

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 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.

PROJECT DESCRIPTION

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

CONTRIBUTIONS & IMPACT

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.

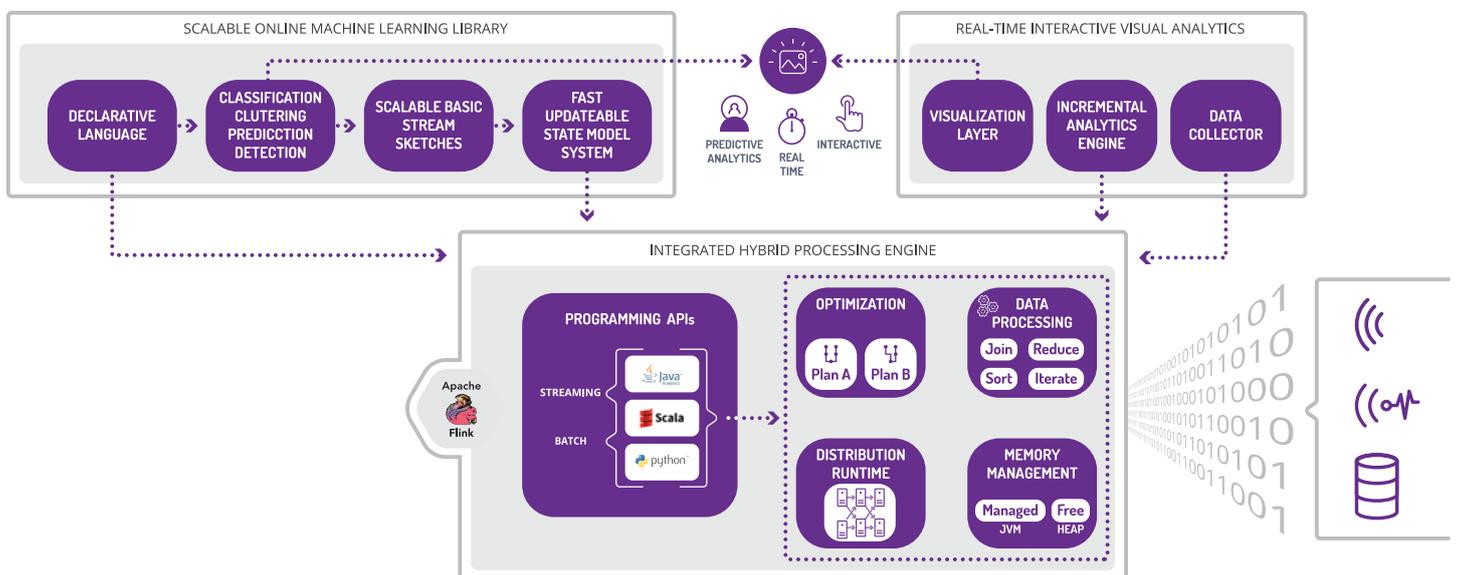
The **PROTEUS** impact is manifold:

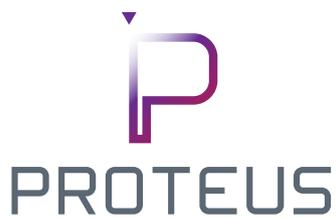
- ▶ strategic, by reducing the gap and dependency from the US technology, empowering the **EU Big Data platform Apache Flink**;
- ▶ economic, by fostering the development of new skills and opportunities towards economic growth;
- ▶ industrial, by demonstrating the outcome on an industrial operational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

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.

DURATION
36
 MONTH
 DEC. 2015
 NOV. 2018





www.proteus-bigdata.com

PARTNERS



SCALABLE ONLINE
MACHINE LEARNING
FOR PREDICTIVE
ANALYTICS AND
REAL-TIME
INTERACTIVE
VISUALIZATION



*This project is funded by the European Union
(Horizon 2020, Ref: 687691)*