



PROTEUS

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

687691

D6.3 Report on Community Engagement and Technology Transfer Activities – V1

Lead Author: DFKI

Reviewer: [TRI]

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	May 31 th 2017
Actual delivery date:	May 31 th 2017
Version:	1.2.2
Total number of pages:	14
Keywords:	Community, Engagement, Technology Transfer

Abstract

In this deliverable, we report the activities which we have done to transfer technologies developed within PROTEUS project to big data communities and industries. In particular, we describe how we engage with the Apache Flink community by taking part in community discussions on two main topics, namely, hybrid computation engine and online machine learning algorithms which are also two core components of PROTEUS project. Furthermore, we list our activities to transfer the technologies developed within PROTEUS project to industrial companies, partners and political delegations from across the globe. Moreover, to promote technology developed in PROTEUS, we describe our plans for participating in BDVA Summit 2017 and FlinkForward conference in September 2017.

Executive summary

This deliverable reports the activities which we have done to transfer technologies developed within PROTEUS project to big data communities and industries. This includes three main parts: engaging and contributing in Apache Flink open source community, building communities by attending big data events and presenting the contributions done by PROTEUS project to industrial companies.

To fulfil the first case, we engaged with Apache Flink community by taking part in community discussions on two main topics, namely, hybrid computation engine and online machine learning algorithms which are also two core components of PROTEUS project. To fulfil the second case, the PROTEUS consortium members have attended three big data summits and conferences in the first half of the project. To transfer technologies developed in the PROTEUS project, DFKI team presented the PROTEUS projects to several industrial companies, partners and political delegations from across the globe such as Italy, China, and Finland among the others who visited DFKI. Furthermore, we presented declarative language achievements in several invited talks in companies (done by DFKI partner) and we attended big data meetups (done by TREELOGIC).

Finally, we describe some of our tentative plans for the future including participating in BDVA Summit 2017 and FlinkForward conference in September 2017.

Document Information

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

Deliverable	Number	D6.3	Title	Report on community engagement and technology transfer activities
Work Package	Number	WP6	Title	Communication, exploitation and dissemination

Date of Delivery	Contractual	M18	Actual	M18
Status	version 2.0		final	<input type="checkbox"/>
Nature	report			
Dissemination level	public			

Authors (Partner)	DFKI			
Responsible Author	Name	Alireza Rezaei Mahdiraji	E-mail	alireza.rm@dfki.de
	Partner	DFKI	Phone	+49 30 23895 6627

Abstract (for dissemination)	In this deliverable, we report the activities which we have done to transfer technologies developed within PROTEUS project to big data communities and industries. In particular, we describe how we engage with Apache Flink community by taking part in community discussions on two main topics, namely, hybrid computation engine and online machine learning algorithms which are also two core components of PROTEUS project. Furthermore, we list our activities to transfer the technologies developed within PROTEUS project to industrial companies and partners. Moreover, to promote technology developed in PROTEUS, we describe our plans for participating in BDVA Summit 2017 and FlinkForward conference in September 2017.
Keywords	Community, Engagement, Technology Transfer

Version Log			
Issue Date	Rev. No.	Author	Change
May 3	1.0.0	Alireza Rezaei Mahdiraji	First version
May 9	1.1.0	Rachel Finn	First review: few major and some minor changes requested
May 17	1.2.0	Alireza Rezaei Mahdiraji	Fix issues mentioned in 1.1.0
May 26	1.2.1	Rachel Finn	Second review: few editorial issues
May 26	1.2.2	Alireza Rezaei Mahdiraji	Fix issues mentioned in 1.2.1

Table of Contents

Executive summary.....	3
Document Information.....	4
Table of Contents.....	5
1. Introduction.....	6
2. Community Engagement	7
In this Section, we describe our activities to engage into Apache Flink community and build PROTEUS community by attending big data and machine learning events.	7
2.1 Engaging in Apache Flink Community.....	7
2.1.1 Discussing Hybrid Processing on Flink Mailing List	7
2.1.2 Discussing SOLMA on Flink Mailing List.....	8
2.2 Community Building via Big Data and ML Events	8
2.2.1 BDVA Summit 2016.....	8
2.2.2 Participation in FlinkForward 2016 Conference.....	8
2.2.3 Big Data Excellence in UK and Germany Workshop.....	9
3. Technology Transfer	10
3.1 Invited Talks.....	10
3.2 Workshops and Conferences.....	10
3.2.1 Big Data Spain 2016	10
3.2.2 Technology Transfer at EuroPro 2017 Workshop	10
3.3 Industrial Transfer at DFKI.....	11
3.4 Future Engagement.....	11
3.4.1 BDVA Summit 2017.....	12
3.4.2 FlinkForward 2017.....	12
4. Conclusions.....	13
References.....	14

1. Introduction

The goal of WP6 of the PROTEUS project (i.e., communication, exploitation and dissemination) is to connect the project to external worlds. A major step to reach this goal is to coordinate activities to engage with developers of open-source communities and ease involvement of external stakeholders to extend the infrastructure developed within PROTEUS project by providing tutorials, talks, and workshops about the project achievements.

This deliverable reports the activities that the PROTEUS consortium has done to transfer technologies developed within the PROTEUS project to the open-source community, big data communities and industrial companies. We categorize our efforts in three main categories: 1) engaging and contributing in Apache Flink open-source community, 2) building communities by attending big data events and 3) presenting the contributions achieved by the PROTEUS project to stakeholders.

As part of the first category, we engaged with Apache Flink open-source community by taking part in community discussions on two main topics, namely, hybrid computation engine and online machine learning algorithms. These are two core components of PROTEUS project. We engaged in online developer community discussions about these topics and shared our ideas and designs, offered our help, and showed interests in adding more features and improvements related to the topics.

To fulfil the second case, the PROTEUS consortium members attended three big data summits and conferences. These events consist of audiences from companies and academia and provided ample opportunities to discuss the project and draw attentions to its achievements.

To transfer technologies developed in the PROTEUS project, DFKI team presented the PROTEUS projects to several industrial companies, partners and political delegations from across the globe who visited DFKI. Furthermore, members of DFKI presented declarative language achievements in several invited talks in companies and big data meetups.

In the future, we plan to organize and attend more events to promote PROTEUS to end-users and industries. In particular, we describe our tentative plans for organizing a joint workshop in BDVA Summit 2017 and presenting in FlinkForward conference in September 2017.

This deliverable is organized as follows: Section 2 describes our activities for community engagement. Section 3 presents what we have done to transfer PROTEUS technology to companies. Finally, Section 4 concludes this deliverable.

2. Community Engagement

In this Section, we describe our activities to engage with the Apache Flink community and build a PROTEUS community by attending big data and machine learning events.

2.1 Engaging in Apache Flink Community

The main way to take up a new feature in an open-source community such as Apache Flink is to motivate and promote the feature in the community. The central tool to engage with the Flink community is via the Flink mailing-list [1]. Therefore, we engaged with the Flink community by discussing some PROTEUS features on the developers mailing list. Our efforts were manifold, i.e., we discussed about contributing to Flink through ideas, design documents, feedbacks, reviews, and implementations. We mainly provided our thoughts and offered our help regarding Hybrid Processing (i.e., PROTEUS Engine) and Scalable Online Machine Learning library (i.e., SOLMA).

To familiarize the reader with the life cycle of getting a feature accepted in Apache Flink community, we list all the necessary steps which should be taken to achieve such goal:

- the feature should be discussed in the mailing list
- once agreed on the feature (i.e., is it interesting and required by the community), one has to write a Flink Improvement Plan (FLIP) document which contains a description of the feature, how the APIs might look like, the impact, etc
- when the community agrees on the proposed FLIP, then a Flink PMC/shepherd (along with the feature proposer and anyone else who is willing to contribute) open a JIRA issue [21] and start implementing
- create a pull request on github
- pull request is reviewed by PMCs/shepherd
- if the pull request satisfies the feature FLIP and standard development practices then it gets merged

Based on the steps listed above which are required to actually merge a contribution to Flink engine, one can see that the final decision to merge a feature is in the hand of main Flink contributors and PMC (Project Management Community).

2.1.1 Discussing Hybrid Processing on Flink Mailing List

Hybrid processing is a feature that has been discussed on Flink mailing list [2]. We engaged into the discussion by providing our point of view regarding this feature through a designated design document [3].

The discussion was open and ongoing for several months. Finally, one of the main developers of Apache Flink decided to work on that feature after few months, and this resulted in a Flink Improvement Plan (FLIP), referencing our first design document as an attempt to bring the hybrid processing into Flink [4].

As we were interested in contributing on this feature, we also jumped into the discussion by pointing out our thoughts about the possible approaches to the problem [5].

Our feedback regarding this topic was well received by the community. The main developers and contributors of Flink approved our feedback and decided to implement the feature as we suggested. We also offered our software engineering skills to finally implement the hybrid processing into Flink by taking over a JIRA issue [6].

However, as main Flink developers are currently busy with releasing Flink 1.3, the focus on the hybrid processing has temporally been lost.

2.1.2 Discussing SOLMA on Flink Mailing List

Being a streaming-first processing engine, the Apache Flink community is very interested in streaming scenarios such as online machine learning algorithms.

In the machine learning discussion of Flink mailing list, we informed the community about our effort in developing SOLMA and we offered our help and expertise in building a Flink Online Machine Learning by reusing SOLMA [7]. We have been listed as interested party regarding online machine learning [8].

However, currently Flink community seems to prefer offline learning on Streaming API rather than pure online machine learning due to lack of industrial use-cases. Our plan for the future is to continue monitoring the online ML discussion on the Flink mailing list, provide update about our progress on implementing SOLMA (Scalable Online Machine Learning Algorithm) library of PROTEUS on the mailing list, and offer our skills and experiences in implementing such algorithms in the Apache Flink if the community decided to integrate the algorithms as part of the Flink.

2.2 Community Building via Big Data and ML Events

Members of PROTEUS consortium have attended several big data conferences and summits to learn about other projects and promoted PROTEUS contributions within big data communities. In the following subsections, we briefly explained the events.

2.2.1 BDVA Summit 2016

The Big Data Value Association (BDVA) [9] is a fully self-financed not-for-profit organisation under Belgian law which implements the BDV PPP programme (Big Data Value PPP). The BDVA has over 150 members in Europe composed of industries and research organizations.

BDVA is a natural fit for several purposes, including a research network, dissemination, communication and exploration activities. Two members of the PROTEUS consortium, namely, TRI and TREE were present in the BDVA Summit 2016 (November 30th to December 2nd 2016, in Valencia - Spain). The partners used the summit as an opportunity for networking, awareness raising, community building and exploration of future cooperation and synergies. For more details on BDVA 2016 please refer to the deliverable D6.7 Report on Project Communication and Engagement Activities.

2.2.2 Participation in FlinkForward 2016 Conference

FlinkForward conference (September 2016, Berlin, Germany) [10] was the second conference to bring together the open source stream processing community. It is the ideal venue to learn more about the Apache Flink and other streaming projects and connect with the community. Two members of TREELOGIC and all the members of DFKI PROTEUS team attended the conference.

Participants from several major companies such as Netflix, Alibaba, Zalando, King, ResearchGate, and Uber reported their success stories of deploying Flink in production, processing tens of billions of events per day, and scaling Flink to a large number of computer nodes. They shared the challenges they have faced and the solutions they have discovered while deploying Flink in production at very large scale.

Most of the Apache Flink committers attended the conference and several of them gave in-depth technical talks about improvements and new features of Flink and the future development of Flink.

This provided numerous opportunities for the attendees of PROTEUS project to engage in interesting and inspiring discussions and exchange experiences and ideas on the stream-based big data analytics and its future directions. In particular, we aimed to attend all the talks which were related to two main objectives of PROTEUS, namely, declarative language and online scalable machine learning. There of the most relevant talks are as follows:

- Apache Beam: A Unified Model for Batch and Streaming Data Processing [22]
- Streaming ML with Flink [23]
- AMIDST Toolbox: Scalable probabilistic machine learning with Flink [24]

2.2.3 Big Data Excellence in UK and Germany Workshop

Big Data Excellence in UK and Germany Workshop [11] is a joint event by the Berlin Big Data Center (BBDC) and UK Science and Innovation Network held on March 1st, 2017 in Smart Data Forum which is managed by DFKI. The audience heard about UK's big data landscape and the Berlin Big Data Centre projects, with sessions on data analytics, big data management, machine learning and E-health. There has been discussion about how the UK and Germany are building up national networks for interdisciplinary collaboration and how researchers are collaborating with government and industry in practice. DFKI team exchanged information about ongoing projects in DFKI with attendees from UK and others from Germany. Moreover, members of DFKI presented the contributions related to declarative languages and stream processing. This includes technical details about the Emma and Lara declarative languages (please see the deliverable D3.3 for details about these languages).

3. Technology Transfer

We categorize our activities related to the PROTEUS technology transfer into three categories, namely, invited talks, workshops and conferences, and industrial transfer at DFKI. These categories are explained in the following subsections.

3.1 Invited Talks

Members of DFKI team gave invited talks in several companies. These talks covered aspects of PROTEUS project mostly related to the hybrid processing and declarative language. In the following, we provide the list of the companies where an invited talk was given:

- Amazon Berlin @ Amazon Labs, Berlin, Germany
- Big Data Beers @ SAP Berlin, Germany
- SAP, Berlin, Germany

3.2 Workshops and Conferences

In the following, we describe big data two events in which PROTEUS was presented by members of the PROTEUS consortium.

3.2.1 Big Data Spain 2016

The Big Data Spain [12] is international business-oriented conference with focus on big data technology held yearly in Madrid. Members of TREELOGIC (the PROTEUS coordinator partner) presented the PROTEUS project contributions on performing scalable advanced analytics on data stream and incremental visualization of big data. For more details please refer to the deliverable D6.7 Report on Project Communication and Engagement Activities.

3.2.2 Technology Transfer at EuroPro 2017 Workshop

PROTEUS was present in EuroPro 2017 workshop by an abstract and an oral presentation. EuroPro is the 1st International Workshop on Big Data Management in European Projects collocated with the EDBT/ICDT 2017 joint conference held in Venice, Italy [14]. The main aim of this workshop is “to share experiences and best practices, discuss challenges and effective solutions adopted, and investigate opportunities for collaboration among European projects (various directorates of the European Commission or other European funding agencies) dealing with big data management”. DFKI attended the workshop and presented a brief overview of PROTEUS in front of audiences of sister projects of PROTEUS. A following panel session let presenters and attendees take part into an open discussion regarding Big Data Management. The core topic which was discussed in the panel session was the need for declarative languages for scalable data processing. A DFKI team member led the discussion along this line as a declarative language is one of the main output of PROTEUS. Furthermore, many presenters of other projects were interested in the technical aspects of PROTEUS.

The other projects presented in this workshop were:

- datACron: Big Data Analytics for Time Critical Mobility Forecasting [15]
- Toreador: Trustworthy model-aware Analytics Data platform [16]
- STREAMLINE [17]
- My Health, My Data [18]
- Supersede [19]
- Big Data Europe [20]

3.3 Industrial Transfer at DFKI

DFKI hosts many visitors from companies and political delegations both nationally and internationally. The team members of DFKI / IAM exploit each of these opportunities to promote the contributions provided by our projects and in particular the PROTEUS project.

A list of companies for which the DFKI/IAM members conducted workshops and talks presenting and promoting technology provided by the PROTEUS project (among the other projects) are as follows:

- Amazon Research Germany
- Zalando
- Audi (Audi Production Lab)
- Rolls-Royce Deutschland
- T-Systems
- SAP Corporate Strategy Group
- Teradata (CTO)
- IBM Research (Böblingen, Germany)
- Siemens CT
- Deloitte Analytics Institute
- Volkswagen AG

A list of political delegations which visited DFKI and were introduced to PROTEUS (among other projects) are as follows:

- Stefania Giannini, the Italian Minister for Education and Research, a delegation of representatives of: 28 EU member states;
- Important representatives of industry and administration of the Chinese Region Maanshan with the mayor of Maanshan, Yao Wie;
- Finnish companies;
- Visitors of the National IT-Summits;
- Andrus Ansip, Vice-President of the European Commission for the Member States of the European Communities;
- The Republic of Singapore and of the Civil Service College Singapore.

3.4 Future Engagement

In the future, we continue to leverage DFKI industrial and academic contacts to present and promote the PROTEUS projects. Moreover, we have a planned list of activities all related to the technology transfer by attending big data events, meet ups (e.g., the docker meet-up in Mardird). We also plan to see how the contributions of the PROTEUS project can influence big data standardization. To this end, we will organize meeting with SPEC Research Big Data Working Group [25] to show our achievement in benchmarking hybrid computation.

In the following, we explain our plan to attend two big data events in order to be able to present PROTEUS achievements in more details.

3.4.1 BDVA Summit 2017

PROTEUS is engaged with the Big Data Value Association (BDVA) and is in constant contact with the association since the execution of the project. One example, as introduced in section 2.2.1, is the participation in the annual summit events organised by the platform, bringing together stakeholders of all kind around the Big Data Value chain. Out of the partners in PROTEUS, Treelogic and Trilateral Research are Associate Members of the BDVA, and active in events and activities within this platform.

PROTEUS is planning to have a more active and central role in the BDVA, and in particular we are planning to organise a workshop in liaison with the BDVA as part of our WP6 activities. The aim is to conduct a joint workshop as a working session with other projects, especially other projects funded under the same call, but also other related initiatives under the umbrella of the BDVA PPP. At the time of this report, we are evaluating all aspects of our participation towards the forthcoming BDVA Summit 2017, being planned and expected to take place in Versailles (France) in November 2017.

3.4.2 FlinkForward 2017

As we briefly described FlinkForward conference in Section 2.2.2 of this deliverable, FlinkForward is an appropriate venue for fulfilling the requirements of WP6 of PROTEUS project. The main reasons for this is that FlinkForward allows us to directly meet Flink community (i.e., developers, end-users, industrial companies), share achievements of PROTEUS, and extend PROTEUS user community.

To this end, we commenced a series of activities to meet requirements for submitting a presentation proposal to FlinkForward 2017 which will be held in Berlin in the coming September. So far, we took two initial step. First, we have collected information about successful proposal for FlinkForward. Based on the previous proposal guideline [13], a successful presentation proposal should tell a story of an innovative application with actual use-cases which uses best practices for solving big data challenges. Second, we have collected a short description of each of consortium member activities within the project.

Our next step will be to compile the contributions of all the partners in a concise abstract which tell a story about the industrial use-case of PROTEUS and how PROTEUS solutions solve challenges of the use-case.

As a final note, we would like to draw the reader attention to a main challenge which we are facing in participating in FlinkForward in the upcoming September. The challenge is that a lot of main technical contributions of PROTEUS which includes most of its novel scalable online machine learning algorithms are not developed yet and will be done in a later stage of the project. This may hinder us to come up with a story as part of a bigger picture (i.e., solving industrial big data challenges) which excites the audiences.

4. Conclusions

In this deliverable, we reported our activities to engage in open-source communities and transfer technologies developed in PROTEUS project to big data communities and industries. Since the project is built on top of Apache Flink open-source project, we aimed to engage in Flink community by discussing hybrid computation and online machine learning algorithms in the Flink mailing list. This led to gain attention from the community about our effort in PROTEUS. We attended big data events and exchanged the project technical details with end-users and industrial companies. We also presented the contributions done by PROTEUS project to industrial companies and political delegations to build communities and draw attentions of end-users and industries to PROTEUS achievements. In the future, we plan to organize and attend more big data conference and summit. In particular, we showed our tentative plans for organizing and participating in several big data events.

References

- [1] Apache Flink Community & Project Info: <https://flink.apache.org/community.html>
- [2] Add Side Input/Broadcast Set For Streaming API: <http://apache-flink-mailing-list-archive.1008284.n3.nabble.com/DISCUSS-Add-Side-Input-Broadcast-Set-For-Streaming-API-td11529.html>
- [3] Apache Flink Side Inputs Proposal: <https://goo.gl/P2Mps5>
- [4] FLIP-17 Side Inputs for DataStream API <https://cwiki.apache.org/confluence/display/FLINK/FLIP17+Side+Inputs+for+DataStream+API>
- [5] FLIP-17 Side Inputs Discussion: <https://lists.apache.org/thread.html/e5fae9bb51ce37eb1d54fa4ecf9a01a663063cd06357c7b2f280ba5f@%3Cdev.flink.apache.org%3E>
- [6] Add buffering service for stream operators: <https://issues.apache.org/jira/browse/FLINK-6141>
- [7] Machine Learning on Flink - Next steps: <http://apache-flink-mailing-list-archive.1008284.n3.nabble.com/Machine-Learning-on-Flink-Next-steps-td16334.html>
- [8] Flink ML Roadmap: <https://docs.google.com/document/d/1afQbvZBTV15qF3vobVWUjxQc49h3Ud06MIRhahtJ6dw/edit#heading=h.txzqce8aw2b6>
- [9] BDVA Summit 2016 in Valencia: <http://www.bdva.eu/?q=valencia-summit>
- [10] FlinkForward 2016: <http://2016.flink-forward.org>
- [11] Big Data Excellence in UK and Germany Workshop: <http://smartdataforum.de/big-data-excellence-in-uk-and-germany/>
- [12] Big Data Spain: www.bigdataspain.org
- [13] Flink Forward Call for Submissions: <http://us12.campaign-archive2.com/?u=0db7e9ab89e059d850c86f41a&id=855d0d953c&e=%5BUNIQID%5D>
- [14] EDBT/ICDT 2017: <http://edbticdt2017.unive.it>
- [15] datACron: <http://ai-group.ds.unipi.gr/datacron/>
- [16] Toreador: <http://www.toreador-project.eu>
- [17] STREAMLINE: <https://streamline.sics.se>
- [18] My Health, My Data: <http://www.myhealthmydata.eu>
- [19] Supersede: <https://www.supersede.eu/project/supersede-at-a-glance/>
- [20] Big Data Europe: <https://www.big-data-europe.eu>
- [21] JIRA Software: <https://de.atlassian.com/software/jira>
- [22] Kenneth Knowles. Apache Beam: A Unified Model for Batch and Streaming Data Processing. http://2016.flink-forward.org/kb_sessions/apache-beam-a-unified-model-for-batch-and-streaming-data-processing/
- [23] Márton Balassi. Streaming ML with Flink. http://2016.flink-forward.org/kb_sessions/streaming-ml-with-flink/
- [24] Ana M. Martinez. AMIDST Toolbox: Scalable probabilistic machine learning with Flink. http://2016.flink-forward.org/kb_sessions/amidst-toolbox-scalable-probabilistic-machine-learning-with-flink/
- [25] SPEC Research Big Data Working Group. <https://research.spec.org/working-groups/rg-big-data.html>