PROTEUS

Scalable online machine learning for predictive analytics and real-time interactive visualization

687691

D6.10 PROTEUS factsheet leaflet

Lead Author: Mª José García Pumarino (TREE)
With contributions from: Rubén Casado, Marcos Sacristán (TREE)
Reviewer: Rachel Finn [TRI]

Deliverable nature:	Websites, patents filling, etc.
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	29th February 2016 (Month 3)
Actual delivery date:	9th February 2016
Version:	1.0
Total number of pages:	14
Keywords:	Leaflet, factsheet, public, dissemination

Abstract

This deliverable defines the main characteristics and design of the public leaflet factsheet of the PROTEUS Project.

The main objective of this leaflet is to provide an early dissemination material for communication purposes, including the most relevant information of the project in a nutshell. It is available from the very beginning of the project activities as an initial public brochure, though new versions will be produced in the future if needed and agreed between the project partners.

Executive summary

This document provides a design of the PROTEUS leaflet. The leaflet is presented in a diptych form, with a cover page, a back page and the main contents inside, and includes the following sections:

- Cover page: including the project's logo, name and acronym.
- Main contents:
    - Abstract.
    - Duration.
    - Contributions & impact.
    - A short project description.
- Back page: including the project's logo, a list of the partners, and acknowledgements.

Document Information

IST Project Number	687691	**Acronym**	PROTEUS
Full Title	Scalable online machine learning for predictive analytics and real-time interactive visualization		
Project URL	http://www.proteus-bigdata.com/		
EU Project Officer	Martina EYDNER		

Deliverable	**Number**	D6.10	**Title**	PROTEUS factsheet leaflet
Work Package	**Number**	WP6	**Title**	Communication, exploitation and dissemination

Date of Delivery	**Contractual**	M03	**Actual**	M03
Status	version 1.0		final ■	
Nature	report □ demonstrator □ other ■			
Dissemination level	public ■ restricted □			

Authors (Partner)	TREE			
Responsible Author	**Name**	Marcos Sacristán	**E-mail**	Marcos.sacristan@treelogic.com
	Partner	TREE	**Phone**	+34 663 24 66 99

Abstract (for dissemination)	This deliverable defines the main characteristics and design of the public leaflet factsheet of the PROTEUS Project. The main objective of this leaflet is to provide an early dissemination material for communication purposes, including the most relevant information of the project in a nutshell. It is available from the very beginning of the project activities as an initial public brochure, though new versions will be produced in the future if needed and agreed between the project partners.
Keywords	Leaflet, factsheet, public, dissemination

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
04/01/2016	0.1	TREE	Main structure & initial draft
08/02/2016	0.2	TREE	Version of internal peer review process
09/02/2016	0.3	TRI	Review of final deliverable
09/02/2016	1.0	TREE	Final release

Table of Contents

Executive summary 3
Document Information 4
Table of Contents 5
List of figures and/or list of tables 6
Abbreviations 7
1 Introduction 8
2 Leaflet structure & contents 9
2.1 Cover page 9
2.2 Main contents 10
2.2.1 Abstract 11
2.2.2 Contributions and impact 11
2.2.3 Project description 11
2.3 Back page 13
3 Conclusions 14

List of figures and/or list of tables

Figure 1: PROTEUS leaflet, cover page 9
Figure 2: PROTEUS leaflet, main page 10
Figure 3: PROTEUS overview diagram 12
Figure 4: PROTEUS leaflet, back page 13

Abbreviations

EU: European Union.

1 Introduction

This document provides a design of the PROTEUS leaflet. The leaflet is presented in a diptych form, with a cover page, a back page and the main contents inside.

This document provides detailed information about the design and contents of the leaflet at the time this report has been issued (Month 3 of the project). Although there are no future versions planned for this deliverable, the option for preparing future versions of a leaflet or other dissemination materials of the project is left open.

2 Leaflet structure & contents

The leaflet is presented in a diptych form, printable in a DIN-A4 format, with a cover page, a back page, and the main contents inside. It includes the following sections:

- Cover page: including the project's logo, name and acronym.
- Main contents:
  - Abstract.
  - Duration.
  - Contributions & impact.
  - A short project description.
- Back page: including the project's logo, a list of the partners, and acknowledgements.

The following sections provide information about the design and contents for each section.

2.1 Cover page

The cover page includes the project's logo, the full name and the acronym, over a background image representing the steelmaking industry (the main demonstration scenario).

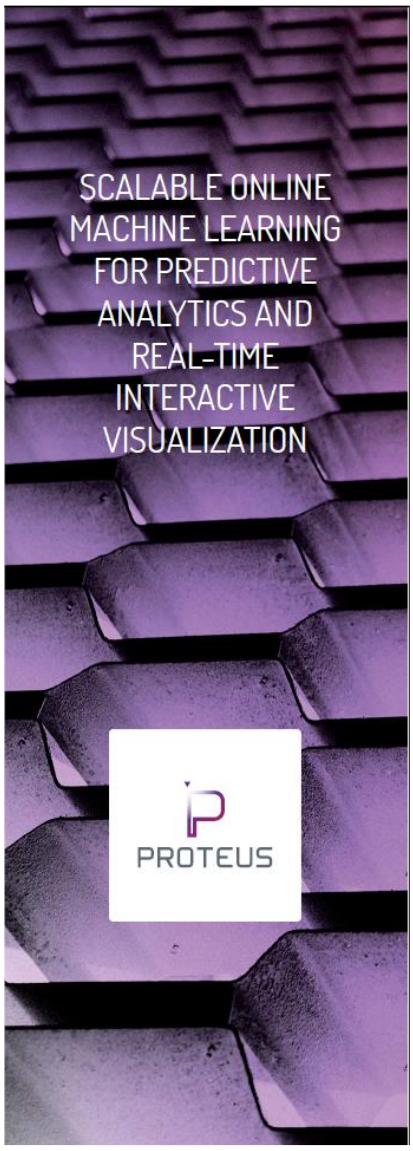


Figure 1: PROTEUS leaflet, cover page

2.2 Main contents

The central part of the leaflet provide the main contents, as shown below:

PROTEUS

ABSTRACT

PROTEUS mission is to investigate and develop ready-to-use **scalable online machine learning** algorithms and real-time interactive visual analytics to deal with extremely large data sets and data streams.

The foundation for the **PROTEUS** advances is the use of an optimized implementation of **combined batch and streaming processing** and building around this later scalable real time processes. The developed algorithms and techniques will form a library to be integrated into an enhanced version of Apache Flink, the EU Big Data platform.

PROTEUS will contribute to the Big Data area by addressing fundamental challenges related to the scalability and responsiveness of analytics capabilities. The requirements are defined by a steelmaking industrial use case. The techniques developed in PROTEUS are however, general, flexible and portable to all data stream-based domains.

DURATION

36 month (Dec. 2015 - Nov. 2018)

CONTRIBUTIONS & IMPACT

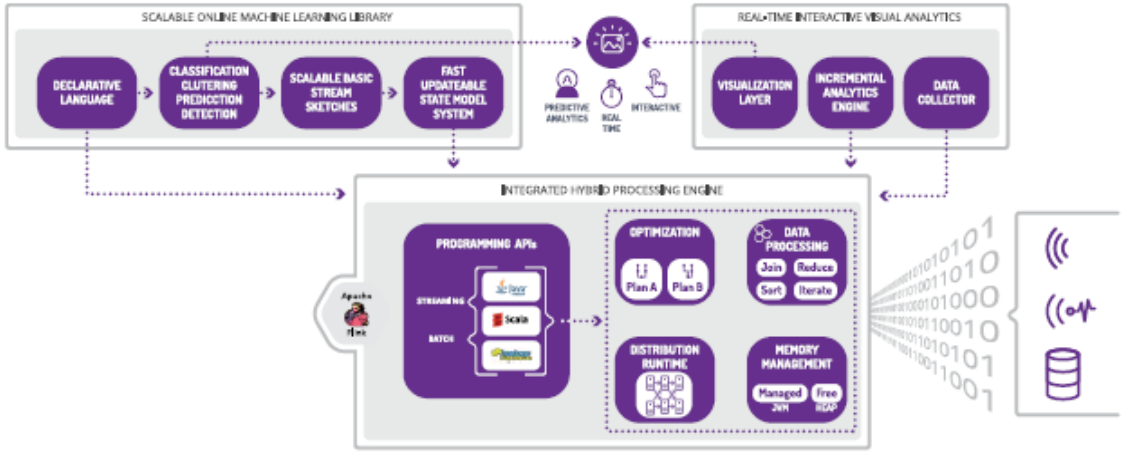
In particular, the project will go beyond the current state-of-art technology by making the following **specific original contributions:**

- New strategies for real-time hybrid computation, batch data and data streams.
- Real-time scalable machine learning for massive, high-velocity and complex data streams analytics.
- Real-time interactive visual analytics for Big Data. Implementation the new advances on top of Apache Flink.
- Real-world industrial validation of the technology developed.

The **PROTEUS** impact is manifold:

- strategic, by reducing the gap and dependency from the US technology, empowering the EU Big Data industry through the enrichment of the EU platform Apache Flink;
- economic, by fostering the development of new skills and new job positions and opportunities towards economic growth;
- industrial, by considering real-world requirements from industry and by validating the outcome on an operational setting, and
- scientific, by developing original hybrid and streaming analytic architectures that enable scalable online machine learning strategies and advanced interactive visualisation techniques that are applicable for general data streams in other domains.

PROJECT DESCRIPTION



PROTEUS is presented as three key technology components (hybrid computation model for both data-at-rest and data-in-motion, scalable online machine learning and real-time interactive visual analytics) integrated into the existing **Apache Flink, the EU Big Data platform**, to provide a validated solution for specific problems in an industrial setting related to steelmaking.

However the project contributions will be generic and context-independent to be applied in all data stream driven domains.

The core innovations and value of **PROTEUS** are based on a new integrated processing engine for being able to apply complex analytics techniques at scale and for batch data (data-at-rest) and data streams (data-in-motion) in a hybrid-merge mode.

In this way, our predictive engine will be able to provide real-time predictions while self-adapts continuously to learn more complex and refined learning models.

Moreover, visual analytics will be scalable with decreasing latency (interactive) demands using a novel incremental approach that represents the information (both data-in-motion and incremental process of batch data) as data streams.

Figure 2: PROTEUS leaflet, main page

2.2.1 Abstract

The contents for the abstract of the project have been reduced to adapt to the structure and size of the leaflet, resulting on the following description:

PROTEUS mission is to investigate and develop ready-to-use scalable online machine learning algorithms and real-time interactive visual analytics to deal with extremely large data sets and data streams.

The foundation is the use of an optimized implementation of combined batch and streaming processing and building around this later scalable real time processes. New algorithms and techniques will form a library to be integrated into an enhanced version of Apache Flink.

PROTEUS addresses fundamental challenges related to the scalability and responsiveness of analytics capabilities. The requirements are defined by a steelmaking industrial use case, but the techniques developed are flexible and portable to other data stream-based domains.

2.2.2 Contributions and impact

Similar to the abstract, the expected contributions and impacts are described in a concise way:

The project will provide the following specific original contributions:

- *New strategies for real-time* hybrid *computation, batch data and data streams.*
- *Real-time scalable machine learning for massive, high-velocity and complex data streams analytics.*
- *Real-time interactive visual analytics for Big Data. Implementation the new advances on top of Apache Flink.*
- *Real-world industrial validation of the technology developed.*

The PROTEUS impact is manifold:

- *strategic, by reducing the gap and dependency from the US technology, empowering the EU Big Data platform Apache Flink;*
- *economic, by fostering the development of new skills and opportunities towards economic growth;*
- *industrial, by demonstrating the outcome on an industrial operational setting, and*
- *scientific, by developing original hybrid and streaming analytic architectures that enable scalable online machine learning strategies and advanced interactive visualization techniques.*

2.2.3 Project description

Finally, the project description is presented via a concise description of the technical approach of the project, together with a general diagram graphically illustrating the concepts.

The descriptive text is presented as follows:

PROTEUS presents three key technology components (hybrid computation model for both data-at-rest and data-in-motion, scalable online machine learning and real-time interactive visual analytics) integrated into Apache Flink, and will demonstrate the solution for specific problems in an industrial setting: steelmaking.

The core innovations and value of PROTEUS are based on a new integrated processing engine able to apply complex analytics techniques at scale for batch data (data-at-rest) and data streams (data-in-motion) in a hybrid-merge mode. This predictive engine will be able to provide real-time predictions while self-adapts continuously to learn more complex and refined learning models.

Moreover, visual analytics will be scalable with decreasing latency (interactive) demands using a novel incremental approach that represents the information (both data-in-motion and incremental process of batch data) as data streams.

The description introduces an overview diagram of the project's concepts:

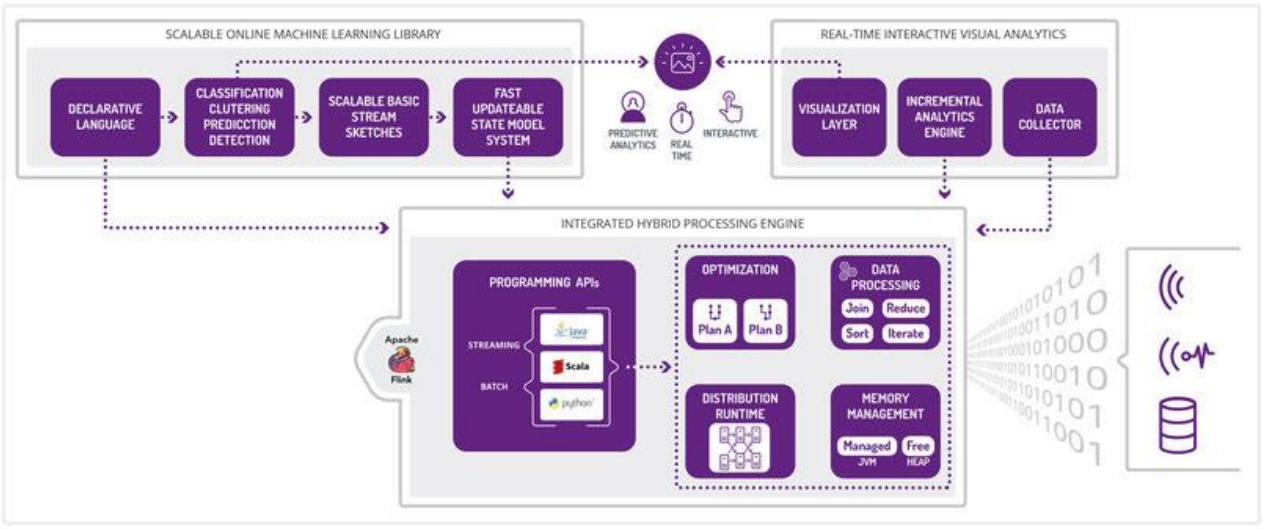


Figure 3: PROTEUS overview diagram

2.3 Back page

The back page includes the project's logo and the website URL, a list of the partners, and the acknowledgements of the project: "*This project is funded by the European Union (Horizon 2020, Ref: 687691)*", together with the EU flag.

Figure 4: PROTEUS leaflet, back page

3 Conclusions

This document provides detailed information about the design and contents of the leaflet at the time this report has been issued (Month 3 of the project). Although there are no future versions planned for this deliverable, the option for preparing future versions of a leaflet or other dissemination materials of the project is left open.