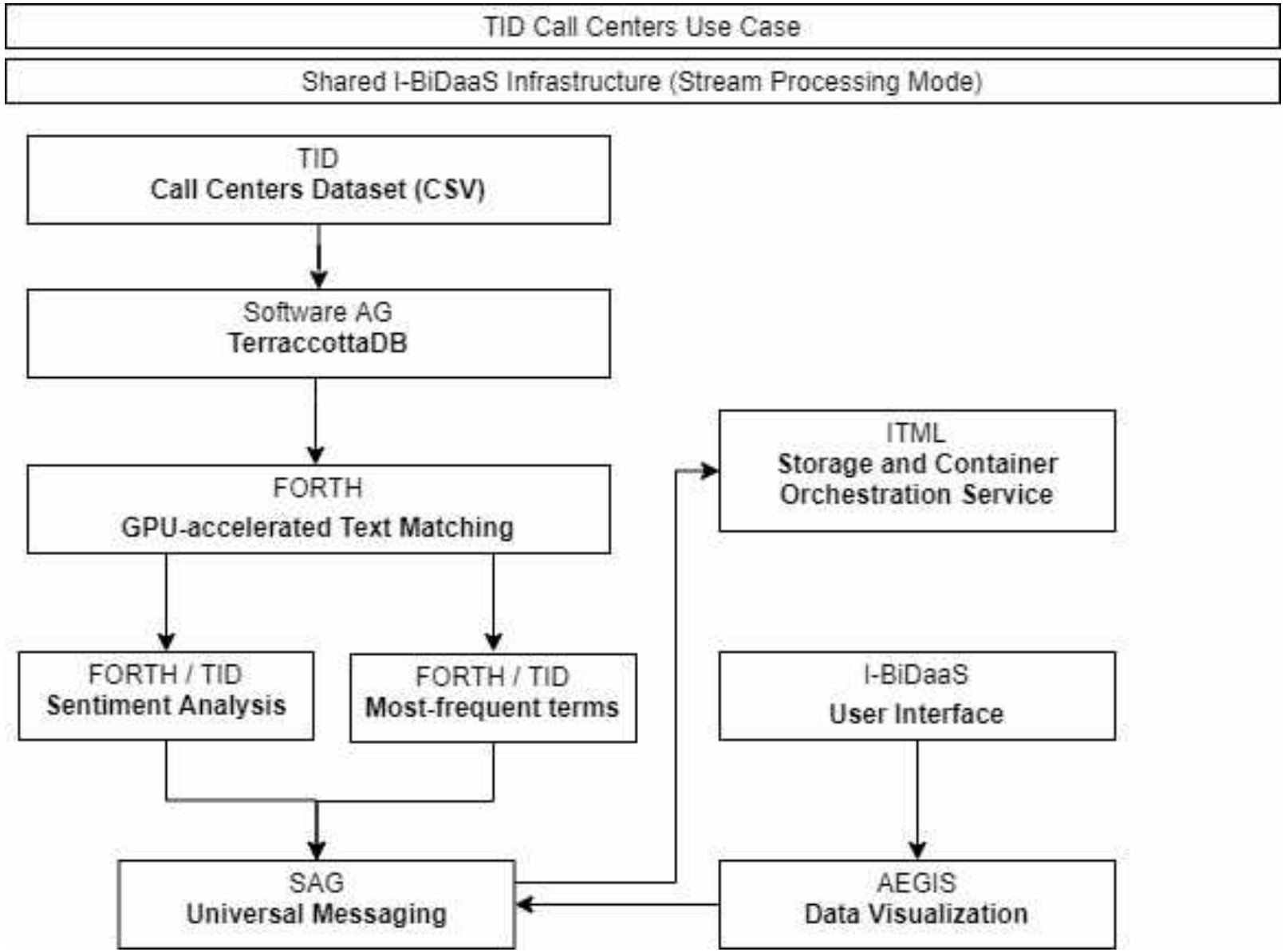
- **High efficiency**



- **Currently**
  - Business Units in TEF manually inspect a small portion of phone calls, less than 1% of total amount of CC calls per year
  - Prioritization of CC operations based on previous insights
- **I-BiDaaS approach**
  - Estimates a sentiment score aggregated by call centers/regions and by time window and a list of more relevant words:
    - Max. real.-time throughput: **42-48K calls/sec**

Benchmark performed with a GPU Nvidia GTX1080

I-BiDaaS solution specialized to call centers



Big Data Pilot Demo Days

I-BiDaaS Application to Telecommunications Sector

Arapakis, Ioannis (Telefonica Research)

Quality of Service in Call Centers

Improve performance of audio calls processing by automatically predicting customer satisfaction

Synthesized Data: ✓
Real Data: ✓

Accurate location prediction with high traffic and visibility

Enable the automatic extraction of behavioural patterns of customers

Synthesized Data: ✓
Real Data: ✓

Optimization of placement of telecommunication equipment

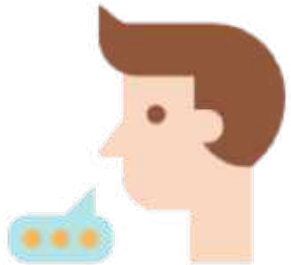
Improve routing and placement of the telecommunication equipment

Synthesized Data: ✓
Real Data: ✓

TID Use Cases:
Quality of Service in Call Centres (CC)

QoS in Call Centres Requirements

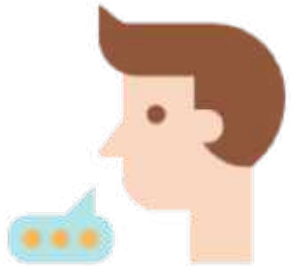
- Indicative examples and questions:
  - Users calling to solve a technical question or to get commercial feedback
  - What is the call topic? How to automate the response?
- Requirements:
  - Customer Satisfaction Index (CSI) is a key parameter for most of the organizations which are offering commercial services
  - The CSI is essentinal to both reducing churn and increasing revenues
  - Quickly get familiar and understand customer's perspective and sentiment
  - Shorten number of calls, duration and waiting time



Telefonica

QoS in Call Centres Overview

- Detect customer satisfaction and dissatisfaction and motivation based on the automatic transcripts of the calls
- Important **challenges** that stem from this use case
  - Employ advanced machine learning techniques (automatic speech recognition, sentiment analysis, ad hoc CSI prediction, acoustic feature extraction) for automatic customer call intent understanding and sentiment extraction
  - Application of streaming-based model for real-time script adaptation
  - Anonymisation of transcripts



I-BiDaaS

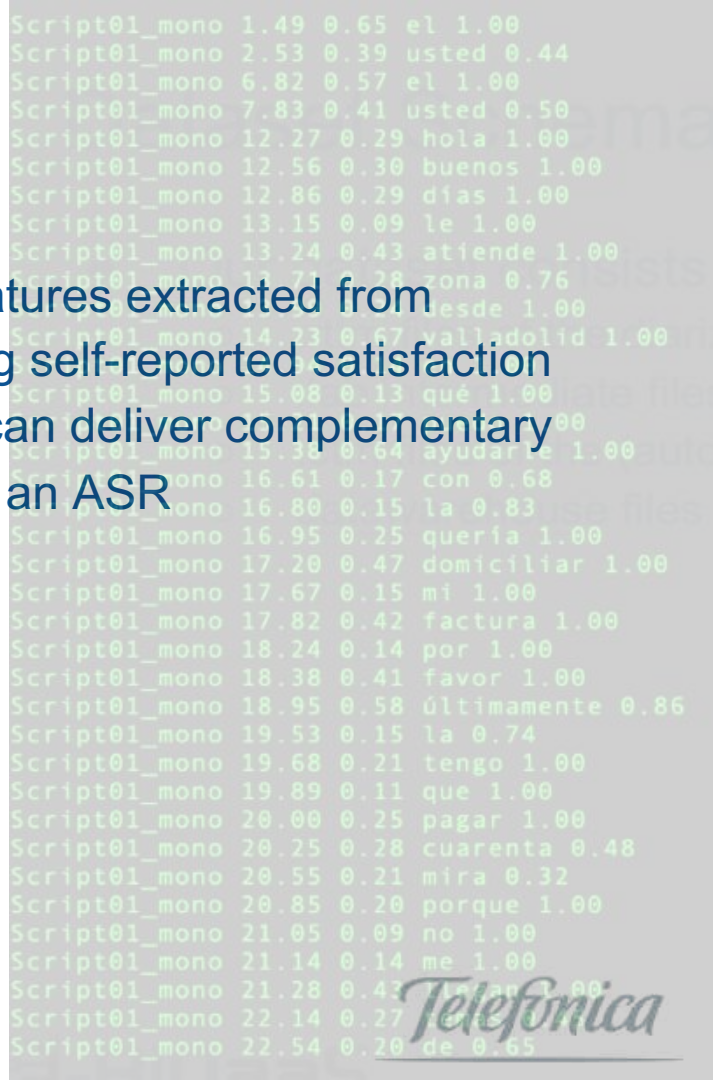
Telefonica

QoS in Call Centres Data

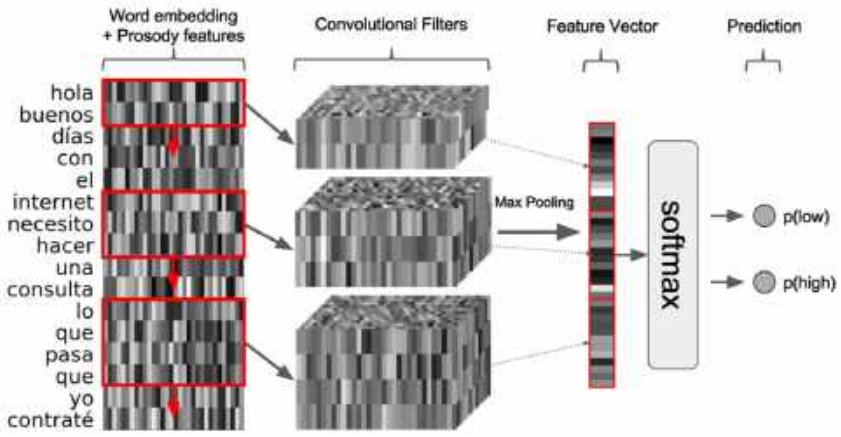
- The data set consists of a **mixture** of **heterogenous**, structured and unstructured data sources
- 20 hours of speech (**manually transcribed** for each language), where speech data is anonymized:
  - Based in automatic algorithm we switch by a tag [·]
  - Switch more common proper names by tag [name]
  - Switch more common products and brand names by tag [name]
  - Switch more popular population names, addresses, … by tag [population]
  - Switch any number by tag [numbers]
- Based in manual supervision of transcripts:
  - Editors vetting the text and validation the automatic anonymization

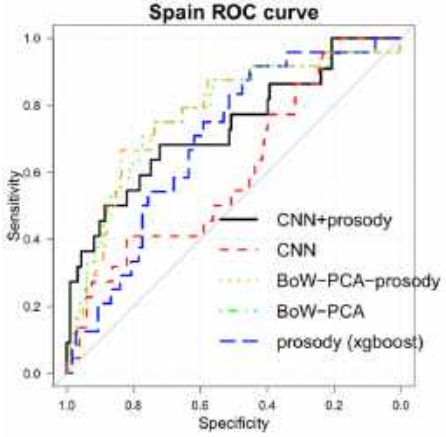
QoS in Call Centres Data

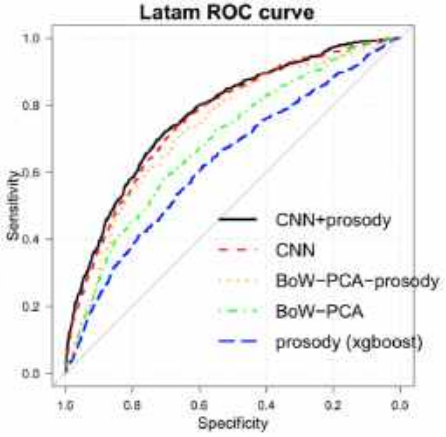
- Analyzed the significance of various acoustic features extracted from customer-agents' spoken interaction in predicting self-reported satisfaction
- Investigated whether speech prosodic features can deliver complementary information to speech transcriptions provided by an ASR
- Our dataset consists of the following:
  - .rttm files of the diarization output
  - .lab intermediate files the ASR generates
  - .ctm files of the (automatic) transcription output
  - data warehouse files (as produced by sql queries)



QoS in Call Centres Research







	Dataset AUC, (F-score)	
	Spain	Latam
BoW-PCA	0.716 (0.324)	0.689 (0.262)
XGBoost (prosody)	0.58 (0.185)	0.610 (0.2398)
BoW-PCA + prosody (MW)	0.7309 (0.3420)	0.701 (0.269)
CNN	0.605 (0.212)	0.759 (0.410)
CNN + prosody	**0.733 (0.242)**	**0.772 (0.427)**

KPIs of QoS in Call Centres

- No automated solution in place; customer audio calls are processed by human agents (**costly** and **time consuming** solution)
- Given an average call duration of 8.6', a human agent following a work schedule of 40 hours per week (160 hours per month), could help process up to **11,520 calls** per year
- This manual process allows to flag **~2,300** low customer satisfaction calls that require (further & immediate) attention and may lead to customer churn

KPIs of QoS in Call Centres

- The I-BiDaaS platform (configuration with 1 core) will process 12 x 8.6' calls per hour; this equals to **105,120** calls per year
- Increase the detection of low customer satisfaction audio calls by human agents from **2,300** to **7,000** (~200% improvement*)
- A solution for processing the customer audio calls in an **automatic** and **scalable manner** can go a long way in increasing the number of low customer satisfaction calls detected

***Assuming a 70% recall**

QoS in Call Centres Demo

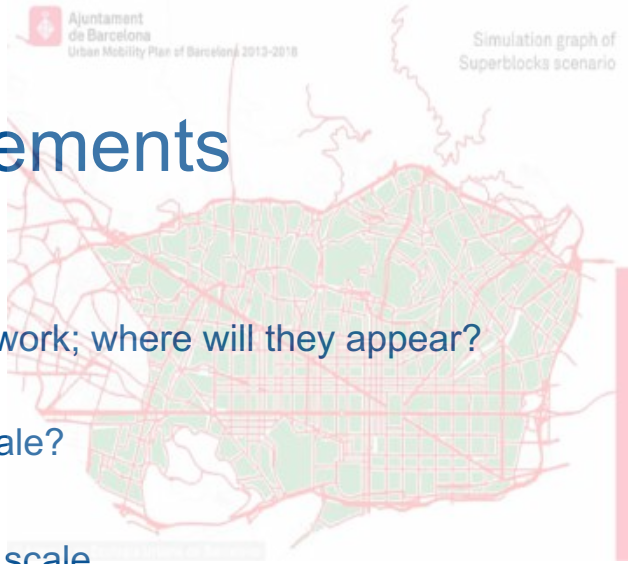
TID Use Cases:

Accurate location prediction with high traffic and visibility

&

Optimization of placement of telecommunication equipment

Accurate Location Prediction Requirements



- Indicative example and questions:
  - Users travel around the city creating traffic congestions in network; where will they appear? When?
  - Can we predict when new events will cause movements at scale?
- Requirements:
  - Forecast immediately next events to anticipate movements at scale
  - Improve the routing and placement of the telecommunication equipment that is already in place, or to arrange accordingly the new equipment obtained

Accurate Location Prediction Overview

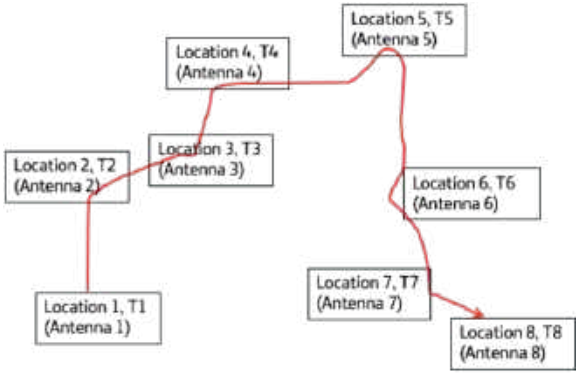
- Predict places with high traffic and congestion events in order to optimise their resource distribution
- Important challenges that stem from this use case:
  - Interpolate missing events to recover plausible event trajectories
  - Minimize processing time with respect to growing data size
  - Maintain real-time delivery of results

Accurate Location Prediction Data

- Our dataset consists of **anonymous traces** collected from a large European cellular network provider with tens of millions of subscribers (cross-sectorial)
- Each trace is a time series of mobile events
- These events contain the encrypted user identifier, a timestamp, and the location of the associated base station used to deliver service to the user
- Base stations have varied coverage (between ~100 m to tens of km) depending on deployment density and radio propagation characteristics like obstacles, hills, or mountains

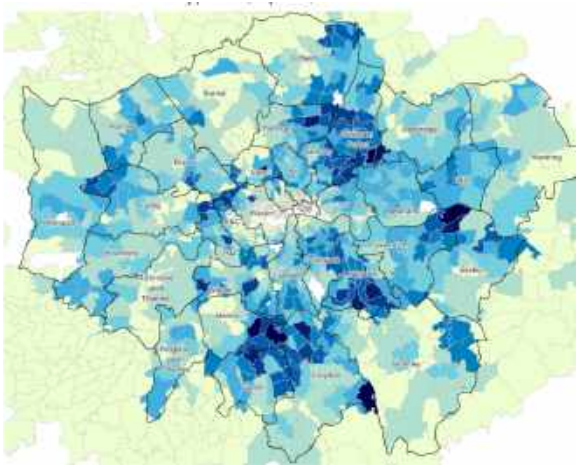
Accurate Location Prediction Data

- A mobile event is generated every time a mobile device:
  - Activates/deactivates in the network (i.e., the user switches on and off their phone)
  - Makes/receives a call or sends/receives an SMS
  - Moves from one location area code to another (i.e., handovers)
  - Changes from one technology to another (i.e., between 2G, 3G, and 4G)
  - Requests access to data (2G/3G) or requests a high-speed data channel (4G/5G)
  - Is actively pinged by the network if no other event is registered for 2 hours (i.e., a timeout to check if the device is still alive)



Accurate Location Prediction Solution

- We will predict movement from one sector to another
- Estimate the delta (%) in connected users per sector x hours in advance
- The forecasting model may incorporate:
  - Information about the other sectors' current status and load
  - Contextual data, such as the weather for the location
  - Occurring events (holiday events or other)



Optimization of Equipment Placement Reqr.

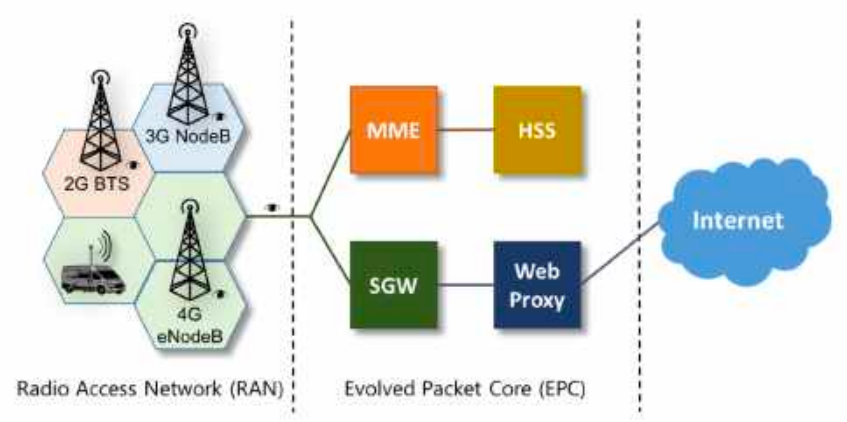
- Indicative examples and questions:
  - To manage large-scale cellular networks, operators need intel on which sectors underperform at any given time
  - Can we avoid deploying new antennas and use existing infra?
  - Can we predict with accuracy which antennas will become the next “hot spots”?
  - What models can be applied?
- Requirements:
  - Minimize processing time with respect to growing data size
  - Maintain real-time delivery of results

Optimization of Equipment Placement Overview

- Predict cell sites with high traffic and congestion events in order to optimise their resource distribution
- Important tasks that stem from these challenges:
  - Analyse streaming data in order to improve the routing and placement of the telecommunication equipment that is avialable or arrange for new equipment to be obtained
  - Study the spatio-temporal patterns and provide insights on the dynamics of cellular sectors
  - Consider DL models and study their performance as a function of time, amount of historical data, and prediction horizon

Optimization of Equipment Placement Data

- Optimization of Placement of Telecommunication Equipment 2G/3G/4G/5G feeds are separated by radio technology Mobile Network Operators (MNOs)
- These feeds continuously provide various Key Performance Indicators (KPIs), such as coverage monitoring, and voice/data service metrics, about each radio sector

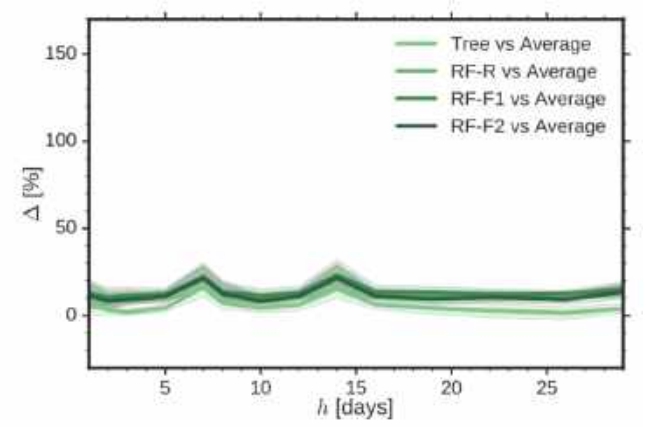


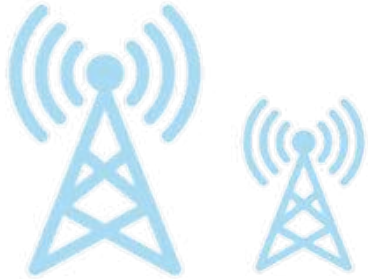
Optimization of Equipment Placement Data

- Such KPIs are one of the key information for MNOs to understand network performance
- The employed indicators correspond to 2G/3G/4G/5G sectors and can be grouped into the following categories:
  - **coverage** (e.g., radio interference, noise level, power characteristics)
  - **accessibility** (e.g., success establishing a voice or data channel, paging success, allocation of high-speed data channels)
  - **retainability** (e.g., fraction of abnormally dropped channels)
  - **mobility** (e.g., handovers' success ratio)
  - **availability and congestion** (e.g., number of transmission time intervals, number of queued users waiting for a resource, congestion ratios, free channels available)

Optimization of Equipment Placement Solution

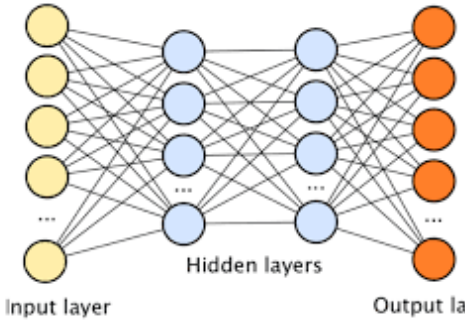
- We want to exploit the KPI information to predict if a sector is going to be a **hotspot** or not
- As the main objective for predictions is intervention, and such interventions cannot be typically made ‘on-the-hour’ by operators, so we aim for a daily resolution





Optimization of Equipment Placement Solution

- We address this problema as a binary classification task; hence, our target variable is a binary label that corresponds to the notion of “being a hot spot” at a certain day
- Alternatively, we explore a multi-class prediction scenario (under-utilised, normal, over-utilised) based on a predefined % threshold of usage, e.g.,
  - Over-utilised: > 5% of average usage
  - Normal: -/+ 5% of average usage
  - Under-utilised: < 5% of average usage
- Use DL models (RNNs, CNNs)



Demo

Accurate location prediction with high traffic and visibility

&

Optimization of placement of telecommunication equipment