Horizon 2020 Grant Agreement No 809965
Contract start date: July 1st 2018, Duration: 30 months

LAMBDA Deliverable 4.3
LAMBDA Learning and Consulting Tools at PUPIN

Due date of deliverable: 30/06/2020
Actual submission date: 30/06/2020

Revision: Version 1.0

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme, H2020-WIDESPREAD-2016-2017 Spreading Excellence and Widening Participation under grant agreement No 809965.

Author(s)	Marko Jelić, Dea Pujić, Dušan Popadić, Dejan Paunović (PUPIN)
Contributor(s)	Hajira Jabeen, Damien Graux
Internal Reviewer(s)	Sahar Vahdati (UOXF)
Approval Date	
Remarks	

Workpackage	WP 4 Experts Exchange Program: Meeting the Big Data Challenges in practice
Responsible for WP	Institute Mihajlo Pupin
Deliverable Lead	Institute Mihajlo Pupin (Valentina Janev)
Related Tasks	Task 4.3 LAMBDA Learning and Consulting Tools at PUPIN

Document History and Contributions

Version	Date	Author(s)	Description
0.1	20.02.2020	Marko Jelić, Dea Pujić	First Draft
0.2	30.03.2020	Hajira Jabeen, Damien Graux	SANSA Tutorial and testing environment
0.3	19.06.2020	Dušan Popadić, Dejan Paunović	Update
0.4	22.06.2020	Marko Jelić, Dea Pujić	Update
0.5	30.06.2020	Sahar Vahdati	Internal review

Executive Summary

This deliverable summarizes the activities of the Mihajlo Pupin Institute's activities regarding the planned activities in Task 4.3 (LAMBDA Learning and Consulting Tools at PUPIN) framework. Task 4.3 (M13-M24) has three main objectives:

- to establish a single environment (BDA Learning and Consulting platform) for learning Big Data related algorithms, methods, tools and prototypes with the help of visiting scholars from the linked institutions. The objective is to establish a playground for early stage researchers for experimentation with open source tools in Big Data scenarios relevant for PUPIN.
- to provide an opportunity for UBO, IAIS and UOXF researchers to learn about real-world challenges from existing 'Big Data' PUPIN clients from government, energy, transport and other sectors.
- to continuously monitor the Big Data & Analytics market.

In the last two year, the PUPIN team succeed to establish a Technology watch activity where the researchers constantly explore the market of Big Data tools and conduct experiments. As a result, several articles were presented and published as conference or journal papers.

The most promising domain for experimentation is the energy sector, based on the availability of data from the PUPIN proprietary VIEW4 SCADA system.

Table of Contents

Abbreviations and Acronyms

API	Application Programming Interface
BDA	Big Data Analytics
HTML	Hypertext Markup Language
RDD	Resilient Distributed Datasets
URI	Unified Recourse Identifier
WP	Work Package

List of Figures

List of Tables

1. Introduction

1.1 Scope

This report discusses the activities in Task 4.3 framework, in Phase 2 of the project, see Figure 1.

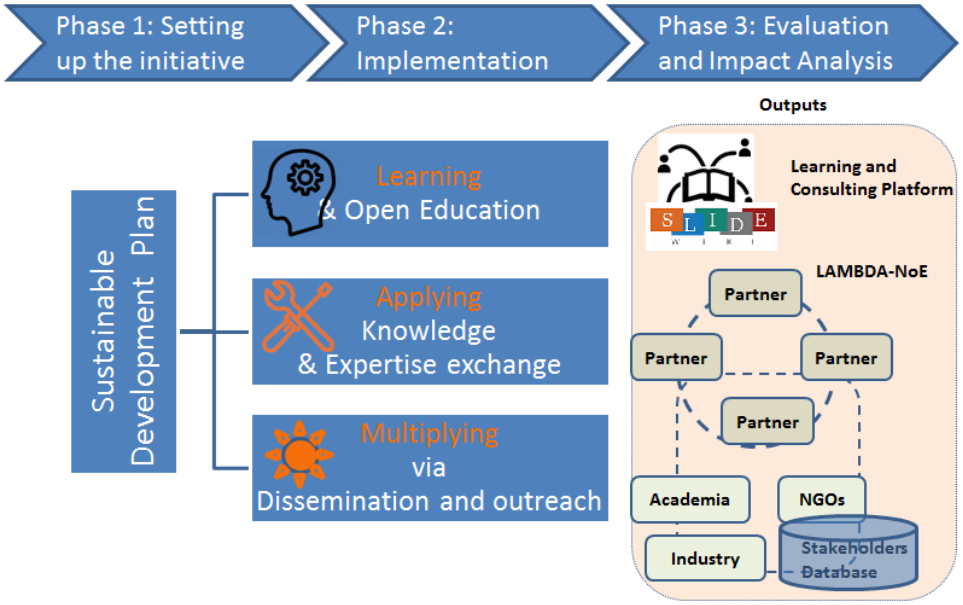

Figure 1. LAMBDA Methodology

The main objective of Work Package 4 (Experts Exchange Program: Meeting the Big Data Challenges in practice) of the LAMBDA (**L**earning, **A**pplying, **M**ultiplying **B**ig **D**ata Analytics, http://www.project-lambda.org/) project is knowledge transfer and expertise exchange on:

- Facilitating fundamentals of Knowledge Graphs and Big Data Analytics;
- Applying Linked Data principles for smart integration and analytics applications in multiple research areas such as e-Government, e-Environment, e-Health, Energy Efficiency, Safety and Security, Smart Cities, and Traffic Management;
- Providing methods and techniques to improve the quality, performance, and security of research results (software tools, pilot applications), and market opportunities (discussion on business solutions relevant for companies).

Task 4.3 LAMBDA Learning and Consulting Tools at PUPIN (M13-M24) has three main objectives:

- The first objective of Task 4.3 is to establish a single environment (BDA Learning and Consulting platform) for learning Big Data related algorithms, methods, tools and prototypes with the help of visiting scholars from the linked institutions. The objective is to establish a playground for early stage researchers for experimentation with open source tools in Big Data scenarios relevant for PUPIN.
- The second objective is to provide an opportunity for UBO, IAIS and UOXF researchers to learn about real-world challenges from existing 'Big Data' PUPIN clients from government, energy, transport and other sectors.

- The third objective of this Task is to continuously monitor the Big Data & Analytics market (see for instance the Big Data Analytics Landscape on Figure 1, source https://mattturck.com/data2019) and benchmarking activities.

1.2 Relation to other Deliverables

This deliverable is related to:

- Deliverable 2.1 Big Data Challenges and Analysis of Scientific and Technological Landscape that gives an overview of the Big Data concepts, outlines some of the relevant challenges in this domain and reviews and describes the current state of the art tools relevant to Big Data applications.

- Deliverable 2.2 Education and RTD Needs that presents the Research and Development activities of the LAMBDA consortium and introduces details about PUPIN's R&D Priorities.

- Deliverable 3.1 The 'Trainers' Network' Infrastructure that describes the adoptions made on the LAMBDA platform (see https://project-lambda.org/) in order to facilitate teachers-trainees cooperation.

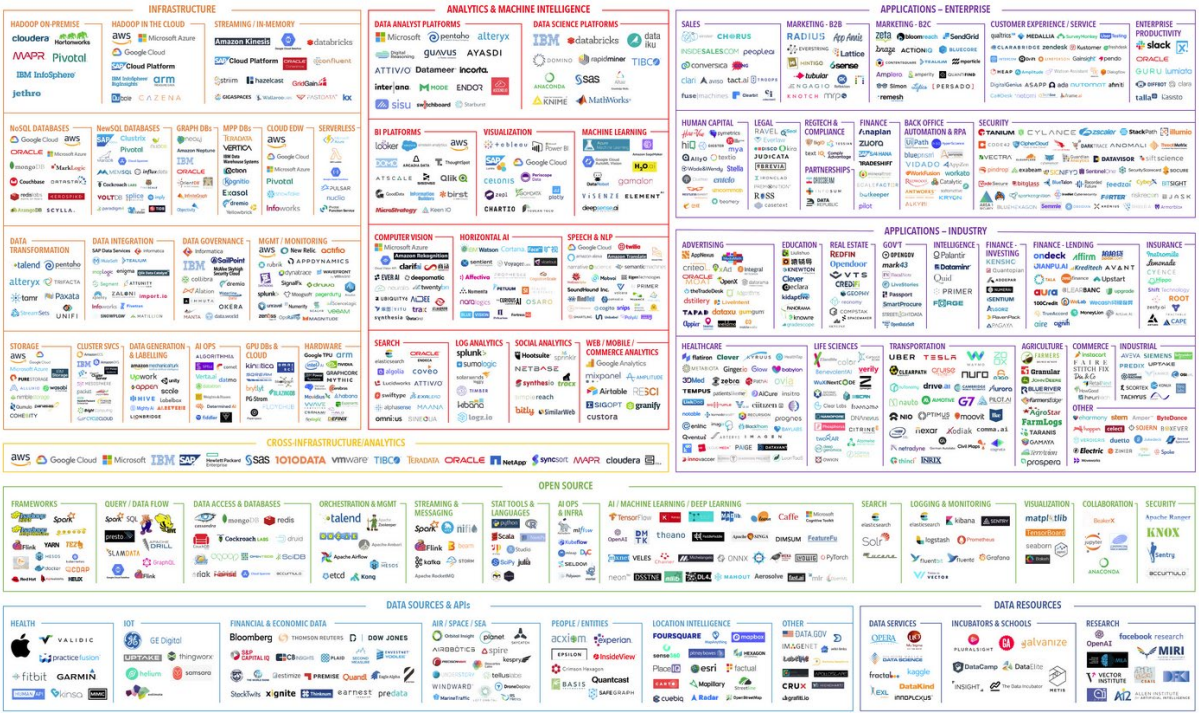

Figure 2. Data & AI Landscape (source: Matt Turck, June 27, 2019)

1.3 Structure of the Deliverable

Section 2 presents the High Level Vision of the LAMBDA Learning and Consulting Platform, the existing learning possibilities (Lectures, Big Data Tool) and the configuration changes implemented in order to register and search the learning items.

2. Big Data Tools

For an easier and more effective collaboration among consortium members (e.g facilitating joined paper and deliverable writing, version management, information sharing, stakeholders data-base management, etc) and with stakeholders, the LAMBDA platform was established in month 1 of the project. The public part of the platform, see https://project-lambda.org/, is relevant for end users interested to exploit the learning capabilities e.g. to retrieve a Lecture (see Figure 2), to link to SlideWiki platform where additional materials can be found, to learn about the Big data tools identified from the consortium as relevant for students and professionals. The platform also serves to present the LAMBDA communication and dissemination activities (see Deliverable 5.2 Dissemination and Communication Strategy and Preliminary Exploitation Plan , Deliverable 5.3 First Report on Stakeholder Engagement and Exploitation Activities, Deliverable 5.5 First Report on Communication activities and Dissemination Events 1.0).

The Lecture repository is accessible via the link https://project-lambda.org/Knowledge-repository/Lectures. More information about the Lectures can be found in Deliverable 3.4 Smart data Analytics.

The Tools repository is accessible via the link https://project-lambda.org/tools-for-experimentation.

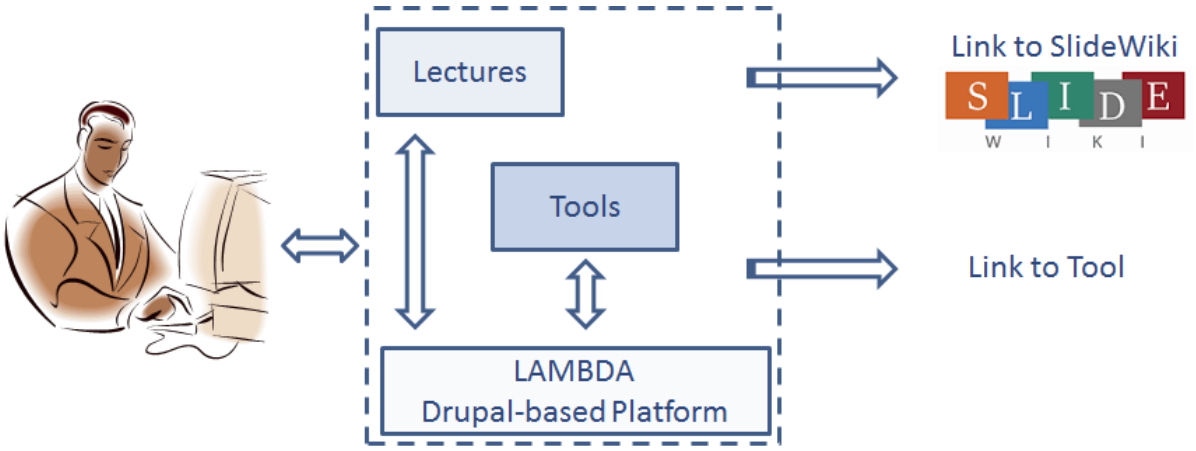

Figure 3. High Level Vision of the LAMBDA Learning and Consulting Platform

2.1 Categorization of Tools

The focus in the first report related to the analysis of Big Data landscape named D2.1 Big Data Challenges and Analysis of Scientific and Technological Landscape (M06) was primarily on the identification, selection and observation of information sources closely related to the Big Data Analytics field. By analysing the most popular frameworks used to handle Big Data, in Section 3, we proposed a characterization of the landscapes in the topics related to Big Data into the following categories

- Big Data Frameworks
- NoSQL Platforms
- Stream Processing Data Engines
- Big Data Preprocessing
- Big Data Analytics
- Big Data Visualization Tools

During the last 18 months, the PUPIN team conducted extensive analysis of functionalities of Big Data frameworks and engines and selected tools for experimentation, see Section 3. Based on our observation, we propose herein the following categorization of the Big Data landscape:

- Cloud Marketplaces
- Hadoop as a Web Service / Platform
- Operational Database Management Systems
- NoSQL/ Graph databases
- Stream Processing Engines
- Analytics Software / System / Platform
- Data Analytics Languages
- Optimization Library for Big Data
- Library / API for Big Data
- ML Library / API for Big Data
- Visualization Software / System
- Distributed Messaging System

2.2 Registering Tools with the LAMBDA Platform

An authorized user can register a new tool in the LAMBDA repository by entering the relevant data about the tool, as is presented in Figure

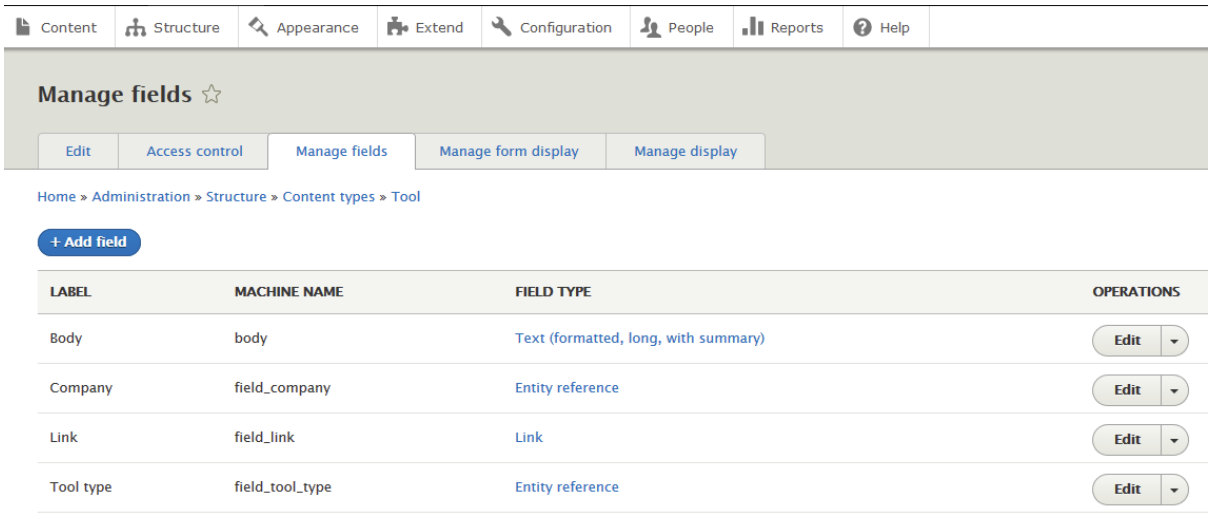

Figure 4. Definition of a Content Type - TOOL

2.3 Overview of Tools

By June 2020, more than 80 tools have been identified that exist in the market, but just several of them were tested with data and in scenarios relevant for the Institute Mihajlo Pupin. Table 1 gives an overview of the tools that have been identified.

Table 1. Identified Tools

Cloud Marketplaces	Alibaba Cloud IBM Cloud Google Cloud Platform Oracle Cloud Marketplace CISCO Marketplace Microsoft Azure Marketplace AWS

	Marketplace
Hadoop as a Web Service / Platform	HDInsight IBM InfoSphere BigInsights MapR Cloudera CDH Amazon EMR
Operational Database Management Systems	IBM (DB2) SAP (SAP HANA) Microsoft (SQL Server) ORACLE (Database)
NoSQL/ Graph databases	Hadoop Distributed File System (HDFS) Amazon Neptune Neo4j TigerGraph MapR Database Ontotext GraphDB AlegroGraph Virtuoso Appache Jena MarkLogic JanusGraph OrientDB Microsoft Azzure Cosmos DB Apache Hbase Apache Cassandra MongoDB
Stream Processing Engines	Apache Flume Apache Apex Amazon Kinesis Streams Apache Flink Apache Samza Apache Storm Apache Spark
Analytics Software / System / Platform	SAS Analytics Software & Solutions MATLAB H2O.ai Accord Framework Apache Hadoop Cloudera Data Platform Vadalog System MATLAB Semantic Analytics Stack (SANSA)
Data Analytics Languages	Scala Julia SPARQL SQL R Python Package Index (PyPI) Python
Optimization Library for Big Data	Facebook Ax HyperOpt IBM ILOG CPLEX Optimization Library
Library / API for Big Data	TensorFlow Serving MLlib BigML Google Prediction API Azure Machine Learning Amazon Machine Learning API IBM Watson Programming with Big Data in R
ML Library / API for Big Data	CAFFE.AI Appache MXNet XGBoost PyTorch Keras TensorFlow
Visualization Software / System	Oracle Visual Analyzer Microsoft Power BI Datawrapper Qlikview Canvas.js Highcharts Fusion chart D3 Tableau Google Chart
Distributed Messaging System	Apache Kafka

3. Experiments

This Section describes in more details the experiments carried out with Big Data tools.

3.1 Python 3 programming language

Python is a general-purpose, versatile and popular **programming language**. It's great as a first **language** because it is concise and easy to read, and it is also a good **language** to have in any **programmer's** stack as it can be used for everything from web development to software development and data science applications. More info at https://www.python.org/

Objective	Development and deployment of several interoperable cloud (server-based) services for multiple projects
Scenario	Providing a basis for effortless integration of different data processing, machine learning and optimization frameworks
Category of the tool	Data Analytics Languages
Datasets used	/
Where is the tool installed	Server (for deployed services) and local machine (for research and development)
Dissemination of results	This tool was used as a basis for multiple research efforts, however it is not the focus of any specific undertaking.

3.2 MATLAB

MATLAB is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages. More info at https://www.mathworks.com/products/matlab.html

Objective	Comparison of state of the art machine learning technique (MATLAB 8.2, R2013b release)
Scenario	Student exam performance prediction
Category of the tool	Analytics Software / System / Platform
Datasets used	Datasets created in the SlideWiki project framework, see also SlideWiki.org platform.
Where is the tool installed	Local machine (for research and development)
Dissemination of results	Nikola Tomasević, Nikola Gvozdenović, Sanja Vraneš, An overview and comparison of supervised data mining techniques for student exam performance prediction, Computers and Education Volume 143, January

	2020, 103676 https://doi.org/10.1016/j.compedu.2019.103676

Objective	Prototyping optimization solutions and data visualization
Scenario	Rapid testing of different models and methodologies
Category of the tool	Analytics Software / System / Platform
Datasets used	/
Where is the tool installed	Local machine (for research and development)
Dissemination of results	This tool was used as a basis for multiple research efforts, however it is not the focus of any specific undertaking. Related projects • **REACT** – **R**enewable **E**nergy for self-sust**A**inable island Communi**T**ies • RESPOND: Integrated demand REsponse Solution towards energy POsitive NeighbourhooDs

3.3 IBM ILOG CPLEX Optimization Library (for MATLAB, Python and Java)

The IBM ILOG CPLEX Optimizer solves integer programming problems, very large linear programming problems using either primal or dual variants of the simplex method or the barrier interior point method, convex and non-convex quadratic programming problems, and convex quadratically constrained problems (solved via second-order cone programming, or SOCP).
The CPLEX Optimizer has a modelling layer called Concert that provides interfaces to the C++, C#, and Java languages. There is a Python language interface based on the C interface. Additionally, connectors to Microsoft Excel and MATLAB are provided. Finally, a stand-alone Interactive Optimizer executable is provided for debugging and other purposes.

The CPLEX Optimizer is accessible through independent modelling systems such as AIMMS, AMPL, GAMS, OptimJ and TOMLAB. In addition to that AMPL provides an interface to the CPLEX CP Optimizer.

The full IBM ILOG CPLEX Optimization Studio consists of the CPLEX Optimizer for mathematical programming, the CP Optimizer for constraint programming, the Optimization Programming Language (OPL), and a tightly integrated IDE.

More info at https://www.ibm.com/analytics/cplex-optimizer

Objective	Determining the optimal energy management and dispatching strategy for projects https://www.inbetween-project.eu/ and http://project-respond.eu/
Scenario	Calculating the optimal set of variables given a predefined criterion

	function
Category of the tool	Optimization Library for Big Data
Datasets used	Proprietary (use-case specific demand measurements), pricing data from https://www.omie.es/en/market-results/daily/daily-market/daily-hourly-price and meteorological data from https://e3p.jrc.ec.europa.eu/articles/typical-meteorological-year-tmy
Where is the tool installed	Server (for deployed services) and local machine (for research and development)
Dissemination of results	Marko Jelić et all (2020) Towards self-sustainable island grids through optimal utilization of renewable energy potential and community engagement (accepted for publication in Energies journal by MDPI, https://www.mdpi.com/journal/energies).

3.4 VADALOG System

Objective	Testing the features of VADALOG for semantic data processing
Scenario	Experimental integration and reasoning using rules for an energy-based ontology
Category of the tool	Analytics Software / System / Platform
Datasets used	/
Where is the tool installed	Server (accesses as a cloud service)
Dissemination of results	VADALOG was extensively used during the staff exchange in Oxford, UK in February 2020, see https://project-lambda.org/Staff-exchange-UOXF-Feb-2020 . A joint publication that would utilize VADALOG in conjunction with knowledge graph embeddings between University of Oxford and Institute Mihajlo Pupin was planned for the Workshop of Knowledge Representation & Representation Learning at ECAI 2020, however due to the ongoing pandemic, this plan could not be realised

3.5 Apache Spark

Apache Spark is a generic, in-memory data processing engine. It provides high-level APIs in Java, Python and Scala. Apache Spark has simplified the programming complexity by introducing the abstraction of Resilient Distributed Datasets (RDD), i.e. a logical collection of data partitioned across machines. The rich API for RDDs manipulation follows the models for processing local collections of data, making it easier to develop complex programs. Spark provides higher-level constructs and libraries to further facilitate users in writing distributed applications. At the time of writing, Apache Spark provides four libraries:

- Spark SQL - Offers support for SQL querying of data stored in RDDs, or an external data source. It allows structured data processing using high-level collections named dataset and

data frame. A Dataset is a distributed collection of data and a DataFrame is a Dataset organized into named columns. It is conceptually similar to a table in a relational database. The DataFrames can be constructed in numerous different ways like reading from structured data files,tables in Hive, external databases, or existing RDDs.

- Spark streaming - Spark implements stream processing by ingesting data in minibatches. Spark streaming makes it easy to build scalable fault-tolerant real-time applications. The data can be ingested from a variety of streaming sources like Kafka, Flume (covered in earlier sections). This data can be processed using complex real-time algorithms using a high-level API.
- MLlib Machine Learning Library - Provides scalable machine learning algorithms. It provides common algorithms for classification, regression, clustering, algorithms for feature extraction, feature selection and dimensionality reduction, high-level API for machine learning pipelines, saving and loading algorithms, and utilities for linear algebra and statistics.
- GraphX - Provides a distributed graph processing using graph-parallel computation. GraphX extends the Spark RDD by introducing \Graph": a directed multigraph with properties attached to each vertex and edge. GraphX comes with a variety of graph operators like subgraph, joinVertices, or algorithms like pageRank, ConnectedComponents, and several graph builders that allow building a graph from a collection of vertices and edges from RDD or other data sources.

Objective	Running tests that depict stream processing
Scenario	Parsing RDF data
Category of the tool	Stream Processing Engines
Datasets used	Proprietary RDF provided by UBO
Where is the tool installed	Local machines
Dissemination of results	Apache Spark was extensively tested during the staff exchange in Bonn, Germany in February 2019, see https://project-lambda.org/Staff-exchange-IAIS-UBO-Feb-2019

3.6 TensorFlow

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. More info at https://www.tensorflow.org/

Objective	This library has been used for neural network training as a part of H2020 inBETWEEN: ICT enabled BEhavioral change ToWards Energy EfficieNt lifestyles (GA. 768776).
Scenario	Models that have been trained were used for disaggregation of total household energy consumption, i. e. for Non-Intrusive Load Monitoring (NILM).

Category of the tool	ML Library / API for Big Data
Datasets used	Both open (REDD[1] and UKDALE[2]) and closed datasets were used
Where is the tool installed	Local computer
Dissemination of results	D. Pujic, N. Tomasevic and M. Batic, Semi-supervised Approach for Improving Generalization in Non-Intrusive Load Monitoring submitted in Neural Computing and Applications (submitted on April, 16th)

3.7 Keras

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. More info at https://keras.io/

Objective	Keras library has been used as an API for neural network training as a part of H2020 inBETWEEN: ICT enabled BEhavioral change ToWards Energy EfficieNt lifestyles (GA. 768776).
Category of the tool	ML Library / API for Big Data
Datasets used	Both open (REDD[3] and UKDALE[4]) and closed datasets were used
Where is the tool installed	Local computer
Dissemination of results	D. Pujic, N. Tomasevic and M. Batic, Semi-supervised Approach for Improving Generalization in Non-Intrusive Load Monitoring submitted in Neural Computing and Applications (submitted on April, 16th)

[1] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.
[2] Jack Kelly and William Knottenbelt. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific Data 2, Article number:150007, 2015, DOI:10.1038/sdata.2015.7
[3] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.
[4] Jack Kelly and William Knottenbelt. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific Data 2, Article number:150007, 2015, DOI:10.1038/sdata.2015.7

3.8 KAFKA

Apache Kafka is a distributed messaging system that uses the publish-subscribe mechanism. It was developed to support continuous and resilient messaging with high throughput at LinkedIn. Kafka is a fast, scalable, durable, and fault-tolerant system. It maintains feeds of messages in categories called topics. These topics are used to store messages from the producers and deliver them to the consumers who have subscribed to that topic.

Kafka is a durable, high volume message broker that enables applications to process, persist and re-process streaming data. Kafka has a straightforward routing approach that uses a routing key to send messages to a topic. Kafka offers much higher performance than message brokers like RabbitMQ. Its boosted performance makes it suitable to achieve high throughput (millions of messages per second) with limited resources.

Objective	Kafka messaging system will be used for building fault-tolerant real-time data pipelines as a part of H2020 project Trinity (http://trinityh2020.eu/)
Scenario	Exchange of information (forecasts of energy production, real time values of energy production) between RES Control Centre and Trinity Coordination Platform
Category of the tool	Distributed Messaging System
Datasets used	/
Where is the tool installed	PUPIN Server and local machine (for research and development)
Dissemination of results	This tool was used as a basis for multiple research efforts, however it is not the focus of any specific undertaking. Related projects: ● **TRINITY – TR**ansmission system enhancement of reg**I**o**N**al borders by means of **I**ntelligen**T** market technolog**Y**

3.9 Virtuoso Server

Virtuoso Universal Server is a middleware and database engine hybrid that combines the functionality of a traditional Relational database management system, Object-relational database, virtual database, RDF, XML, free-text, web application server and file server functionality in a single system, see https://virtuoso.openlinksw.com/.

Objective	In the last five years many European countries put forward Government 3.0 as a new paradigm, and as a result, improved efficiency in the provision of public services, increased transparency and interaction with citizens and society as a whole, but also created new businesses across Europe. This study was motivated by the need to find better strategies for delivering the data from both local and national governments to the public in a powerful, machine-readable and future-proof format.

Scenario	Publishing open data in Linked Data format and interlinking with DBpedia.
Category of the tool	NoSQL/ Graph databases
Datasets used	Government data from https://data.gov.rs/sr/
Where is the tool installed	PUPIN Server
Dissemination of results	Valentina Janev (2019). Open data: Challenges and Opportunities for Serbia. In I. Janev (Ed) Serbia: Current Political, Economic and Social Issues and Challenges. Nova Science Publishers, ISBN: 978-1-53615-060-5 (eBook), pp. 165-184.

3.10 SANSA - Scalable Semantic Analytics Stack

SANSA is a big data engine for **scalable** processing of large-**scale** RDF data. **SANSA** uses Spark and Flink which offer fault-tolerant, highly available and **scalable** approaches to efficiently process massive sized datasets. **SANSA** provides the facilities for **Semantic** data representation, Querying, Inference, and **Analytics**.
SANSA includes **several libraries** for creating applications:

1. **Read / Write RDF / OWL library** for RDF/OWL operations,
2. **Querying library** support a query language on top of distributed RDF/OWL library, as well as querying heterogeneous non-RDF data.
3. **Inference library** implements rule-based reasoning on RDF/OWL data,
4. **ML- Machine Learning core library**

More info at http://sansa-stack.net/

Objective	Testing SANSA with datasets from Serbia in LAMBDA and PLATOON projects framework.
Scenario	More scenarios under development including • Renewable energy forecasting • Load / Demand forecasting
Category of the tool	Distributed Messaging System
Datasets used	Datasets from the PUPIN proprietary SCADA VIEW4.
Where is the tool installed	SANSA was tested at UBO premises during the Staff exchange in February 2019, https://project-lambda.org/Staff-exchange-IAIS-UBO-Feb-2019 Additonally, Interactive Spark Notebooks for running SANSA-Examples were created for the Hands-on session conducted during the Big Data Analytics Summer School 2020. The repository contains a docker-compose.yml for running Hadoop/Spark cluster locally. The cluster also

	includes Hue for navigation and copying file to HDFS. The notebooks run using Apache Zeppelin.
Dissemination of results	Presentation / paper has been submitted to the International Conference on INnovations in Intelligent SysTems and Applications (INISTA), see http://inista.org/call-for-papers.php

3.11 Apache Jena

Apache Jena is an open source Java framework for building semantic web and Linked Data applications. The framework is composed of different APIs interacting together to process RDF data:

- RDF API
  - Ontology API
  - SPARQL API
- Inference API
- Store API

It also contains TDB (high performance RDF store) and Fuseki (SPARQL server) which together provide a robust, transactional persistent storage layer.

More info at https://jena.apache.org/index.html.

Objective	Apache Jena is used to store an ontology containing knowledge about spatial arrangements of the rooms, apartments and buildings at pilot sites which are part of H2020 inBETWEEN: ICT enabled BEhavioral change ToWards Energy EfficieNt lifestyles (GA. 768776).
Scenario	Improving energy efficiency in buildings.
Category of the tool	NoSQL/ Graph databases
Datasets used	Closed dataset from inBETWEEN project.
Where is the tool installed	PUPIN Server
Dissemination of results	Dusan Popadic, Lazar Berbakov, Marko Jelic, Marko Batic,"Ontology Enabled Internet of Things System for Smart Buildings", ICIST 2020, ISBN: TBC, 10th International Conference on Information Society and Technology, Vol. X, pp. XX-XX, 2020

3.12 SPARQL (SPARQL Protocol and RDF Query Language)

SPARQL is a RDF query language and is one of the key technologies of the semantic web.

Objective	SPARQL language is used to query the ontology stored on Jena Fuseki server as a part of H2020 inBETWEEN: ICT enabled BEhavioral change ToWards Energy EfficieNt lifestyles (GA. 768776).
Scenario	Querying the ontology.
Category of the tool	Data Analytics Languages
Datasets used	Closed dataset from inBETWEEN project.
Where is the tool installed	locally
Dissemination of results	Dusan Popadic, Lazar Berbakov, Marko Jelic, Marko Batic,"Ontology Enabled Internet of Things System for Smart Buildings", ICIST 2020, ISBN: TBC, 10th International Conference on Information Society and Technology, Vol. X, pp. XX-XX, 2020

4. Conclusion

This deliverable summarizes the activities of the Mihajlo Pupin Institute's activities in Task 4.3 (LAMBDA Learning and Consulting Tools at PUPIN) framework. In the last two year, the PUPIN team succeed to establish a Technology watch activity where the researchers constantly explore the market of Big Data tools and conduct experiments. As a result, several articles were presented and published as conference or journal papers.

The most promising domain for experimentation is the energy sector, based on the available data from the PUPIN proprietary VIEW4 SCADA system.