

Project Title	High-performance data-centric stack for big data applications and operations
Project Acronym	BigDataStack
Grant Agreement No	779747
Instrument	Research and Innovation action
Call	Information and Communication Technologies Call (H2020-ICT-2016-2017)
Start Date of Project	01/01/2018
Duration of Project	36 months
Project Website	<a href="http://bigdatastack.eu/">http://bigdatastack.eu/</a>

## D7.3 – Exploitation Plan and Business Potential-V2

Work Package	WP7 – Communication, Exploitation, Standardisation, & Roadmapping
Lead Author (Org)	Ana Belén González (ATOS)
Contributing Author(s) (Org)	Yosef Moatti (IBM), Pavlos Kranas (LXS), Luis Tomar Bolivar (RHT), Marieke Williems (TRUST-IT), Orlando Avila (ATOS), Maurizio Megliola (GFT), Sophia Karagiorgou (UBI), Nikitas Sgouros (UPRC) Maria Patiño (UPM), Ainhoa Azqueta (UPM), Amaryllis Raouzaiou (ATC), Bernat Quesada (WL), Richard Maccreadie (GLA), Anestis Sidiropoulus (ATC), Mauricio Fadel (NEC), Anestis Sidiropoulos (ATC), Christos Doulkeridis (UPRC), Jean-Didier Totow (UPRC), Stavroula Meimetea (UPRC), Stathis Plitsos (DANAOS)
Due Date	30.10.2020
Date	09.12.2020
Version	1.0

### Dissemination Level

<input checked="" type="checkbox"/>	PU: Public (*on-line platform)
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

### Versioning and contribution history

Version	Author	Notes
	Ana Belén González (ATOS)	Elaboration of Individual Exploitation Template to collect partners' final exploitation plans. Sent to all partners to be filled in.
	Ana Belén González (ATOS)	Elaboration of Components Template to collect updated information from SW components. Sent to all partners to be filled in.
0.1	Ana Belén González (ATOS)	ToC and initial draft
0.2	Ana Belén González (ATOS)	Integration of Partners' Individual Exploitation Plans from all partners (Section 5.2). Contributions from: Yosef Moatti (IBM), Pavlos Kranas (LXS), Luis Tomar Bolivar (RHT), Marieke Williems (TRUST_IT), Orlando Avila (ATOS), Maurizio Megliola (GFT), Sophia Karagiorgou (UBI), Maria Patiño (UPM), Ainhoa Azqueta (UPM), Amaryllis Raouzaïou (ATC), Bernat Quesada (WL), Richard Maccreadie (GLA), Anestis Sidiropoulos (ATC), Mauricio Fadel (NEC), Anestis Sidiropoulos (ATC), Christos Doukeridis (UPRC), Jean-Didier Totow (UPRC), Stavroula Meimetea (UPRC), Stathis Plitsos (DANAOS)
0.3	Ana Belén González (ATOS)	Integration of information about SW components from all partners (section 4.3) and Workshop outcomes (4.2). Contributions from: Yosef Moatti (IBM), Pavlos Kranas (LXS), Luis Tomar Bolivar (RHT), Marieke Williems (TRUST_IT), Orlando Avila (ATOS), Sophia Karagiorgou (UBI), Maria Patiño (UPM), Ainhoa Azqueta (UPM), Amaryllis Raouzaïou (ATC), Richard Maccreadie (GLA), Anestis Sidiropoulos (ATC), Mauricio Fadel (NEC), Anestis Sidiropoulos (ATC), Christos Doukeridis (UPRC), Nikitas Sgouros (UPRC), Jean-Didier Totow (UPRC)
0.4	Ana Belén González (ATOS)	Elaboration of sections: 2, 3, 4.1, 4.2, and 5.1 and 5.3
0.5	Ana Belén González (ATOS)	All partners provide information about ongoing projects and initiatives, and check information from exploitable assets
0.6	Ana Belén González (ATOS)	Include Executive Summary and Conclusions
0.7	ATOS	Internal quality check from ATOS
0.8	Yosef Moatti (IBM), Dimosthenis Kyriazis (UPRC)	Internal Review
0.9	Ana Belén González (ATOS)/ Dimosthenis Kyriazis (UPRC)	Comments from internal reviewers addressed
1.0	Ana Belén González (ATOS)	Final version for submission

**Disclaimer 1**

This document contains information that is proprietary to the BigDataStack Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the BigDataStack Consortium

**Disclaimer 2**

This report contains information extracted from the previous deliverable D7.2 Exploitation Plan and Business Potential-V2 also led by ATOS, which was the first one of the series of deliverables resulting from Task 7. 1 Market Analysis, Business Scenarios and Exploitation.

The information from BigDataStack assets have been provided by the BigDataStack partners and may have similarities with the information included in deliverables from technical WPs.

## Table of Contents

1. Executive Summary .....	9
2. Introduction .....	10
2.1. Purpose and Scope .....	10
2.2. Structure of the document .....	10
3. Exploitation Objectives and Activities .....	11
3.1. Activities with external stakeholders .....	11
4. BigDataStack Exploitable Assets and MVPs .....	18
4.1. BigDataStack Exploitable Assets .....	18
4.2. BigDataStack MVPs .....	22
4.3. Updated information for BigDataStack Exploitable Assets .....	29
4.3.1. SQL Data Skipping .....	29
4.3.2. Dynamic Orchestration .....	30
4.3.3. Network Policy support at Kuryr .....	32
4.3.4. Kuryr Cluster Network Operator .....	33
4.3.5. NVME-mdev Kernel driver .....	33
4.3.6. Integration of ovn distributed load-balancer at Kuryr for E/W traffic ....	34
4.3.7. Infrastructure management through API.....	35
4.3.8. QoS Evaluator .....	36
4.3.9. Information Driven Networking .....	37
4.3.10. Process Modelling Framework .....	38
4.3.11. Visualization Environment .....	39
4.3.12. Seamless Analytics .....	40
4.3.13. Adaptable Distributed Storage.....	41
4.3.14. Data Toolkit .....	42
4.3.15. Data Quality Assesment .....	43
4.3.16. Application Dimensioning Workbench .....	43
4.3.17. Process Mapping.....	45
4.3.18. Complex Event Processing (CEP).....	46
4.3.19. Beta-RecSys Recommendations Development Framework.....	46
4.3.20. Realization Engine .....	48

5.	BigDataStack Exploitation and Sustainability Strategy.....	50
5.1.	Exploitation and Sustainability Strategy .....	50
5.1.1.	Partners Exploitation Plans for BigDataStack assets .....	50
5.1.2.	Joint exploitation in ongoing and further R&D projects.....	52
5.1.3.	Commercial Exploitation.....	54
5.1.4.	Patents .....	55
5.1.5.	Technology Transfer.....	55
5.1.6.	BigDataStack OS approach and upstream contribution to OS Communities .....	55
5.1.7.	BigDataStack joint to BDVA Innovation Marketplace .....	57
5.1.8.	EOSC-DIH Collaboration Agreement .....	57
5.1.9.	Future R&D activities and education .....	58
5.2.	Partners’ Final Exploitation Plans.....	60
5.2.1.	IBM Israel – Science and Technology LTD for SQL Data Skipping.....	60
5.2.2.	NEC Europe LTD .....	61
5.2.3.	RED HAT Israel LTD .....	62
5.2.4.	ATOS Spain SA.....	64
5.2.5.	Worldline Iberia SA.....	65
5.2.6.	GFT Italia SRL.....	67
5.2.7.	DANAOS Shipping Company LTD .....	69
5.2.8.	SingularLogic Romania Computer Applications SRL .....	70
5.2.9.	Athens Technology Center SA .....	71
5.2.10.	LeanXcale SL .....	73
5.2.11.	UBITECH Limited .....	74
5.2.12.	TRUST-IT Services Limited.....	75
5.2.13.	University of Piraeus Research Center.....	79
5.2.14.	Universidad Politecnica de Madrid .....	81
5.2.15.	University of Glasgow.....	82
5.3.	BigDataStack joint exploitation.....	86
5.3.1.	BigDataStack Commercialization Business Model .....	86
5.3.2.	Joint Exploitation Strategy .....	87
6.	Conclusions.....	91

## List of tables

Table 1. BigDataStack Exploitable Assets analysis .....	20
Table 2. BigDataStack Exploitable Use Cases .....	22
Table 3. MVPs from BigDataStack Offering .....	24
Table 4. Discussion on MVPs from BigDataStack exploitable assets (I).....	26
Table 5. Discussion on MVPs from BigDataStack exploitable assets (II).....	28
Table 6. BigDataStack MVPs .....	28
Table 7. Most representatives exploitation activities from Partners' exploitation plans .....	51
Table 8. BigDataStack commercialization Business Model Canvas .....	87

## List of figures

Figure 1. BigDataStack Key Offering.....	18
Figure 2. Mural from Activity 1.....	23
Figure 3. Murals from Activity 2.....	25
Figure 4. BigDataStack Exploitation and Sustainability strategy.....	50

## Acronyms

API	Application Program Interface
ATC	Athens Technology Center SA
B	Billions
BDVA	Big Data Value Association
BM	Business Model
BMC	Business Model Canvas
CE	European Commission
CNI	Container Network Interface
CO2	Carbon Dioxide
DANAOS	DANAOS Shipping Company Limited
DB	Data Base
DNS	Domain Name System
EBDVF	European Big Data Value Forum
EOSC	European Open Science Cloud
EOSC-DIH	European Open Science Cloud – Digital Innovation Hub
EU	European

EW	East/West
GDPR	General Data Protection Regulation
GFT	GFT Italia SRL
GLA	University of Glasgow
HTTP	Hypertext Transfer Protocol
H2020	Horizon 2020
IBM	IBM Israel-Science and Technology LTC
ICB	Impact Creation Board
ICT	Information and Communication Technology
IoT	Internet of Things
IPR	Intellectual Property Right
IT	Information Technology
KPI	Key Performance Indicators
LXS	LeanXcale SL
M	Month
ML	Machine Learning
MVP	Minimum Valuable Product
NEC	NEC Europe LTD
OS	Open Source
OSI	Open System Interconnection
PoC	Proof of Concept
QoS	Quality of Service
R&D	Research and Development
RHT	Red Hat Israel LTD
RL	Reinforced Learning
SDN	Software defined networking
SILO	Singularlogic Romania Computer Application SRL
SLO	Service Level Objectives
SQL	Structured Query Language
SME	Small and Medium Enterprise
SW	Software
SWOT	Strengths, Weaknesses, Opportunities, and Threats
TRL	Technology Readiness Level
UBI	Ubitech Limited
UI	User Interface
UK	United Kingdom

UPM	University of Madrid
UPRC	University of Piraeus Research Center
URL	Uniform Resource Locator
US	United States
UX	User Experience
WP	Workpackage

## 1. Executive Summary

This deliverable reports the exploitation activities carried out by BigDataStack partners during the second period of the project and the achievements in terms of exploitation and commercialization of the BigDataStack project.

During the project lifetime, twenty-one exploitable assets have been identified, eight Minimum Valuable Products (MVP), as well as three exploitable business Use Cases. Also, three patents have arisen, and a technology transfer activity is in progress.

Even though the COVID-19 pandemic situation, all exploitation activities have been properly performed. All the partners have developed their exploitation plans after the project, the partners have also already achieved remarkable exploitation results during the project, such as: commercialization of Data Skipping component by IBM and the Kuryr and Infrastructure API by RHT, Use Case partners are trying to leverage the use cases prototypes with their customers, most partners are already participating together in several Research & Development (R&D) projects to jointly exploit their Software (SW) components and BigDataStack has contributed actively to Open Source (OS) communities.

Different exploitation activities have been organized in the second period of the project among partners as well as with external stakeholders. Among them, a remote workshop to discuss the exploitable assets and identifying the MVPs and two templates regarding exploitation aspects of the components have been filled in by partners, and finally, a discussion panel with big data experts organized within the EBDVF2020<sup>1</sup> event to present BigDataStack outcomes and gather feedback on the adoption of BigDataStack results.

Additionally, BigDataStack Consortium has laid the ground for the sustainability of the project and the involvement of project partners after the project, by joining to the BDVA Innovation Marketplace<sup>2</sup> and the EOSC-DIH<sup>3</sup>.

---

<sup>1</sup> <https://www.european-big-data-value-forum.eu/>

<sup>2</sup> <https://marketplace.big-data-value.eu/>

<sup>3</sup> <https://eosc-dih.eu/>

## 2. Introduction

### 2.1. Purpose and Scope

---

This deliverable is the second of a set of two Exploitation and Business Potentials deliverables included in the BigDataStack Grant Agreement.

The objective of this deliverable is to report on the exploitation activities carried out within Task 7.1<sup>4</sup>, from M18-M36. The report is focused on the exploitation paths for BigDataStack components, partners' exploitation plans and BigDataStack joint exploitation.

### 2.2. Structure of the document

---

This document is structured in 3 main sections:

- **Section 3. Exploitation Objectives and Activities:** This section provides an overview of the exploitation activities carried out among partners and with external stakeholders during the second period of the project.
- **Section 4. BigDataStack Exploitable Assets and MVPs:** It provides updated information on the BigDataStack exploitable assets as well as which exploitable assets can be grouped into MVPs with high exploitation potentiality. Moreover, we include more business-related updated information from the exploitable assets.
- **Section 5. BigDataStack Exploitation and Sustainability Strategy:** This section presents the BigDataStack exploitation and sustainability strategy, and detailed information at the consortium and individual levels. It also provides the partners' final exploitation plans and BigDataStack joint exploitation strategy.

Finally, in the last section, we present the conclusions of the exploitation activities carried out during the project lifetime.

---

<sup>4</sup> Task 7.1 Market Analysis, Business Scenarios and Exploitation

## 3. Exploitation Objectives and Activities

BigDataStack project defined a series of Exploitation objectives for T7.1 and T7.4<sup>5</sup> at the beginning of the project. The exploitation objectives were firstly reported within the deliverable D7.2<sup>6</sup> in M18.

During the first period of the project, the exploitation team focused on conducting a market analysis, stakeholder analysis and value chain, SWOT analysis and define the BigDataStack value proposition. This information was mainly used to establish the project context and its positioning. BigDataStack partners also provided initial information of the exploitable assets and their preliminary exploitation intentions and interests on the exploitable outcomes of the project. Moreover, three business models were identified to exploit BigDataStack outcomes considering the partners' profiles.

Throughout the second period, the BigDataStack consortium principal focus was to identify and carry out exploitation opportunities for BigDataStack exploitable assets as well as establishing a realistic sustainability plan for project outcomes. The outcomes of these activities are extensively described within this deliverable.

The exploitation team has been also involved within T7.4 to define the Adoption Roadmap for BigDataStack solution, what has been reported within D7.5<sup>7</sup>.

It is worth mention that due COVID-19 situation, some planned activities with the Impact Creation Board (ICB) were cancelled and replaced by other activities with external stakeholders which are detailed below.

### 3.1. Activities with external stakeholders

---

BigDataStack had planned a workshop with the members of the Impact Creation Board (ICB). Due to the pandemic situation and problems with the agenda of some of the members, this workshop couldn't be organized, although BigDataStack consortium looks for other opportunities to collect feedback from external stakeholders.

The BigDataStack consortium organized a discussion panel with experts on big data domain within the EBDVF2020<sup>8</sup> event held 3-5 November 2020 titled "A BigDataStack for the industry – a focus on retail, shipping and insurance: Is Big Data the real future of emerging business?"<sup>9</sup>.

---

<sup>5</sup> T7.4 Roadmapping & Impact on Adoption

<sup>6</sup> D7.2 Exploitation plan and business potential I

<sup>7</sup> D7.5 Dissemination, Standardisation, and Adoption Roadmap

<sup>8</sup> <https://www.european-big-data-value-forum.eu/>

<sup>9</sup> <https://bigdatastack.eu/content/bigdatastack-industry-focus-retail-shipping-and-insurance-big-data-real-future-emerging>

Within this event, BigDataStack consortium presented project outcomes to the industry while gathering industry insights on the barriers and issues of adopting big data.

In this section, we present the feedback provided by the experts to the discussion held with them during the EBDVF2020 Event. (This information is extracted from the panel discussion and the records of the session, that are available within the BDVA website and BigDataStack website<sup>10</sup>).

## 1. So much data for industry. But how do you avoid the disadvantage of negative data exhaustion?

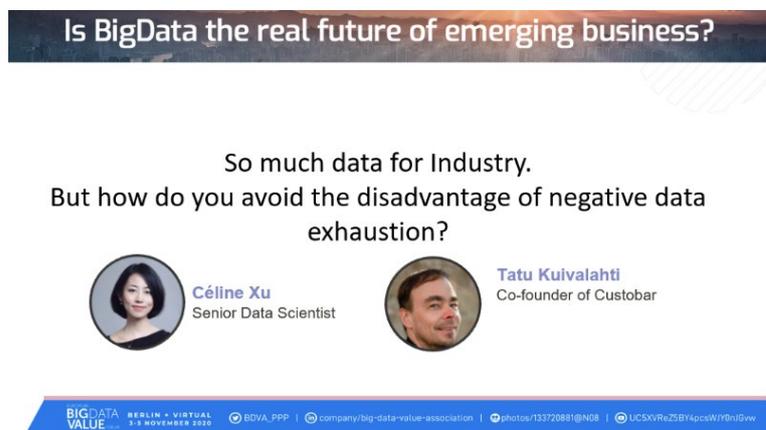
Celine Xu: “Data Exhaustion can refer to invaluable data or data that goes to waste.

Regarding data is wasted, there are five or 6 possible reasons:

- The data is not accurate which link with data quality issue, which leads to a lack of trust and people aren't able to use the data or they are not willing to use it.
- Retrieve: The data is stored in a non-accessible place
- Data is not in the right format and is not able to integrate with other existing data or the data that users have.
- Users don't know how to use those data and incorporate with their business cases
- Customers don't know that data exist

Some solutions to avoid this can be: Efficiency storage of the data and provision of easy access, right data format and standardized format as the industry have, provide guidelines on how to standardize existing data and workers' education.

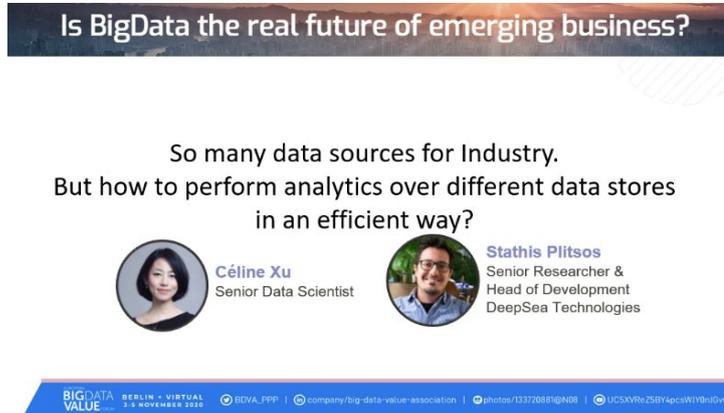
Additionally, Tatu Kuivalahti remarked “the importance of “visualize the data” and have tools for users and retailers (in the case of Retail sector) to visualize customers behaviour and have as much information as possible”. At this sense, Tatu Kuivalahti showed an interesting example of visualization of customer behaviour from a big data project of his company. The example aimed at showing the power of visualization tools to provide insights.



<sup>10</sup> <https://bigdatastack.eu/events/bigdatastack-industry-focus-retail-shipping-and-insurance-bigdata-real-future-emerging>

## 2. So many data sources for Industry. But how to perform analytics over different data stores in an efficient way?

Céline Xu: “All the analytics serve the same purpose: generate insights and actions, and for make analytics effectively, she remarks three points: a) understand which problem is going to solve, b) data needed for that, and c) how to use that data to get actionable insights. Celine Xu gave the example of one of her company’s project, in which they use different data to get insights for different target



customers; and how based on the problem we wanted to solve, we would retrieve data from different data sources and accordingly use different models to predict. Other issues to bear in mind is the different definition of the data across the databases, and the aggregation at the same level as the right time zoon what is crucial for analytics.

Data quality, for different analysis we don’t need the same level data quality for performance, i.e. for report level, maybe 20% of missing insignificant data is enough, but for predicting the customers’ trend, a 5% of missing data can make the model worse.

Another risk is the limitation of the data, for example, for big data projects in which we use the PIC of the mobile phone to know the situation of the people, maybe people have more than one mobile, so we need to understand possible risks and the limitations of the data, for efficient analysis.”

Both experts agreed that that is crucial to clearly define the problem we want to solve with analytics and the importance of make the right questions and select the right data.

## 3. Where does Europe sit in terms of its digital sovereignty towards Big Data?

## 4. Where are investments still needed in your industry sector?

Ray Walsh: “To have data sovereignty, we need to have control over our data, and for that purpose, the CE has developed the General Data Protection Regulation (GDPR<sup>11</sup>) to protect citizens and ensure that personnel identifiable data and information if it has been stored and restored securely and safely.

Digitalization has made that industry stakeholders are worldwide and when dealing with customers, data passing back and forth all the time and as Europeans, we want control over

<sup>11</sup> <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

our data and make sure that data has been treated the same way as if we would like to have treated within our jurisdiction. The emergence of 5G has also raised concerns at this respect. EU is evangelizing the rest of the world that this is a basic human right.

We need standardized playing field, we need everybody to subscribe to best practices when it comes to data, we need data governance and sovereignty, we need to ensure that the international community work together to achieve standardization, certification, regulation, legislation, concerning citizens data. GDPR is the starting point, but other EU initiatives are putting in place, such as white paper, EU data strategy. OS and OS standards are very important at this regard”.

Regarding the need of investments in the industry, for the Shipping sector, Stathis Plitsos mentioned “two pillars that should be addressed regarding investments: a) *Digitalization of the companies*, what includes, IoT solutions onboard, decision support systems, the integration on both tools, digital twins, etc. Stathis Plitsos commented how the Shipping sector is affected by the pandemic, and if they were invested in digitalization, some challenges they are facing now it would have been solved, and the other pillar is b) *Decarbonization* (also closely related with Digitalization): There is a regulation by IMO<sup>12</sup>, where we have to reduce, by 2050, by 50% of CO2 emissions. This includes an alternative fuse, components in the vessel to filter the gases, etc. that must be monitored. So, the Shipping sector needs digitalization to address the challenge of decarbonization”.

For Retail sector, Tatu Kuivalahti stressed “the utilization of data in retail is in early stage and not mature; although e-commerce is speed growing, the majority of shopping is done in physical stores, and that data is not much utilized, so there is a room for improving the utilization of data in this sector. There is also a lack of understanding of the big data possibilities for decision-makers and stakeholders in the retail sector. Regarding investments, Tatu thinks that it should start with education on how to leverage big data; educating people that are going to work in the retail industry in marketing or sales areas, and of course, executives. Additionally, he thinks that there is also a lack of research projects between industry and research community, and retail companies are not investing in research. BigDataStack is a good example of how retail companies are pushing big data to be more competitive and invest in research”.



**Is BigData the real future of emerging business?**

Where does Europe sit in terms of its digital sovereignty towards BigData?

**Ray Walshe**  
Assistant Professor  
University of Dublin

Where are investments still needed in your industry sector ?

**Stathis Plitsos**  
Senior Researcher &  
Head of Development  
DeepSea Technologies

**Tatu Kuivalahti**  
Co-founder of Custobar

BIG DATA VALUE BERLIN • VIRTUAL 21-8 NOVEMBER 2020 | BDVA\_PPP | @company/big-data-value-association | @photos/13720891@N08 | @UCS@W@Z5B@4@cc@W@Y@3@G@w

<sup>12</sup> <https://www.imo.org/>

## 5. On October 21st, the EC approved the new Open Source Software Strategy 2020-2023 “Think Open”. So much software out there for industry. How important are Open Source Solutions in Big Data analytics?

Ray Walshe: “Within H2020, CE is supporting IT standard developments and OS solutions and software and standardization is part of CE strategy.

OS can contribute in the big data ecosystem and it is very important since provides a platform to develop more fast solutions. Ray encourages companies to develop

standards because when Standards is published and adopted worldwide is positive for market access; this is for all emerging technologies including Big Data”.

Céline Xu: “The importance of OS is increasing and can be seen from the users and suppliers sides. From the supplier side, i.e. Microsoft, Apple, Facebook or IBM, actively participate in OS communities and even contribute to their OS solutions. Another big firm, Mackenzie also has an OS solution. From the users' side and according to a RedHat report, 75% surveyed companies say that OS is quite important for their business and OS adoption acts as a major accelerator.

The OS phenomenon is due to different reasons: OS technological tools are cost-effective, the “try before you buy”, make the development process quite short and the time to market much shorter. Another important factor is the “sharing knowledge phenomenon” and how OS communities share knowledge, developers already use OS solutions, and finally, the last factor is the flexibility of the integration of OS solutions, including data collection, aggregation and analysis; when you use OS solutions you have more flexibility to connect to the database, different processes, platforms or clouds”.

### Is BigData the real future of emerging business?

So much software out there for industry.  
How important are Open-Source Solutions in Big Data analytics?



**Céline Xu**  
Senior Data Scientist

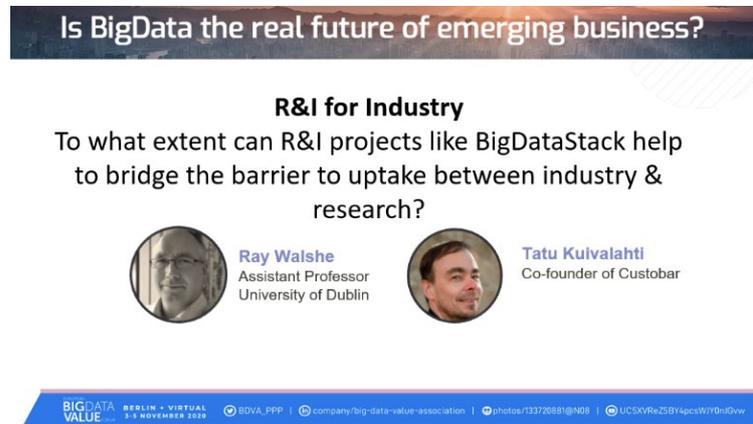


**Ray Walshe**  
Assistant Professor  
University of Dublin

**BIG DATA VALUE** BERLIN + VIRTUAL 3-6 NOVEMBER 2020 [BDVA\\_PFP](#) | [company/big-data-value-association](#) | [@rhdtor/13372088@NB8](#) | [UC3XVReZ5BYCpcW10nZvr](#)

## 6. R&I for Industry. To what extent can R&I projects like BigDataStack help to bridge the barrier to uptake between industry & research?

Ray Walshe: “To break barriers among industry and research and encourage participation in R&D projects, we need to encourage collaboration and that is what CE has done in the H2020 and previous programmes. Billions of € have been spent to get the best of the best from academia, industry, government, OS communities, etc. to work together and collaborate”.



**Is BigData the real future of emerging business?**

**R&I for Industry**

To what extent can R&I projects like BigDataStack help to bridge the barrier to uptake between industry & research?

**Ray Walshe**  
Assistant Professor  
University of Dublin

**Tatu Kuivalahti**  
Co-founder of Custobar

BIG DATA VALUE BERLIN • VIRTUAL 13 NOVEMBER 2020 | BDVA, PPP | company/big-data-value-association | photos/13372081@n08 | UCSXReZ5BYapcsWjY0aDww

From the business perspective, Tatu Kuivalahti thinks “that companies are running too fast and don’t looking around for help in research or innovation and are too busy running the business and operations. There is not enough understanding of possibilities of utilization of research and it is needed to have a more active dialogue with the research community to explain which industrial problems companies are currently facing”.

### Is Big Data the real future of emerging business?

Finally, to conclude the session, experts were requested to provide a takeaway from this discussion.



**Céline Xu**  
Senior Data Scientist

*“Big Data will help a lot of emerging businesses”*

*“The utilization of a large amount of data is, definitely, the future for most industries, and it is needed to actively collaborate with different stakeholders, specifically with the research community”*



**Tatu Kuivalahti**  
Co-founder of Custobar



**Ray Walshe**  
Assistant Professor  
University of Dublin

*“Digitalization is accelerating due to the Covid-19. Data is everywhere and we need to develop mechanisms to protect, standardize, governed properly, legislate, and certificate while protecting the rights of citizens. This is a challenge but also an opportunity to generate new business and new ways of generates revenues”*

*“The important thing is to transform the data into actionable information. Big data is open our eyes to the options we didn’t know we have”*



**Stathis Plitsos**  
Senior Researcher &  
Head of Development  
DeepSea Technologies

## Conclusions

The discussion among experts led to remarkable conclusions that helped to validate the adoption roadmap proposed for BigDataStack. The conclusions were reported within D7.5, and here, we summarize them:

- Digitalization is needed for the adoption of Big Data technologies
- It is needed Big Data tools that help to correctly define the business cases and transforming data into actionable insights for companies
- Visualization tools are very important to better know the insights of analytics in real-time
- Data management, data cleaning, data retrieving, data governance, etc. is critical for big data analytics. So, mechanisms to overcome barriers related to data must be put in place
- Involvement of the different departments of the company (business department, operation department, sales, etc.) to help to define the use case
- OS solutions and technologies are a major driver for the adoption of Big Data and facilitate companies “try before you buy” and reduce “time-to-market”.
- Highlight the importance of education and big data awareness within the company before starting the process of adoption of Big Data solutions

As mentioned in the summary above and thorough the discussion among partners, we conclude that BigDataStack solution overcomes the concerns and main barriers mentioned by the expert.

BigDataStack project delivers “a complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that addresses the emerging needs of data operations and applications. To achieve this overall goal, the project will provide a set of key offerings ranging from the underlying *Data-driven Infrastructure Management* approach, the *Data as a Service* approach and the offerings on top towards different stakeholders: the Data Toolkit towards data scientists, the Application Dimensioning Workbench towards application owners and engineers, and the Process Modelling Framework towards business analysts”<sup>13</sup>.

---

<sup>13</sup> DoA, BigDataStack Consortium, 2017

## 4. BigDataStack Exploitable Assets and MVPs

This section provides information on the final exploitable assets derived from BigDataStack project. In addition to the SW components and Use Cases developed within BigDataStack, eight MVPs were identified.

BigDataStack partners participated in a remote workshop aimed at discovering possible groups of exploitation assets which could constitute a MVP. The workshop’s outcomes are also included and explained in this section.

### 4.1. BigDataStack Exploitable Assets

BigDataStack aims “at providing a **complete infrastructure management system**, which will base the management and deployment decisions on data from current and past application and infrastructure deployments. This complete infrastructure management system is delivered as a full “stack” that facilitates the needs of operation data and application as well as facilitate it in an optimized way”.

BigDataStack key offerings / assets are depicted in the next figure.

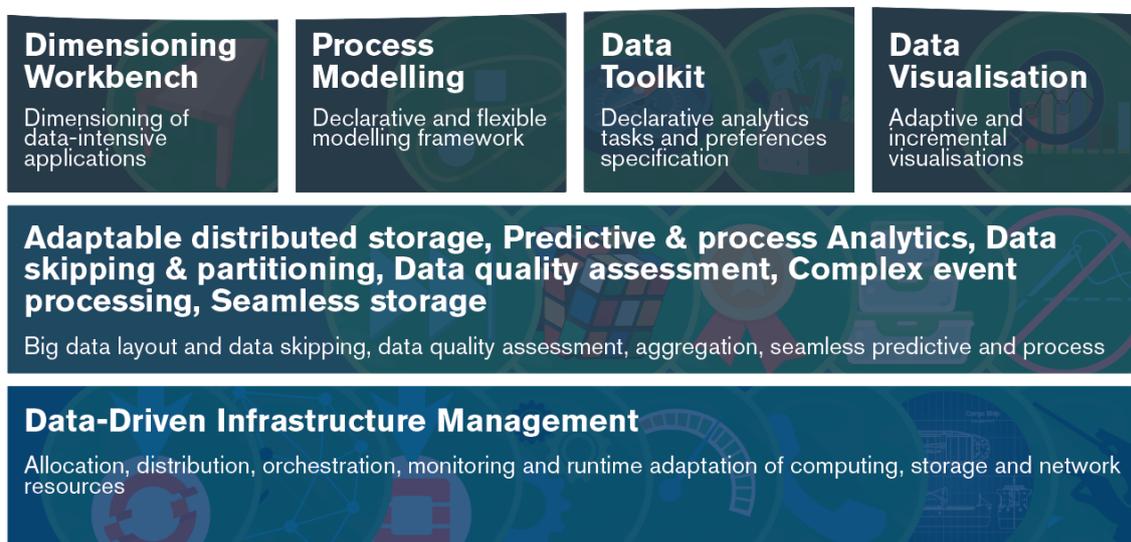


Figure 1. BigDataStack Key Offering<sup>14</sup>

To develop the functionalities of the BigDataStack offering, a series of SW components have been developed within the project which represents the BigDataStack exploitable assets.

<sup>14</sup> D2.4 Conceptual Model and Reference Architecture

D7.2<sup>15</sup> submitted in M18 included the initial list BigDataStack exploitable assets and information about them. Here, we present the final list of 21 exploitable assets and updated information on OS license, TRL, ownership and the link to public repositories.

---

<sup>15</sup> D7.2 Exploitation Plan and Business Potential v1

	Exploitable Assets	OS License	TRL	Owner	Link to publicly accessible code
<b>Data-driven Infrastructure management</b>	Information Driven Networking	Apache License 2.0	TRL5	SILO	<a href="http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istioyaml/">http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istioyaml/</a> , <a href="http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istioproxy/">http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istioproxy</a> , <a href="http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istiopod/">http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istiopod</a> , <a href="http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istiopodconsumer">http://bigdatastack-tasks.ds.unipi.gr/ppetrouubi/istiopodconsumer</a>
	Tripe Monitoring Engine	Apache License 2.0	TRL5	UPRC	<a href="https://github.com/jdtotow/tme">https://github.com/jdtotow/tme</a>
	Realization Engine	Apache License 2.0	TRL5	GLA	<a href="https://github.com/terrierteam/realizationengine">https://github.com/terrierteam/realizationengine</a>
	Beta-RecSys Recommendations Development Framework	MIT	TRL6	GLA	<a href="https://github.com/beta-team/beta-recsys">https://github.com/beta-team/beta-recsys</a>
	Dynamic Orchestrator	BSD 3-clause "New" or "Revised" License	TRL7	NEC	<a href="http://bigdatastack-tasks.ds.unipi.gr/mauricio.fadel/dynamic-orchestrator">http://bigdatastack-tasks.ds.unipi.gr/mauricio.fadel/dynamic-orchestrator</a>
	QoS Evaluator	Proprietary	TRL8	ATOS	If necessary, ATOS will provide to BigDataStack partners with a OS licence only for research purposes
	Kuryr integration into OpenShift and Cluster Network Operator	Apache License 2.0	TRL9	RHT	<a href="https://github.com/openshift/cluster-network-operator">https://github.com/openshift/cluster-network-operator</a> , <a href="https://github.com/openshift/installer">https://github.com/openshift/installer</a>
	Network Policy Support at Kuryr	Apache License 2.0	TRL9	RHT	<a href="https://opendev.org/openstack/kuryr-kubernetes">https://opendev.org/openstack/kuryr-kubernetes</a>
	NVME-mdev Kernel driver	Apache License 2.0	TRL7	RHT	This did not make it upstream, upstream submission can be found here: <a href="https://lore.kernel.org/patchwork/patch/1052204/">https://lore.kernel.org/patchwork/patch/1052204/</a>
	ovn loadbalancer integration	Apache License 2.0	TRL9	RHT	This involves different code in different repos: <a href="https://opendev.org/openstack/kuryr-kubernetes">https://opendev.org/openstack/kuryr-kubernetes</a> , <a href="https://opendev.org/openstack/octavia">https://opendev.org/openstack/octavia</a> , <a href="https://opendev.org/openstack/neutron">https://opendev.org/openstack/neutron</a> , <a href="https://github.com/openshift/cluster-network-operator">https://github.com/openshift/cluster-network-operator</a>

	Infrastructure API	Apache License 2.0	TRL9	RHT	<u>This involves different code in different repos:</u> <a href="https://github.com/openshift/cluster-network-operator">https://github.com/openshift/cluster-network-operator</a> , <a href="https://github.com/openshift/installer">https://github.com/openshift/installer</a> , <a href="https://github.com/kubernetes-sigs/cluster-api-provider-openstack">https://github.com/kubernetes-sigs/cluster-api-provider-openstack</a> , <a href="https://github.com/openshift/machine-config-operator">https://github.com/openshift/machine-config-operator</a> , <a href="https://github.com/openshift/machine-api-operator">https://github.com/openshift/machine-api-operator</a>
Data as a Service	Complex Event Processing (CEP)	Proprietary	TRL3	UPM	
	Seamless	Apache License 2.0	TRL5	LXS / IBM	<a href="http://bigdatastack-tasks.ds.unipi.gr/pavlos_LXS/lx-bigdatastack/-/tree/master/seamless-query-federator">http://bigdatastack-tasks.ds.unipi.gr/pavlos_LXS/lx-bigdatastack/-/tree/master/seamless-query-federator</a>
	Adaptable Distributed Storage	Proprietary	TRL5	LXS	
	Data Quality Assessment	Apache license 2.0	TRL5	UPRC	<a href="https://pypi.org/project/forma/">https://pypi.org/project/forma/</a>
	SQL Data Skipping	Proprietary -> OS	TRL9	IBM	
Dimensioning Workbench	Application Dimensioning Workbench	Apache License 2.0	TRL5	UPRC	<a href="http://bigdatastack-tasks.ds.unipi.gr/gkousiou/adw">http://bigdatastack-tasks.ds.unipi.gr/gkousiou/adw</a>
Process modelling and optimisation framework	Process Modelling Framework	MIT	TRL5	ATC	<a href="http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-process-modeler">http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-process-modeler</a>
Data Visualisation	Data Visualisation	Apache License 2.0	TRL5	ATC	<a href="http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-ui">http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-ui</a>
Data Toolkit	Data Toolkit	MIT	TRL5	UBI	<a href="http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-ui">http://bigdatastack-tasks.ds.unipi.gr/Sidiropoulos/bigds-ui</a>
	Process Mapping	Apache License 2.0	TRL4	UPRC	<a href="http://bigdatastack-tasks.ds.unipi.gr/gpoul/process_mapping">http://bigdatastack-tasks.ds.unipi.gr/gpoul/process_mapping</a>

Table 1. BigDataStack Exploitable Assets analysis

In addition to the SW components developed within BigDataStack project, three use cases have been implemented through the project lifetime, which represents exploitable assets in themselves.

Next, we present the BigDataStack exploitable use cases<sup>16</sup>.

Connected Consumer Use Case (Retail Sector)	Real-time Ship Management Use Case (Transport Sector)	Smart Insurance Use Case (Insurance Sector)
<p>Connected Consumer Use Case provides a recommender system for grocery market.</p> <p>It leverages the BigDataStack capabilities to create a collaborative-filtering recommender system that produces recommendations of new products to the customers</p>	<p>Real-time Ship management Use Case provides information on preventive maintenance, scheduling of orders, and visualization of current status for the decision-making process of shipping companies.</p>	<p>Smart Insurance Use Case exploitable knowledge includes the definition of the model and the development of two Machine Learning-based services: Product Recommender and Churn Predictor.</p> <p>It allows Insurance companies to improve the management of their customers' portfolio, by providing personalized services recommendation</p>

Table 2. BigDataStack exploitable Use Cases

## 4.2. BigDataStack MVPs

The term MVP is used in different ways. It can be used to describe “a methodology that achieves the maximum amount of learning from the least amount of resourced effort”<sup>17</sup>. In product development, MVP is the product with only core features that make possible the highest return on investment versus risk.

Within BigDataStack project, we have assumed that an MVP “*is the minimum product with enough features to engage early-adopter customers and validate a product idea*”<sup>18</sup>, and partners have work to identify the groups of exploitable assets that can work together and have a joint exploitation potential. BigDataStack partners participated in a remote workshop to discuss and explore MVP possibilities. To that end, the exploitation team used the MIRO<sup>19</sup> tool for remote activities with distributed teams.

The outcomes of the workshop are depicted in the next tables.

<sup>16</sup> This information has been extracted from *D6.2 Use Case description and implementation Y3*

<sup>17</sup> <https://www.sciencedirect.com/topics/computer-science/minimum-viable-product>

<sup>18</sup> <https://www.productplan.com/glossary/minimum-viable-product/>

<sup>19</sup> <https://miro.com/>

**Activity 1. Which BigDataStack offering is an MVP?**

This activity aimed at identifying which modules from the BigDataStack offering could be considered as an MVP, and discuss its features, early adopters, etc.

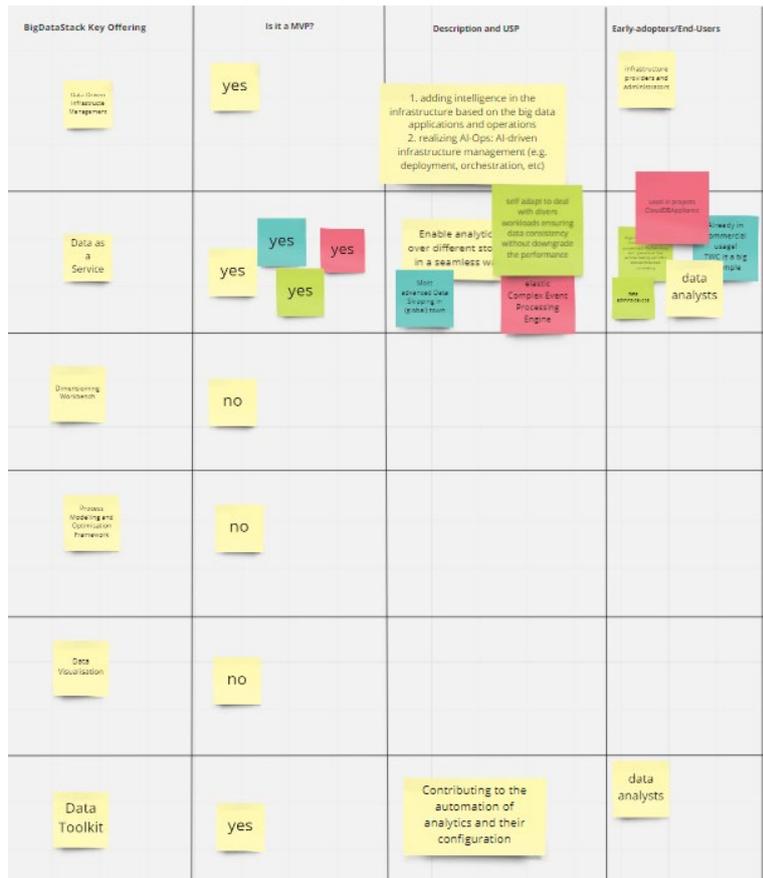


Figure 2. Mural from Activity 1

BigDataStack Offering	Is it an MVP?	Description and USP	Early adopters/end-users
<b>Data-driven Infrastructure Management</b>	<b>YES</b>	Adding intelligence in the infrastructure based on the big data applications and operations Realizing AI-Ops: AI-driven infrastructure management (e.g. deployment, orchestration, etc)	Infrastructure providers and administrators
<b>Data as a Service</b>	<b>YES</b>	Self-adapt to deal with divers workloads ensuring data	Data Analysts Data Admins/DevOps

		<p>consistency without downgrade the performance</p> <p>Enable analytics over different stores in a seamless way</p> <p>Most advanced Data Skipping in (global) town</p> <p>Dynamic elastic Complex Event Processing Engine</p>	<p>Organizations that need to combine analytical processing in BigData along with operational data without dealing with ETLs and sacrifice data consistency</p> <p>Currently used in projects</p>
<b>Data Visualization</b>	<b>YES</b>	Contributing to the automation of analytics and their configuration	Data Analysts

Table 3. MVPs from BigDataStack Offering

### Activity 2. MVPs from BigDataStack Exploitable Assets

This activity aimed at identifying which group of exploitable assets from the BigDataStack project could be considered as an MVP, and discuss its features, early adopters, etc.

BigDataStack MVP	Description and USP	Features	Early adopters/End-Users	Exploitation Paths
<p>Seamless</p> <p>Sql Data Skipping</p>	<p>Enables query execution on databases that have been split among an operational and analytical databases, ensuring data consistency.</p> <p>The reason of this improvement is that, thanks to the skipping mechanism, the analytical query does not access the operational data.</p>		<p>Organizations that need to perform analytics over operational datasets that contain BigData</p>	<p>AI to test and further develop on the AWS Ray2020</p>
<p>Realization Engine</p> <p>Tripe Monitoring Engine</p> <p>Systems Orchestration</p>	<p>1. AI Ops: AI driven management of the infrastructure / operations.</p> <p>2. Runtime adaptable managers of resources and services</p>	<p>Applicable to different scenarios / applications</p>	<p>Infrastructure providers and administrators</p> <p>Developers of applications (Big Data, Edge)</p>	<p>Further develop and modify for its use on FogFlow</p>
<p>Tripe Monitoring Engine</p> <p>QoS Evaluator</p>	<p>Enhanced levels of dynamism (across all layers = resources adaptation, storage adaptation, and) by utilizing real world dynamics</p>	<p>Support for different types of metrics and the respective QoS parameters. Exemplified in terms of those metrics and parameters</p>	<p>Infrastructure providers and administrators</p>	
<p>Network Policy Support at Kurje</p> <p>Kurje integration into OpenShift</p> <p>Kurje integration into OpenShift</p> <p>Information Driven Networking</p>	<p>Significant performance increases (7x)</p>		<p>Infrastructure providers and administrators</p>	

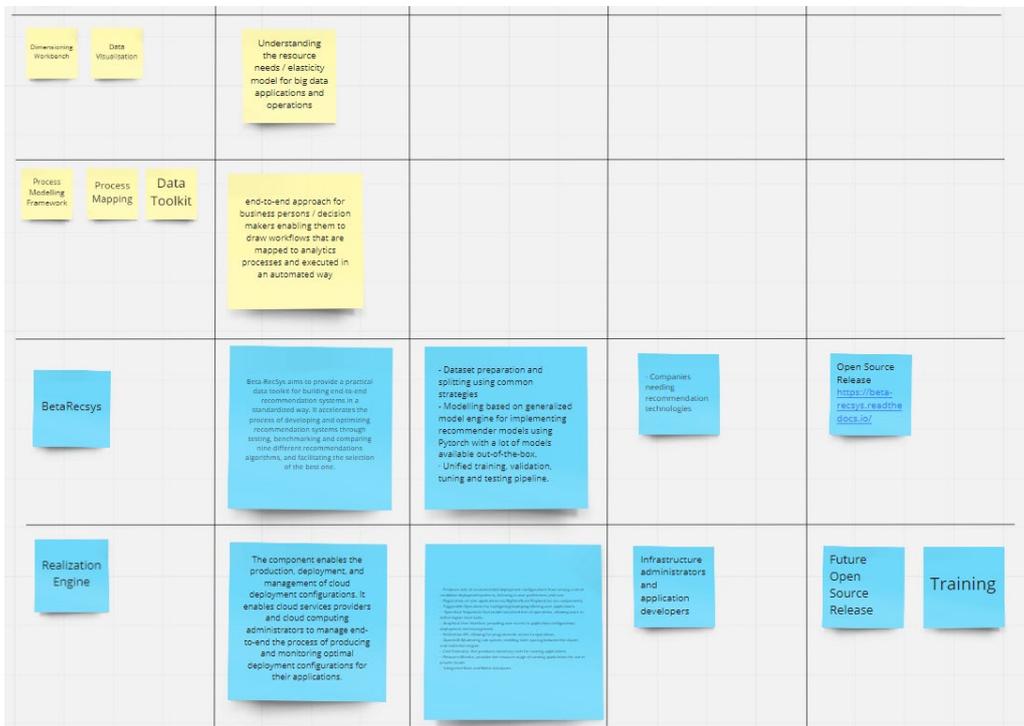


Figure 3. Murals from Activity 2

BigDataStack MVPs	Description and USP	Features	Early-adopters/End-users
<p><b>Seamless SQL Data Skipping</b></p>	<p>Enables query execution on datasets that have been split among an operational and analytical datastore, ensuring data consistency.</p> <p>The support of JOIN operation on the object store via the data skipping technology allows for the efficient execution of all standard SQL statements</p>	<p>Organizations that need to perform analytics over operational datasets that contain Big Data</p>	<p>This group of components will be used and further developed under the H2020 PolicyCloud project</p>
<p><b>Realization Engine</b></p> <p><b>Deployment Recommendation Service</b></p> <p><b>Triple Monitoring Engine</b></p> <p><b>Dynamic Orchestrator</b></p>	<p>AI-Ops: AI-driven management of the infrastructure/operations</p> <p>Runtime adaptable management of resources and services</p>	<p>Applicable to different scenarios/applications</p>	<p>Infrastructure providers and administrators</p> <p>Developers of applications (Big Data, Edge)</p>

			Further develop and modify for its use on FogFlow <sup>20</sup>
<b>Triple Monitoring Engine</b> <b>QoS Evaluator</b>	Enhanced levels of dynamicity (across all layers = resources adaptation, storage adaptation, etc) by utilizing real-world dynamics	Support for different types of metrics and the respective QoS parameters.  Extendable in terms of these metrics and parameters	Infrastructure providers and administrators

Table 4. Discussion on MVPs from BigDataStack exploitable assets (I)

BigDataStack MVPs	Description and USP	Features	Early-adopters/ End-users
<b>Network Policy Support at Kuryr</b> <b>Kuryr integration into OpenShift and into Cluster Network Operator</b>	Significant performance increases (7x)	Information-driven networking for big data applications and operations deployed on infrastructures	Infrastructure providers and administrators
<b>Dimensioning Workbench</b> <b>Data Visualization</b>	Understanding the resource needs/elasticity model for big data applications and operations	Load injection (in an automated way) and elasticity model creation for big data apps	Big data analysts, application owners
<b>Process Modelling</b> <b>Process Mapping</b> <b>Data Toolkit</b>	End-to-end approach for businesspersons/decision makers enabling them to draw workflows that are mapped to analytics processes and executed in an automated way	Generation of business workflows including analytics  Automated mapping of processes to analytics	Business analysts, decision makers
<b>BetaRecsys</b>	Beta-RecSys aims to provide a practical data toolkit for building end-to-	Dataset preparation and splitting using common strategies	Companies needing recommendation technologies

<sup>20</sup> <https://www.nec.com/en/global/techrep/journal/g18/n01/180110.html>

	<p>end recommendation systems in a standardized way</p>	<p>Modelling based on generalized model engine for implementing recommender models using Pytorch with a lot of models available out-of-the-box,</p> <p>Unified training, validation, tuning and testing pipeline.</p>	
<p><b>Realization Engine</b></p>	<p>The component enables the production, deployment, and management of cloud deployment configurations. It enables cloud services providers and cloud computing administrators to manage end-to-end the process of producing and monitoring optimal deployment configurations for their applications</p>	<p>Produces sets of recommended deployment configurations from among a set of candidate deployment patterns, factoring in user preferences and cost.</p> <p>Registration of user applications via BigDataStack Playbook (or via components).</p> <p>Triggerable Operations for configuring/deploying/altering user applications.</p> <p>Operation Sequences that enable serialized lists of operations, allowing users to define higher-level tasks.</p> <p>Graphical User Interface, providing user access to application configuration, deployment and management.</p> <p>Realization API, allowing for programmatic access to operations.</p> <p>Openshift Monitoring subsystem, enabling state-syncing between the cluster and realization engine.</p> <p>Cost Estimator, that produces monetary costs for running applications.</p> <p>Resource Monitor provides live resource usage of</p>	<p>Infrastructure administrators and application developers</p>

		<p>running applications for use in private clouds.</p> <p>Integrated State and Metric databases.</p>	
--	--	--	--

Table 5. Discussion on MVPs from BigDataStack exploitable assets (II)

The eight possible MVPs from BigDataStack exploitable assets is outlined in the next table.

<p>Seamless SQL Data Skipping</p>	<p>Realization Engine</p> <p>Deployment Recommendation Service</p> <p>Triple Monitoring Engine</p> <p>Dynamic Orchestrator</p>	<p>Triple Monitoring Engine</p> <p>QoS Evaluator</p>	<p>Network Policy Support at Kuryr</p> <p>Kuryr integration into OpenShift and Cluster Network Operator</p> <p>Kuryr support for own distributed load balancing</p> <p>Information Driven Networking</p>
<p>Dimensioning Workbench</p> <p>Data Visualization</p>	<p>Process Modelling</p> <p>Process Mapping</p> <p>Data Toolkit</p>	<p>BetaRecsys</p>	<p>Realization Engine</p>

Table 6. BigDataStack MVPs

## 4.3. Updated information for BigDataStack Exploitable Assets

All the exploitable assets developed within BigDataStack were described in detail in the previous exploitation deliverable D7.2. In this section, all partners have provided updated information more focused on the exploitation aspects of the components.

### 4.3.1. SQL Data Skipping

#### Brief Description

**SQL Data Skipping** is part of the “Data as a Service” BigDataStack offering. It is one of the major success of BigDataStack project since this technology has been currently included within IBM portfolio, specifically in four commercialized products.

#### Main Features

According to today’s best practices, cloud compute and storage services should be deployed and managed independently. This means that potentially huge datasets need to be shipped from the storage service to the compute service to analyse the data. This is problematic even when they are connected by a fast network and highly exacerbated when connected across the WAN e.g. in hybrid cloud scenarios. To address this, minimizing the amount of data sent across the network is critical to achieve good performance and low cost. Data skipping is a technique which achieves this for SQL analytics on structured data. (copied from deliverable D4.3<sup>21</sup>)

#### Areas of Application

Big Data Analytics

#### Market Trends and Opportunities

All products/services/projects with which data is retrieved from Object Storage through SQL queries are part of the potential market.

Within the “Innovation Radar project”<sup>22</sup>, the EC identified the Market Creation Potential of the innovation as addressing the needs of existing markets. Additionally, IBM is on its way to Open Source the Data Skipping component, which will help to gain global visibility and foster adoption beyond IBM.

#### Customer Benefits

1) Better performance 2) lower price due to reduced data read.

<sup>21</sup> D4.2 WP4 Scientific Report and Prototype description

<sup>22</sup> <https://www.innoradar.eu/>

## Technological Novelty

Data skipping techniques are novel in the industry, in addition, envisioned synergy with data layout technology may lead to further advances in big data analytics. IBM solution is the most complete and advanced, and hence it has been selected by the EU “innovation radar”<sup>23</sup> as a key innovation. SQL Data Skipping has also been identified as “Tech Ready”.

### 4.3.2. Dynamic Orchestration

#### Brief Description

The Dynamic Orchestrator (DO) triggers the redeployment of BigDataStack applications during runtime to ensure they comply with their Service Level Objectives (SLOs.) The DO utilizes an in-house Reinforcement Learning-based logic, which combines domain knowledge with machine learning for bootstrapping performance and obtaining a more robust and stable behaviour.

#### Main Features

The DO, alongside the Triple Monitoring Engine, monitors and triggers redeployment changes in the application during runtime, ensuring the compliance with SLOs. This improves the performance of applications effortlessly and can help to reduce costs related to overprovisioning of resources in applications with variable load. The main features of the DO are:

**Flexible:** it is possible to define different SLOs that the DO will monitor, and it is also possible to define different redeployment actions that the DO will trigger when a change is necessary. Also, thanks to the use of RL, the DO will be able to manage newly defined actions and learn from its experience when these actions should be applied.

**Stable and robust:** the DO implements a logic that combines machine learning and domain knowledge, obtaining the best of both worlds: on one side, we ensure the agents’ performance is reasonable at startup using heuristics, but the agent can improve upon these heuristics as it gathers more experience, thanks to its RL-based logic.

**Rapid-reaction:** redeployment changes are triggered in real-time when an SLO violation is detected, reducing the time in which any SLO is violated.

#### Areas of Application

The DO can be used by all applications’ developers in BigDataStack, and other platforms in which the DO is integrated such as FogFlow. In particular, the applications that will benefit the most from the DO are applications with varying load: for applications that need to serve

<sup>23</sup> <https://www.innoradar.eu/innovation/35322>

a varying quantity of requests and/or users, the DO provides an important mechanism to enable applications to comply with their SLOs at all times, without incurring into unnecessary costs such as overprovisioning.

### Market Trends and Opportunities

Cloud applications that serve mobile and web applications with a highly variable load are the norm nowadays, and its use will only increase. Additionally, this trend is also seen nowadays in machine learning applications run on the Cloud to serve users' requests, and in particular, these applications are highly affected by a changing load of requests and users. What is more, ML applications many times cannot be replicated as easily as other applications, so the automatic mechanisms incorporated in commercial platforms such as Amazon Web Services and Azure, many times fall short of expectations and require high effort from developers. The DO is an essential tool for managing these applications and thanks to its flexibility, can implement dynamic adaptation with little effort from developers.

DO is classified into the "Data Processing Architectures" within the Big Data Reference categories.

### Customer Benefits

Developers benefit from optimizing the performance of their applications effortlessly, reducing costs and avoiding SLOs violation, which can cause economical penalties and losing users.

### Technological Novelty

The DO implements a novel approach called Tutor4RL<sup>24</sup> that combines domain knowledge with Reinforcement Learning. This approach has been designed and developed in BigDataStack, implementing into the DO. With Tutor4RL, we address two common problems in the use of RL for real-world use cases:

1. Initial poor performance of the agent: in Tutor4RL, we implement domain knowledge in the form of programmable functions that will guide the behaviour of the agent in its initial steps, providing a decent performance while the agent gathers experience to learn from the application and its context, in BigDataStack, these programmable functions are already implemented following heuristics that represent "rules of thumb" for managing the deployment of applications on the platform.
2. Robustness of the agent's behaviour: because we are dealing with real applications and errors should be reduced to the minimum, we use programmable functions for constraining the actions of the agent during its whole lifetime, this ensures the agent will not take actions

---

<sup>24</sup> <https://www.iofu.org/publication/argerich-2020-tutor-4-rl/>

that will surely hurt the performance of the application during runtime, reducing errors caused by the exploration of the agent.

### 4.3.3. Network Policy support at Kuryr

#### Brief Description

Support for Kubernetes<sup>25</sup> Network Policies API at Kuryr

#### Main Features

It adds support for defining and enforcing Kubernetes Network Policies at Kuryr. The Kuryr support creates the needed OpenStack<sup>26</sup> Security Groups and Security Group Rules to ensure the defined Network Policy is enforced

#### Areas of Application

Deploying containerized applications on OpenShift<sup>27</sup>/Kubernetes which require specific fine-grain network access policies. The support at Kuryr is needed when OpenShift/Kubernetes is deployed on top of OpenStack cloud (which avoids double encapsulation and hence greatly improves network bandwidth)

#### Market Trends and Opportunities

Make Kuryr project a more solid and appealing Container Network Interface (CNI) for OpenShift/Kubernetes deployment

#### Customer Benefits

Possibility of using Network Policies while at the same time getting a big boost into the network performance (up to 9x).

#### Technological Novelty

Use of OpenStack Security Groups to back up Kubernetes Network Policies

---

<sup>25</sup> <https://kubernetes.io/>

<sup>26</sup> [www.openstack.org](http://www.openstack.org)

<sup>27</sup> [www.openshift.com](http://www.openshift.com)

#### 4.3.4. Kuryr Cluster Network Operator

##### Brief Description

Integrate Kuryr at the OpenShift Cluster Network Operator

##### Main Features

Integration of Kuryr at Cluster Network Operator. This ensures we follow the Operators pattern and deploying OpenShift cluster with Kuryr as CNI (Container Networking Interface) is straightforward, as well as day-2 operators (updates, upgrades, fault-recovery, ...)

##### Areas of Application

Deploying OpenShift on top of OpenStack with Kuryr CNI. This allows to easily deploy and manage OpenShift cluster with Kuryr so that its improvements due to avoiding double encapsulation are easy to obtain.

##### Market Trends and Opportunities

Make Kuryr project a more solid and appealing as a CNI for OpenShift/Kubernetes deployments

##### Customer Benefits

Possibility of easily deploying/managing OpenShift clusters with Kuryr as SDN/CNI and therefore get the big boost in the network performance (up to 9x).

##### Technological Novelty

Adapt Kuryr deployment model to the Kubernetes Operator model

#### 4.3.5. NVMe-mdev Kernel driver

##### Brief Description

NVMe-mdev kernel driver to have fast storage virtualization.

##### Main Features

This is a Kernel module which implements a virtual NVMe device which VMs can use as a drop-in replacement for PCI assignment. It allows sharing of NVMe devices between server VMs, with simple configuration and roughly the same performance as SPDK.

##### Areas of Application

Virtual Machines that require access to virtualized storage, such as the OpenShift on OpenStack VM use case.

## Market Trends and Opportunities

If merged on the Linux Kernel it will be available on every Linux distribution with a recent kernel version. However unclear if it will be accepted upstream in its current form.

## Customer Benefits

Speed up when accessing virtualized storage without complex configuration.

## Technological Novelty

New kernel driver implementing NVMe devices.

### *4.3.6. Integration of ovn distributed load-balancer at Kuryr for E/W traffic*

#### Brief Description

Integration of ovn load-balancer through its octavia driver into Kuryr for E/W traffic

#### Main Features

ovn distributed load balancer integration at kuryr for E/W traffic. Kubernetes uses Kube-proxy to handle the services to pod traffic. This is handled through IPTables. However, Kuryr uses Octavia to create OpenStack load-balancers to back it up the services. With the default Octavia Amphora driver, there was a need for creating a load-balancer Virtual Machines (VM) (amphora) per each OpenShift/Kubernetes service. This has implications related to resource consumption (high number of VMs), control plane actions (slow, need time to create the VM), fault-tolerance (need to recreate the VM if it dies), and not performance efficient (traffic need to go through that VM which besides the extra hop in the network, could lead to network bottlenecks).

By adding support for fully ovn base load balancers at Kuryr, we avoid all those problems as 1) there is no need to create VMs per service; 2) oad balancing is fully based on ovs flows, which means it scales better than IPTables and it is distributed, thus avoiding the fault tolerance issue; 3) it is faster as it only needs to add a few flows, not to create a new VM; 4) thus also reducing the resource consumption; 5) and it performs better as it avoids both the extra hop in the network as well as the network bottleneck issues.

#### Areas of Application

Deploying OpenShift on top of OpenStack with Kuryr CNI. All applications following Kubernetes model of exposing applications through services will benefit from this, by requiring fewer resources on the OpenStack side, as well as having a more scalable, resilience, and performance environment.

### Market Trends and Opportunities

Make Kuryr project a more solid and appealing as a CNI for OpenShift/Kubernetes deployments, with less resource consumption, better scalability and improved performance.

### Customer Benefits

When creating services, the number of resources needed for them is hugely decreased. Also, pod to service communication is improved (higher bandwidth, reduced latency).

### Technological Novelty

Add support for ovn octavia provider which is distributed by nature for E/W traffic when Kuryr CNI is used.

## 4.3.7. Infrastructure management through API

### Brief Description

Unified API for infrastructure resources to make infrastructure management easy and abstracted from the real infrastructure.

### Main Features

Extension of Kubernetes Cluster API project to have an OpenStack abstraction together with its own operator/actuator, named Cluster API Provider OpenStack. This allows automating creation/scaling actions regarding OpenShift/Kubernetes nodes when running on top of OpenStack. Therefore, helping to manage the infrastructure (OpenShift/Kubernetes nodes) through the abstracted API.

### Areas of Application

Deploying OpenShift on top of OpenStack . Allows to easily adapt the size of the OpenShift cluster to the current needs. Also, it allows having upper layers managing that, like the OpenShift autoscaler (based on CPU and Memory) or the Big Data Stack components with more advance mechanisms.

### Market Trends and Opportunities

Managing Infrastructure (as well as applications) in a declarative manner is the current trend, and this is enabling this direction for Infrastructure management on top of OpenStack.

### Customer Benefits

Easy to scale up/down the Kubernetes/OpenShift clusters through the Kubernetes API, as well as improved fault tolerance and resilience support.

## Technological Novelty

Managing Infrastructure through APIs.

### 4.3.8. QoS Evaluator

#### Brief Description

The QoS Evaluator is part of the Triple Monitoring & QoS Evaluation subsystem of the Data-Driven Infrastructure Management capability of BigDataStack. The component takes an agreement between the service provider and the application developer, describing the expected level of performance of the application as well as the platform services provided by BigDataStack. The service level is described in terms of Service Level Objectives (SLOs). An SLO specifies a constraint on Non-Functional Requirements. An SLO may also describe a business penalty to apply in case of violation. The QoS component will be responsible for managing and evaluating SLOs as well as notifying third parties when any of them is not fulfilled.

#### Main Features

The purpose of the component is to evaluate and ensure a set of QoS attributes (SLOs) at different layers of the system architecture: applications, data services (e.g., storage, processing) and infrastructure (e.g., networking, computing and storage). The QoS Evaluator receives the metrics collected by the monitoring framework and checks them individually or aggregated. The aggregation of metrics enables the specification of complex SLOs (Service Level Objectives) for elaborated conditions or constraints, such as establishing a relation between SLOs at different levels, such as establishing a relation between the response time of the application (application-level metric) and the storage capacity of the system (infrastructure level metric).

#### Areas of Application

QoS in Cloud Computing and QoS in Big Data Analytics.

#### Market Trends and Opportunities

Its open-source nature gives the QoS Evaluator the possibility to be adapted and used by developers in many different fit-for-purpose scenarios and solutions.

Advanced monitoring and alerting systems such as Intelligent Alerts are tightly tied to the proprietary Instrumental monitoring solution (vendor lock-in) and can't be distributed not used separately.

Advanced monitoring and alerting systems such as Anturis Alerting, allows for rules are quite simple, based on thresholds, in comparison to the complex logic than being specified in QoS Evaluator rules.

### Customer Benefits

Developers who want to add QoS evaluation capability to Big Data software applications and systems.

### Technological Novelty

QoS Evaluator let you create complex conditions integrating not only different variables and thresholds but context variables or conditions (including those reflecting business aspects).

QoS Evaluator provides an API to create and manage the lifecycle of QoS agreements, including the specification of their lifespan and/or conditions for activation, deactivation and deprecation.

Also, penalization information related to the violation may be attached and computed (i.e. economic costs).

## 4.3.9. Information Driven Networking

### Brief Description

The Information Driven Networking component provides a set of network engineering methods combined with software-defined networking technologies over containers and virtual machines for the enforcement of targeted policies according to the data (real-time, near real-time and offline), security requirements and application (response time, requests per second, traffic prioritization, etc.) requirements. It supports a set of mechanisms (i.e. proxies, sidecars, etc.) operating at services layer to understand the virtual hosts, URLs and other HTTP headers and at the network layer to understand the workloads in storage services, DNS and a plethora of other services that do not use HTTP.

### Main Features

The main features of the component are, as follows:

- Fully parameterized and easily configured network policies through YAML<sup>28</sup> files;
- Automatic service/pods discovery through sidecar injection;
- Deployable at any cloud environment.

### Areas of Application

Virtual Networking; Software Defined Networking; Network Policies Enforcement.

---

<sup>28</sup> <https://yaml.org/>

## Market Trends and Opportunities

Diverse network engineering capabilities; Multiple policies enablement at different OSI layers; and Diverse protocols and runtimes support.

## Customer Benefits

Automation; Narrowing services technical specificities; Ease of installation; Control app panel architecture of microservices.

## Technological Novelty

Service discovery; traffic management; service-to-service and origin-to-service security and access control; observability (including telemetry and distributed tracing); monitoring of rolling releases and resiliency.

### 4.3.10. *Process Modelling Framework*

#### Brief Description

Process Modelling is a component of the dimensioning, modelling and interaction services building block of BigDataStack, which is required to exploit the added-value services of the “underlying” BigDataStack data-driven infrastructure management and the Data as a Service offering.

#### Main Features

The Process Modelling component allows for declarative and flexible modelling of process analytics. Process Modelling component provides a wide palette of available processes that can model the uses cases. Using this palette, a Business Analyst can create a flow of processes, edit available fields depending on the type of the process, define the overall objective of the Process Modeller graph to be generated and create a high-level graph of processes. The generated graph can be saved, exported, imported or edited and is used as input for Data Toolkit Component.

#### Areas of Application

All the areas that need declarative and flexible modelling of process analytics can favor from Process Modelling Framework, as long as this is part of an integrated solution.

## Market Trends and Opportunities

Process Modelling Framework can acquire a place in the market only as part of BigDataStack solution.

### Customer Benefits

Process Modelling Framework is a very useful component for the business analyst who wants to describe her scenario using a graph, but it should be used in the framework of BigDataStack solution.

### Technological Novelty

Following the user requirements and the architecture defined at early stages of the project, ATC, after extended discussions with other members of BigDataStack consortium, developed Process modelling component. It was initially implemented by utilizing as a baseline Node-RED. Subsequently, taking the UI/UX into account migration and refactoring was performed towards the VueJS framework by using ReteJS library.

## 4.3.11. Visualization Environment

### Brief Description

Visualization Environment is a component of the dimensioning, modelling and interaction services building block of BigDataStack, which is required to exploit the added-value services of the “underlying” BigDataStack data-driven infrastructure management and the Data as a Service offering.

### Main Features

Visualization Environment presents graphs and reports of data and analytics outcome adaptively and interactively. Based on the form and the size of the data, different visualizations are dynamically presented. Performance aspects such as computing, storage and networking infrastructure data, data sources information, and data operations outcomes are visualized. It also provides a multi-view/multi-role unified and structured User Interface that either consumes or integrates various components such as Process Modeller, Data Toolkit, Process Mapping, Benchmarking, Application Dimensioning Workbench, CEP, Triple Monitoring Engine, Data Quality Assessment and Predictive Maintenance components.

### Areas of Application

Visualization Environment can be used in different cases, to “illustrate” the output of an integrated solution.

### Market Trends and Opportunities

Visualization Environment can acquire a place in the market only as part of BigDataStack solution.

## Customer Benefits

The Visualization Environment supports the visualization of data analytics for the applications deployed in BigDataStack and provides a visual application performance monitoring dashboard of the data operations and the applications. It also provides all the necessary tools for the visualization of data analytics.

## Technological Novelty

Following the user requirements and the architecture defined at early stages of the project, ATC, after extended discussions with the whole BigDataStack consortium, developed the Visualization Environment, taking into account the requirements of different components and the three use cases. The component is continuously adapted in close collaboration with the different partners, to meet the new technological challenges.

### 4.3.12. *Seamless Analytics*

#### Brief Description

The component permits to distribute datasets between a traditional ACID data base and an Object Store and to retrieve and query the datasets seamlessly through a single entry point

#### Main Features

This component permits to aggregate the LeanXcale DB and an Object Store into a single logical component. This has two main aspects: a) “historical” data is moved from the LeanXcale DB towards the Object store without user intervention b) the user may perform single SQL queries towards the aggregated data set without even knowing where the dataset is laid out (could be in LeanXcale DB only or both the LXS DB and the Object Store)

#### Areas of Application

Traditional data base client who already offload historical data to Object Store

#### Market Trends and Opportunities

A typical customer would be a customer who both needs the advantages of a traditional relational DB and also the advantages of an Object Store for the older part of its data.

#### Customer Benefits

An enterprise can reduce its operational cost by better allocating the resources needed by the storage layer. The enterprise can also avoid the operational maintenance of its data store, as this can be automatically done by this component.

Additionally, it provides Elastic scalability while ensuring data consistency and transactional

semantics

### Technological Novelty

This is a first of kind 2 tiered datastore solution which combines a relational database with an Object Store. It removes the necessity to rely on ETLs for moving data from an operational datastore to a data warehouse. This movement can be now done in real-time, without any sacrifice of performance or consistency of the data.

### 4.3.13. *Adaptable Distributed Storage*

#### Brief Description

The Adaptable Distributed Storage component consists of a novel mechanism that allows the data storage layer to be adapted to diverse workloads at run-time

#### Main Features

This technology can allow the storage layer to scaly out automatically, without downtime or decreased performance. Moreover, data allocation is achieved automatically, without the need of a DB administrator to explicitly perform any corrective action.

#### Areas of Application

Any enterprise that deals with BigData and it needs a storage layer that can handle diverse workloads

#### Market Trends and Opportunities

Enterprises that want to reduce its operational cost by better allocate the resources needed by the storage layer. The enterprise can also avoid the operational maintenance of its datastore, as this can be automatically done by this component

#### Customer Benefits

Any enterprise that deals with BigData and it needs a storage layer that can handle diverse workloads, and/or their data can get increased in a high rate while additionally, the need for data consistency and transactional assurance is crucial

### Technological Novelty

A novel mechanism has been implemented that allows the storage system of LeanXcale relational datastore to provide elastic scalable capabilities, thus, to enable the adaption to diverse workloads on the run-time

#### 4.3.14. Data Toolkit

##### Brief Description

The Data Toolkit is the component, which takes care to design an end-to-end Big Data application graph. Through the Data Toolkit is feasible to execute valid data analytics pipelines with auto-fill fields and services descriptions minimising the manual configuration required by the data scientist. It supports Spark and Spark Machine Learning (ML) library, but it may be also extended to support other machine learning frameworks (namely indicatively Tensorflow<sup>29</sup>, Scikit-learn<sup>30</sup>, PyTorch<sup>31</sup>, etc.).

##### Main Features

The main features of the component are, as follows:

- Automation;
- Online graph composition and application constraints enforcement;
- Auto-fill capabilities with services descriptions from the cloud infrastructure.

##### Areas of Application

Machine Learning and Application Setup Toolkit

##### Market Trends and Opportunities

End-to-end Big Data orchestration applications and machine learning pipelines set up and deployment.

##### Customer Benefits

Easiness; Automation; Advanced User Experience / UI.

##### Technological Novelty

N/A

<sup>29</sup> <https://www.tensorflow.org/>

<sup>30</sup> <https://scikit-learn.org>

<sup>31</sup> <https://pytorch.org/>

### 4.3.15. *Data Quality Assessment*

#### **Brief Description**

A domain-agnostic data quality assessment and an improvement framework, that can identify valid records in a relational database and establish data veracity.

#### **Main Features**

The main contribution of the data quality assessment and improvement module is that can provide a domain-agnostic framework for automating the data cleaning

#### **Areas of Application**

Data Cleaning and data analysis

#### **Market Trends and Opportunities**

Data cleaning is an area that has not gone far, in contrast to other components of the data analysis pipeline, such as the modelling phase. Thus, there are a lot of opportunities to expand and create value

#### **Customer Benefits**

Data cleaning is the first step in a data analysis pipeline. It often the task engineers spend most of their time. By automating this tedious process, we can free up time to more meaningful and creative tasks.

#### **Technological Novelty**

The main innovation in this component has to do with the keyword “domain-agnostic”. There should be no prior knowledge about the dataset and the mechanism should not work as an expert system, but more like a probabilistic system

### 4.3.16. *Application Dimensioning Workbench*

#### **Brief Description**

The component has two purposes: a) generate load against target services via easily configured and automated parameter sweep tests, thus gathering the necessary performance data b) extend experimentation to other aspects such as multitenancy overheads that can affect SaaS business models

#### **Main Features**

- Automated test setup, launch, lifecycle management and result gathering
- embedded service inclusion in the benchmark graph to detect performance overheads

- UI and REST API based test submission
- Test scalability through the usage of containerized deployment platforms and clusters
- Test abstraction through automated test configuration and modular architecture

### Areas of Application

- Load injection/Stress Testing as a Service
- Performance Engineering
- Business and Cost model validation

### Market Trends and Opportunities

Big Data technologies need to generate sufficient load sizes to demonstrate their ability to handle an increased amount of work. Such load sizes go beyond the abilities of a single node and setup and require a substantial effort in managing the load creation and test execution process.

Furthermore, different types of technologies or services may require diverse load generation processes and baseline tools. The flexibility of ADW to incorporate such baseline tools is key for extending the scope of the tool and/or adapting it per case of needed stress testing.

Software as a Service is a key business model for companies to reutilize their software developments as well as automate software management (updates, versioning etc) for their offerings towards their clients. However, in many cases of SaaS deployment options (e.g. offerings deployed in multitenant infrastructures), the performance degradation from the underlying sharing of resources may lead to undesired QoS levels and poor service performance. To this end, service owners must conduct multitenancy experiments to highlight the limits of each multitenancy scenario and thus drive deployment options and/or SaaS packages based on the trade-off between performance and cost. This also dictates the baseline cost of the service and thus the finally offered price towards their customers.

### Customer Benefits

- Accurate knowledge of the performance of their services
- Accurate knowledge of limits of multitenancy deployments
- Investigation of performance vs cost trade-offs and definition of baseline SaaS packages

### Technological Novelty

- Abstracted test design and execution
- Modular and extendible architecture
- The exploitation of large-scale clusters for load injection

Stress testing as a Service offering

### 4.3.17. *Process Mapping*

#### **Brief Description**

The Process Mapping component targets the problem of selecting the best algorithm along with a set of values for the algorithm's input parameters, from a set of candidate algorithms, given a specific data analysis task, in an automatic way. Its role is to automatically map a step of a process to a specific algorithmic instance from a given pool of algorithms, thereby achieving so-called "process mapping".

#### **Main Features**

- Data pre-processing
- Feature extraction from datasets
- Algorithm selection
- Hyperparameter tuning

#### **Areas of Application**

Automated Machine Learning.

#### **Market Trends and Opportunities**

Existing frameworks for AutoML (research prototypes or systems: AutoML, AutoWeka, TPOT, Google's Vizier) mainly provide solutions in the context of supervised learning, such as classification and regression but none of those frameworks provides a solution for the problem of Automated Machine Learning in an unsupervised context.

#### **Customer Benefits**

The Process Mapping delivers benefits for Data analysts, Data scientists and Business analysts. In all the above cases, the analyst/scientist can gain significant time when applying machine learning techniques to new datasets. The reason is that the Process Mapping component automatically selects the best algorithm and its parameterization, thus eliminating the need for testing various algorithms and configurations manually, until acceptable performance is observed.

#### **Technological Novelty**

Solving the problem of AutoML for unsupervised learning (clustering) is considered as the main technological novelty addressed by the Process Mapping component.

### 4.3.18. *Complex Event Processing (CEP)*

#### **Brief Description**

Complex Event Processing (CEP) engine is a parallel distributed data system for processing continuous infinite flows of data. CEP can run in a different type of hardware and adapts to the available resources and load.

#### **Main Features**

CEP runs in a distributed system made out of heterogeneous devices (servers, laptops, Raspberry Pi...). The CEP takes into consideration the resources of the device to deploy and run the queries and where data is generated to minimize network usage and latency.

#### **Areas of Application**

Any application that needs to process continuous (infinite) flows of data without storing them.

#### **Market Trends and Opportunities**

CEP engines are data processing components for analysing data on the fly that are used in any application that needs to process data as soon as it is produced. With the increasing number of connected devices (IoT) the flows of data are increasing exponentially, and distributed CEP engines will be more and more used in new applications.

#### **Customer Benefits**

The main benefits are the adaptation to the available resources and load without stopping the system.

#### **Technological Novelty**

CEP main novelty is the dynamic reconfiguration of queries according to resource consumption and load either deployed in a local area network or in wide area network without stopping the system.

### 4.3.19. *Beta-RecSys Recommendations Development Framework*

#### **Brief Description**

Beta-RecSys aims to provide a practical data toolkit for building end-to-end recommendation systems in a standardized way. It accelerates the process of developing and optimizing recommendation systems through testing, benchmarking and comparing nine different recommendations algorithms, and facilitating the selection of the best one.

#### **Main Features**

Beta-RecSys supports the following steps of a machine learning pipeline:

- Dataset preparation and splitting using common strategies
- Modelling based on a generalized model engine for implementing recommender models using Pytorch with a lot of models available out-of-the-box,
- Unified training, validation, tuning and testing pipeline.

More information is available at: <https://beta-recsys.readthedocs.io/>

### Areas of Application

- Retail (H2020 BigDataStack)
- Finance and Insurance (H2020 BigDataStack, H2020 INFINITECH)

### Market trends and Opportunities

- Growing momentum of Machine Learning and Artificial Intelligence System. BetaRecSys provides the means for developing advanced recommendation systems using various Machine Learning algorithms, including Deep Learning techniques.
- Automatic Machine Learning (AutoML) has a rising popularity. There are many frameworks that automate the benchmarking of multiple algorithms and the selection of the best one. BetaRecSys can be considered as an AutoML framework for recommendation systems.
- Most AutoML frameworks are focused on ML problems like classification and prediction. There is a potential gap in AutoML for recommendation systems, which can be filled by BetaRecSys.

### Customer Benefits

- **Automation and Cost-Effectiveness:** BetaRecSys automates and facilitates the development of recommendation systems. Therefore, it saves time, effort, and costs for its users.
- **Optimization and Efficiency:** With BetaRecSys end-users end-up with effective recommendation systems that are selected among many candidate models that are benchmarked and evaluated against their ability to solve the problem at hand.

### Technological Novelty

BetaRecSys's technological novelty lies in the standardization of the data preparation, modelling and execution of different ML models for recommendation systems, which enables their automated benchmarking and evaluation.

### 4.3.20. Realization Engine

#### Brief Description

The component enables the production, deployment, and management of cloud deployment configurations. It enables cloud services providers and cloud computing administrators to manage end-to-end the process of producing and monitoring optimal deployment configurations for their applications.

It is used in conjunction with other components of the BigDataStack platform.

#### Main Features

The Realization Engine provides the following features:

- Produces sets of recommended deployment configurations from among a set of candidate deployment patterns, factoring in user preferences and cost.
- Registration of user applications via BigDataStack Playbook (or via components).
- Triggerable Operations for configuring/deploying/altering user applications.
- Operation Sequences that enable serialized lists of operations, allowing users to define higher-level tasks.
- Graphical User Interface, providing user access to application configuration, deployment and management.
- Realization API, allowing for programmatic access to operations.
- OpenShift Monitoring sub-system, enabling state-syncing between the cluster and realization engine.
- Cost Estimator, that produces monetary costs for running applications.
- Resource Monitor provides live resource usage of running applications for use in private clouds.
- Integrated State and Metric databases.

#### Areas of Application

- Retail (H2020 BigDataStack)
- Maritime Transport (H2020 BigDataStack)
- Finance and Insurance (H2020 BigDataStack)

#### Market Trends and Opportunities

- Dynamic resource management in cloud computing and edge computing.
- Optimization of cloud computing infrastructures for BigData and AI applications.

### Customer Benefits

- **Optimization and Efficiency:** The Realization Engine facilitates the identification and deployment of optimized cloud configurations for data-centric applications.
- **Cost-Effectiveness – User Friendliness:** The Realization Engine facilitates cloud service providers and cloud infrastructure providers to manage deployment configurations in a user-friendly way.

### Technological Novelty

Cloud resources optimization taking into account different criteria such as the required resources and cloud costs.

## 5. BigDataStack Exploitation and Sustainability Strategy

### 5.1. Exploitation and Sustainability Strategy

The BigDataStack exploitation and sustainability strategy relies on the next pillars:

- Partners’ exploitation plans to fully exploit, after the project, the exploitable assets developed within BigDataStack and ensure their future sustainability
- Exploitation activities among partners to further jointly develop or exploit BigDataStack exploitable assets
- Commercial exploitation, patents and technology transfer of some BigDataStack SW Components
- Open Source approach and upstream contributions to Open Source Communities
- Joining to BDVA Marketplace
- Use of the SW Components and knowledge gained during the project for further R&D and educational activities
- BigDataStack Joint Exploitation

<b>Exploitation</b>	Partners Exploitation Plans for BigDataStack Assets	Joint Exploitation in ongoing and further R&D projects	Commercial exploitation Patents Technology Transfer
<b>Sustainability</b>	BigDataStack OS approach and upstream contribution to OS Communities	BDVA Innovation Marketplace EOSC-DIH Collaboration Agreement	Future R&D and educational activities

Figure 4. BigDataStack Exploitation and Sustainability strategy

#### 5.1.1. Partners Exploitation Plans for BigDataStack assets

BigDataStack partners elaborated their initial exploitation plans within the first period of the project, in which they included their preliminary intentions and interest in exploiting the SW components and exploitation activities.

Those exploitation plans paved the way for more elaborated exploitation plans and for partners to be aware of the exploitation opportunities that may arise.

Next section provides detailed information regarding the partners’ exploitation plans; here we present a summary of the most representative exploitation actions foreseen by partners to exploit BigDataStack exploitable assets.

	OS approach	Future R&D activities and projects	Educational activities	Commercial exploitation	Patents	Technology Transfer
	X	X		X	X	
	X	X			X	
	X	X		X		
	Proprietary	X		X		
	N/A	X		X		
	N/A	X		X		
	N/A	X				
	X	X				
	X	X				
	Proprietary	X		X	X	
	X	X				
	N/A	X				
	Proprietary	X	X			X
	X	X	X			
	X	X	X			

Table 7. Most representatives exploitation activities from Partners’ exploitation plans

### 5.1.2. Joint exploitation in ongoing and further R&D projects

During the last period of the project, BigDataStack has identified groups of SW components with the potential to work together and constitute an MVP, as described in section 3. BigDataStack partners are working together to identify new opportunities for the BigDataStack MVPs, and some partners are already working in R&D projects in which they intend to evolve the BigDataStack SW components.

The joint activities are:



IBM, LXS and UPRC are participating in the EU H2020 project **PolicyCloud**. They are working in Seamless analytics and data cleaning. PolicyCloud project “aims to harness the potential of digitisation, big data and cloud technologies to improve the modelling, creation and implementation of policy”<sup>32</sup>.



GLA, GFT, UPRC and LXS are participating in the EU H2020 project **INIFINITECH**, in which they plan to use BetaRecsys, Adaptable distributed storage and Realization Engine components. Inifinitech is “a joint effort of global leaders in ICT and finance towards lowering the barriers for BigData/IoT/AI-driven innovation, boosting regulatory compliance and stimulating additional investments”<sup>33</sup>.



RHT, ATOS and UPM will participate in the EU H2020 project **PHYSICS**, starting in January 2021, in which they are working in the resource management operators and runtime adaptation. Also, **UPRC** team members in BigDataStack, including Prof. Kousiouris, will follow-up and contribute to the activities of PHYSICS project as members of the research team of Harokopio University. The respective components have been released as open-source by UPRC and will be further researched and developed in the scope of PHYSICS project.

<sup>32</sup> <https://policycloud.eu/>

<sup>33</sup> <https://www.inifinitech-h2020.eu/>



UBI, LXS and UPRC are working together into the EU project CYBELE, in which they are reusing and evolving the Data Toolkit component. CYBELE’s aim “is to generate innovation and create value in the domains of agri-food by implementing Precision Agriculture (PA) and Precision Livestock Farming (PLF) methods”<sup>34</sup>.



ATC and IBM are working in the H2020 FogProtect project, in which User Authentication of the Adaptive Visualisations, the core component of the BigDataStack Dashboard, has been incorporated into the FogProtect End-to-End Data Protection Framework.

FogProtect aims at developing “A set of generic solutions to be implemented in multiple contexts and support many types of applications and services addressing four matters of innovation of technology”<sup>35</sup>.



The **UPRC team** is also participating as members of Harokopio University within the project H2020 **ACCORDION**. This project aims at “establishing an opportunistic approach in bringing together edge resource/infrastructures (public clouds, on-premise infrastructures, telco resources, even end-devices) in pools defined in terms of latency, that can support NextGen application requirements”<sup>36</sup>.



In addition to those R&D project, NEC, ATOS, Worldline, GLA and UPRC are currently contributing to the BDVA Book “**Leveraging Data-driven Infrastructure Management to facilitate AIOps for Big Data Applications and Operations**”.

<sup>34</sup> <https://www.cybele-project.eu/>

<sup>35</sup> <https://fogprotect.eu/>

<sup>36</sup> <https://www.accordion-project.eu/>

### 5.1.3. Commercial Exploitation

One of the main objectives of BigDataStack project is to bring solutions to the market. In that sense, BigDataStack partners have succeeded in meeting that goal and next we show the most representative examples:

BigDataStack industrial partners, such IBM, LXS, RHT, are already using project outcomes in their commercial offering, and others such as Danaos, GFT, ATOS, Worldline and Trust-IT, are exploring realistic commercial opportunities.

- **IBM** component, Data Skipping **has** reached the General Availability (GA) level for four IBM services:
  - GA-ed in [IBM Cloud Query](#)
  - GA-ed in [IBM Analytics Engine](#)
  - GA-ed in [IBM Watson Studio](#)
  - GA-ed in [IBM Cloud Pack for Data](#)
- **LXS** component, Adaptable Distributed Storage has been incorporated into LXS datastore and it will be used by two paying customers of LXS. The technology will be in production by end of Jan 2021.
- **RHT** is already making all the Kuryr and Infrastructure API components part of the RHT supported portfolio. In fact, companies are starting using them, e.g., the one covered in the project web page: <https://bigdatastack.eu/news/services-australia-tested-bigdatastack-used-openshift-openstack-kuryr-software>
- **TRUST-IT** is leveraging the knowledge gained and network built in the Big Data domain, the knowledge gained on the added value of standards and OS as well as the identification of OS Communities & events. This knowledge has contributed that TRUST-IT is currently participating in H2020 PolicyCloud and has added this project to its portfolio. TRUST-IT has extended its offering to Big Data projects.
- **WL** is currently tuning/experimenting with the Connected Consumer use case prototype in a commercial cloud environment (Google Cloud), This project is being financed by EROSKI. This is an opportunity to evolve the prototype developed during the project to use it in the real applications of EROSKI and migrate the current architecture of the EROSKI eCommerce to a containerized environment in the future
- **GFT** Italia and Spain are already talking with an initial set of customers in the Financial

Industry (e.g. Sulamerica<sup>37</sup>, HDI<sup>38</sup>, NEXI<sup>39</sup>). Also, GFT plans to expand the technological solution offering with the development of ML products at Big Data scale, integrable with third parties platforms and with a high degree of automation in different sectors.

- **Danaos** intends to use Real-time Shipping management use case inhouse only as a predictive maintenance and purchasing tool and as a tool to distributed deploy other shipping domain-related applications
- **ATOS** is exploring the possibility to include the QoS component along with others ATOS assets within the ATOS Computing Continuum offering. In addition to this, ATOS has passed successfully a PoC (Proof of Concept) jointly with SIEMENS to include que QoS component within a future Edge device.

#### 5.1.4. Patents

During the project lifetime, 3 patents have been filed related to BigDataStack outcomes:

- **IBM** filed a patent, for the novelty of its “Extensible Data Skipping” component.
- **LXS** component, Adaptable Distributed Storage is under a proprietary license, and currently, LXS is preparing the documentation to fill a patent for its component.
- **NEC** has applied for patenting the approach developed in BigDataStack, Tutor4RL, that combines RL with domain knowledge. This has been done in the application “Weakly Supervised Reinforcement Learning” (US 16/785,692) on February 2nd 2020, which is currently awaiting the response of the patents’ office.

#### 5.1.5. Technology Transfer

For Academia partners, technology transfer is the way for its components to reach the market:

- **UPM** has a large experience in technology transfer signing contracts with companies for the exploitation of research results. UPM has previously signed exploitation agreements with LXS, and now UPM is considering signing an exploitation agreement for CEP component. This component is under a proprietary license.

#### 5.1.6. BigDataStack OS approach and upstream contribution to OS Communities

At present, Open Source strategy for Big Data Analytics software and solutions is a major accelerator for the BDA adoption. Currently, organizations are either using an entire OS software or available OS tools to perform different tasks in their analytics framework.

---

<sup>37</sup> <https://portal.sulamericaseguros.com.br/>

<sup>38</sup> <https://www.hdi.global/us/en>

<sup>39</sup> <https://www.nexi.go.jp/en/>

Another reason for companies to adopt OS solutions is the lack of data scientist. OS solutions provide a fast adoption with low cost, and a “try before you buy” what is beneficial for small companies.

BigDataStack project has reinforced the OS strategy from the beginning of the project, so that:

- Most SW components are released as Open Source, just three components are proprietary (Adaptable Distributed Storage from LXS , Complex Event Processing from UPM, and QoS from ATOS)
- IBM, one of the largest industrial partner of the BigDataStack consortium, has started its path to open-sourced the component “SQL Data Skipping”
- BigDataStack is not only releasing its SW components as OS into public repositories but also some components from partners like RHT are actively upstream contributing to well established Open Source Communities. RHT model is to contribute to upstream open source communities as well as to try to foster them. This means all RHT code is not only open source, but it is firstly contributed to the communities (e.g., OpenStack, Kubernetes, OpenShift, Linux Kernel, etc.) and then made part of RHT products/portfolio.

RHT has contributed code to the next communities/repositories to provide the functionalities developed during the Big Data Stack project:

- Kuryr-Kubernetes OpenStack project
- Octavia OpenStack project
- OpenShift-installer
- Cluster Network Operator
- Cluster API
- Cluster API provider OpenStack
- Machine Config Operator
- Linux Kernel (though not accepted so far)
- Gophercloud
- Terraform
- Kubernetes (part related to network policy testing, enhancing it to cover more cases)

(Refers to D7.5<sup>40</sup> for more information on upstream contributions)

---

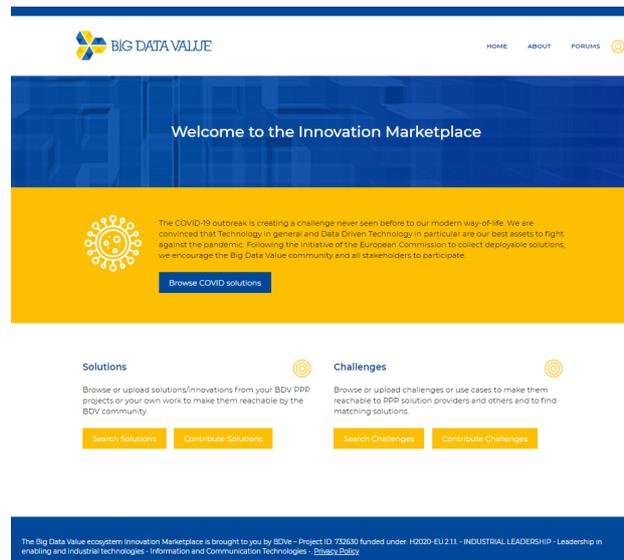
<sup>40</sup> D7.5 Dissemination, standardization and adoption roadmap

### 5.1.7. BigDataStack joint to BDVA Innovation Marketplace

BDVA is an industry-driven international not-for-profit organisation composed of industrial organizations, research community and user organizations. BDVA is the private counterpart to the EU Commission to implement the Big Data Value PPP program<sup>41</sup>.

The mission of BDVA is to develop “the Innovation Ecosystem that will enable the data and AI-driven digital transformation in Europe, and, achieving and sustaining Europe’s leadership on Big Data Value creation and Artificial Intelligence”<sup>42</sup>.

BDVA has created an Innovation Marketplace, in which EU R&D projects under the umbrella of the BDVA can join and promote their projects outcomes. BDVA innovation Marketplace aims at connecting industry with research community to increase the opportunities for exploitation and commercializing research results.



BigDataStack project has joined BDVA Innovation Marketplace to explore new exploitation and sustainability opportunities within the EU Big Data ecosystem. The link to the BigDataStack page within BDVA Marketplace is [here](#).

ATOS is an active member of BDVA and is part of the BDVA Board of Directors. ATOS will continue representing BigDataStack project in the BDVA events and initiatives to explore future exploitation opportunities. Other BigDataStack partners will also participate in BDVA events: UBI represents BigDataStack project in the BDVA events capitalizing over the Data Toolkit with improved AutoML and hybrid Big Data-HPC capabilities, moreover, GFT is currently collaborating with BDVA by leading the Task Force TF7.SG10 Finance.

### 5.1.8. EOSC-DIH Collaboration Agreement

EOSC-DIH<sup>43</sup> objective is to provide support to companies to easily access to digital technologies and services offered by the EOSC Portal<sup>44</sup>. In addition, EOSC-DIH provides a

<sup>41</sup> <http://www.bdva.eu/about>

<sup>42</sup> <https://bdva.eu/about>

<sup>43</sup> <https://eosc-dih.eu/>

<sup>44</sup> <https://www.eosc-portal.eu/>

portfolio of services aims at accelerating the adoption by SMEs and the industry of R&D project results, through training, support and visibility.

BigDataStack consortium is planning to sign a collaboration agreement with EOSC-DIH that covers the following activities:

- Dissemination of BigDataStack offering through the EOSC-DIH channels and foster the collaboration among providers, industry and BigDataStack consortium
- Explore future opportunities to conduct a pilot with an SME in the big data domain to validate and test BigDataStack results

### 5.1.9. Future R&D activities and education

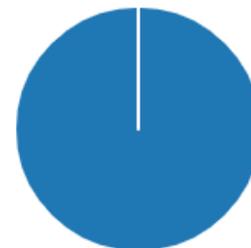
BigDataStack partners have elaborated their exploitation plans for the SW components after the project, in which they have also included the future R&D and educational activities that they are currently carrying out or they expect to perform in the next few years, and that can support the maintenance of project results.

However, partners have shown their willingness to maintain with their own means SW components in case they don't get additional R&D funding.

The exploitation team has circulated a questionnaire among partners to know the partners' intentions. The outcome of the questionnaire is depicted next.

1. Are you willing to maintain the results you have developed within BigDataStack once the project ends?

<span style="color: blue;">●</span> Yes	11
<span style="color: orange;">●</span> No	0
<span style="color: green;">●</span> Maybe	0



2. Is your organisation willing to dedicate personnel resources to the maintenance of the source code?

<span style="color: blue;">●</span> Yes	9
<span style="color: orange;">●</span> No	1
<span style="color: green;">●</span> Maybe	1



3. In case your organization doesn't get public financing to maintain and evolve the project results, is your organization willing to dedicate money to its maintenance?

- Yes, the organization is very in... 5
- No, the organization only dedi... 1
- Maybe 5



Regarding educational activities, partners from Universities are leveraging the knowledge gained during the project to offer courses and masters:

- **UPRC** is currently working on:
  - MSc on Big Data & Analytics: AutoML (process mapping) and Data cleaning
  - MSc on Advanced Information Systems: Resource management for big data applications
  - 3 MSc Theses
  - 2 cPhDs
  - Issuing of 1 MSc thesis topic (October 2020)
  - Incorporation of produced tool and experimentation aspects in 2 PR graduate and 1 post graduate courses:
    - Systems Performance
    - Urban computing
    - Cloud services
- **GLA** is currently producing an online tutorial for use by GLA undergraduate/masters and postdoctoral researchers (Students Enrolled: 94) and is also Considering commercialize the BetaRecsys component through platform training (e.g. via MOOCs (Massive open online courses)).

## 5.2. Partners’ Final Exploitation Plans

BigDataStack partners provided an initial version of their exploitation intentions in M18 that were included in D7.2. Here we present the partners’ final exploitation plans once the project finishes. They have also included the exploitation achievements accomplished during the project lifetime.

### 5.2.1. IBM Israel – Science and Technology LTD for SQL Data Skipping

BigDataStack project is for IBM the natural continuation of the IOStack project, in which they laid the foundation of data skipping and data partitioning for performant SQL queries against big data. With BigDataStack their goal is to further research and develop these techniques which are at the heart of performant data skipping for SQL queries when the dataset resides in an Object Store. The results are a natural candidate for contribution to the IBM SQL Query service. Optimized data layout is critical to the performance of SQL queries against datasets within Object Stores which itself is central in the IBM Cloud business.

The main exploitation strategy planned by IBM during the first period of the project was the integration of the developed technology into the IBM SQL Query service (and possibly other IBM services). In the first period of the project, IBM had just filled a patent for the SQL Data Skipping component developed within the project and integrated the extended data skipping technology within the IBM Cloud SQL query service.

During the second period of the project, IBM has reached its exploitation purposes as depicted in the next table.

Exploitation plan	<p>Data Skipping technology has reached the General Availability (GA) level for four IBM services:</p> <ul style="list-style-type: none"> <li>○ GA-ed in <a href="#">IBM Cloud SQL query service</a>,</li> <li>○ GA-ed in <a href="#">IBM Analytics Engine</a></li> <li>○ GA-ed in <a href="#">IBM Watson Studio</a></li> <li>○ GA-ed in <a href="#">IBM Cloud Pack for Data</a></li> </ul> <p>Additionally, IBM plans to leverage the knowledge gained during the project and the Data Skipping to participate in further R&amp;D projects. Currently, Data Skipping component is being used in the EU H2020 project Policy4Cloud.</p>
Potential risk, barriers or limitations to the	<p>So far, the technology has gained solid ground in IBM products/services. BigDataStack has allowed IBM to position as the front runner for Data Skipping technology. Currently, <b><u>IBM is on its way to Open Source the Data Skipping component, in order to gain global</u></b></p>

exploitation and sustainability	<b><u>visibility and impact beyond IBM and ensure the sustainability of the component.</u></b>
---------------------------------	--

### 5.2.2. NEC Europe LTD

NEC has developed the Dynamic Orchestrator (DO); the component in charge of triggering the redeployment of BigDataStack applications during runtime to ensure they comply with their Service Level Objectives (SLOs.) The DO utilizes an in-house Reinforcement Learning-based logic, which combines domain knowledge with machine learning for bootstrapping performance and obtaining a more robust and stable behaviour.

In the second period of the project, NEC has developed a method called Tutor4RL, which leverages domain knowledge to improve the performance and robustness of Reinforcement Learning. The use of this method addresses two common problems in RL:

- The initial poor performance of the agent: we implement domain knowledge in the form of programmable functions that will guide the behaviour of the agent in its initial steps, providing a decent performance while the agent gathers experience to learn from the application and its context, in BigDataStack, these programmable functions are already implemented following heuristics that represent “rules of thumb” for managing the deployment of applications on the platform.
- Robustness of the agent’s behaviour: because we are dealing with real applications and errors should be reduced to the minimum, we use programmable functions for constraining the actions of the agent during its whole lifetime, this ensures the agent will not take actions that will surely hurt the performance of the application during runtime, reducing errors caused by the exploration of the agent.

NEC has applied for patenting the approach developed in BigDataStack, Tutor4RL, that combines RL with domain knowledge. This has been done in the application “Weakly Supervised Reinforcement Learning” (US 16/785,692) on February 2nd 2020, which is currently awaiting the response of the patents’ office. The software for the DO has been open-sourced with the license BSD 3-Clause.

The exploitation plan envisioned by NEC is outlined in the next table.

Exploitation plan	NEC plans to include the Dynamic Orchestrator component as part of FogFlow. FogFlow is a standard-based IoT fog computing framework that supports serverless computing and edge computing with advanced programming models <sup>45</sup> .
-------------------	--

<sup>45</sup> <https://github.com/smartfog/fogflow>

	<p>The integration of the DO into FogFlow has been analysed and its currently in progress. The design of the DO has been made for BigDataStack and depends on the Triple Monitoring Engine and the Quality of Service components, because of this, its implementation in FogFlow will take longer than expected.</p> <p>However, the knowledge acquired during the development of the DO, has been used to improve FogFlow’s management of serverless FogFunctions, as well as the use of ML models in FogFunctions.</p> <p>NEC also plans to leverage the knowledge acquired during the project for R&amp;D purpose. Currently, NEC is working alongside GLA and ATOS on the BDVA book “Leveraging Data-driven Infrastructure Management to facilitate AIOps for Big Data Applications and Operations.” In this book, NEC collaborates providing its knowledge on using machine learning for adapting deployment of Big Data applications.</p> <p>NEC plans on continuing its development of Tutor4RL, and its application in different use cases. This will be done through future R&amp;D projects of NEC, and we also foresee its use and further development through other public funding sourced projects. Currently, NEC is also exploring the opportunity of developing further Tutor4RL and applying it to other use cases in a co-joint project with UPRC and RHT.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>The DO’s design depends on the input of metrics and violations coming from other components in BigDataStack, and it is currently limited because of this. However, while implementing it for FogFlow, NEC will also develop subcomponents of the DO that can generate or consume metrics from other sources. In addition, as the DO is open-sourced, its updates will be provided by the community in a best-effort policy.</p>

### 5.2.3. RED HAT Israel LTD

As the world largest open source company, RHT has pushed BigDataStack partners to participate and develop in the open-source communities. RHT development methodology is based on working on upstream projects while later productizing them to enterprise-level products.

The exploitation plan foreseen by the RHT for its components is as follow.

<p>Exploitation plan</p>	<p>RHT exploitation model relies on contributing to upstream open source communities as well as to try to foster them. This means all RHT code is not only open source, but it is firstly contributed to the</p>
--------------------------	--

communities (e.g., OpenStack, Kubernetes, OpenShift, Linux Kernel, etc.) and then made part of RHT products/portfolio.

This engagement with the community ensures sustainability since:

- Other companies can contribute to it and use it, which increases the chances of adoption and growth of it well after the Big Data Stack project finishes
- Code gets reviewed
- Code gets tested over time, ensuring new modifications won't break it

This exploitation model is not just for the BigData Stack exploitation, but the way Red Hat works in the daily basis. So, RHT have also followed that model in the BigDataStack project and everything we have done has been (or it is being) contributed to their respective open source communities.

Following that pattern, RHT has contributed code to the next communities/repositories to provide the functionalities developed during the Big Data Stack project:

- Kuryr-Kubernetes OpenStack project
- Octavia OpenStack project
- OpenShift-installer
- Cluster Network Operator
- Cluster API
- Cluster API provider OpenStack
- Machine Config Operator
- Linux Kernel (though not accepted so far)
- Gophercloud
- Terraform
- Kubernetes (part related to network policy testing, enhancing it to cover more cases)

In addition, RHT is already making all the Kuryr and Infra as API contributions part of the Red Hat supported portfolio. In fact there are companies starting using them, e.g., the one covered in the project

	web page: <a href="https://bigdatastack.eu/news/services-australia-tested-bigdatastack-used-openshift-openstack-kuryr-software">https://bigdatastack.eu/news/services-australia-tested-bigdatastack-used-openshift-openstack-kuryr-software</a>
Potential risk, barriers or limitations to the exploitation and sustainability	Other Kubernetes CNI or SDN may appear and get more impact, thus having fewer people focus on maintaining/evolving Kuryr-kubernetes, limiting its adoption.

#### 5.2.4. ATOS Spain SA

ATOS is responsible for the QoS Evaluator that is part of the Triple Monitoring & QoS Evaluation subsystem of the Data-Driven Infrastructure Management capability of BigDataStack.

This component is distributable and deployable in isolation to the rest of the components of BigDataStack; therefore, there is a chance to be exploited separately.

However, together with monitoring systems such as Prometheus, QoS Evaluator comprises a complete and powerful monitoring and QoS evaluation solution to be commercialized together. Moreover, QoS Evaluator is key in any solution based on a data-driven dynamic adaptation of Cloud infrastructure, in particular to the decision-making component such as Dynamic Orchestrator in BigDataStack.

The exploitation plan developed by ATOS is as follow.

Exploitation plan	<p>The QoS Evaluator is part of the “Edge Orchestrator”, one of the most prominent components developed by ATOS during the last few years. The Edge Orchestrator is composed by several capabilities that have been developed within EU R&amp;D projects in the Cloud and Edge computing domains, such as mF2C<sup>46</sup>, and DITAS<sup>47</sup>, within ongoing projects, CLASS<sup>48</sup> and DECENTER<sup>49</sup>, and BigDataStack project.</p> <p>The ATOS new policy for R&amp;D assets with TRL greater than 6 is that the asset will be released as proprietary and not OS, although if necessary, ATOS will provide to BigDataStack partners with an OS licence only for research purposes. Therefore, the QoS will be proprietary.</p> <p>The Edge Orchestrator is aligned with “ATOS Vision for Future Cloud” strategy and currently, ATOS is exploring the possibility to include it into the ATOS Computing Continuum offering. ATOS Computing Continuum is the ATOS solution for intermediation and aggregation across diverse private and public clouds. Cloud Continuum is an</p>
-------------------	---

<sup>46</sup> <https://www.mf2c-project.eu/>

<sup>47</sup> <https://www.ditas-project.eu/>

<sup>48</sup> <https://class-project.eu/>

<sup>49</sup> <https://www.decenter-project.eu/>

	<p>enabler for SWARM Computing. Swarm computing combines network and cloud principles to create an on-demand, autonomic and decentralised computing and storage management layer that transparently interoperates among diverse and disperse Edge and cloud models and typologies<sup>50</sup>.</p> <p>At present, Edge Orchestrator is being analysed for:</p> <ul style="list-style-type: none"> <li>• Being a standalone product within the ATOS Swarm Computing offering</li> <li>• Being integrated as software in the ATOS Bull Sequana Edge Computing Box<sup>51</sup></li> </ul> <p>This component has been also included in a PoC jointly with SIEMENS with the aim of being included in a future Edge device. It has passed the PoC phase successfully, and now, both companies are working on designing the next phase of validation.</p> <p>Additionally, the Edge Orchestrator and its components, will be used in future R&amp;D EU projects.</p> <p>Regarding exploitation activities, ARI, the Research &amp; Innovation hub from ATOS, organizes an annual event, Digital Show, in which presents the research results from R&amp;D projects to ATOS business units, to explore synergies among research results and ATOS commercial portfolio. This year, the Edge Orchestrator has been presented internally within ATOS.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>Popular open-source and commercial monitoring and alerting solutions provide similar components. Not backed by a community of developers and/or users.</p>

### 5.2.5. Worldline Iberia SA

ATOS Worldline is the European leader in the payments and transactional services industry. Worldline delivers new-generation services, enabling its customers to offer smooth and innovative solutions to the end consumer. ATOS Worldline has an experienced Center of Excellence providing business-oriented mobile solutions to any company or public institution. Mobile Competence Center leads R&D and Innovation of Worldline Iberia, focused on mobile and digital enablers technologies.

<sup>50</sup> [https://european-iot-pilots.eu/wp-content/uploads/2019/11/10\\_20191105\\_European-Industry-Partnerships-Lighthouses-to-Thrive-in-the-New-Digital-Age\\_ATOS\\_E.Pages.pdf](https://european-iot-pilots.eu/wp-content/uploads/2019/11/10_20191105_European-Industry-Partnerships-Lighthouses-to-Thrive-in-the-New-Digital-Age_ATOS_E.Pages.pdf)

<sup>51</sup> <https://atos.net/en/solutions/bullsequana-edge>

ATOS Worldline is in charge of implementing one of the use-cases that will help to test the capabilities of BigDataStack and participates in the project along with a partner, EROSKI<sup>52</sup> who is one of the largest retail companies in Spain. EROSKI has provided the needed dataset for the implementation of Connected Consumer use-case.

The exploitation plan developed by ATOS Worldline is as follow.

Exploitation plan	<p>BigDataStack has allowed ATOS Worldline both to get a solid background on BigData technologies and strengthen its relationship with one of its main customers, EROSKI. Furthermore, the knowledge and experienced obtained during the project can help ATOS Worldline to participate and implement other similar Big Data projects for its customers.</p> <p>All of the components produced for the prototypes are expected to be exportable so that ATOS Worldline partner could use them in the implementation of a recommender system running over a commercial Big Data platform.</p> <p>The interest on this part is to reuse the code implemented for the BigDataStack prototypes of the use-case during the implementation of the recommender system over a commercial platform. Thus, avoiding re-coding software already implemented during the prototypes.</p> <p>Thanks to the usage of new technologies such as Docker/Kubernetes/Openshift/Spark, ATOS Worldline team has achieved skills that are being exploited within ecommerce projects:</p> <ul style="list-style-type: none"> <li>• ATOS Worldline is currently considering migrating to Openshift some of their existing ecommerce projects, due to the expertise gained during BigDataStack project in this technology</li> <li>• During the second half of the project, EROSKI has shown interest in deploying the prototype created for the use case in Google Cloud. As a project aside BigDataStack, ATOS Worldline is currently tuning/experimenting the prototype in a commercial cloud environment (Google Cloud), This project is being financed by EROSKI.</li> </ul> <p>This is an opportunity to evolve the prototype developed during the project to use it in the real applications of EROSKI and migrate the current architecture of the EROSKI eCommerce</p>
-------------------	--

<sup>52</sup> <https://www.eroski.es/>

	to a containerized environment in the future.
Potential risk, barriers or limitations to the exploitation and sustainability	No use of commercial cloud services during the project. The prototype created has been targeted for EROSKI features, and it is not trivial to transform the same code for other Retail company.

### 5.2.6. GFT Italia SRL

The GFT Italy serves as a strategic information technology partner, which helps companies to optimize their business processes with intelligent and innovative IT solutions and highly skilled specialists and to transform cutting-edge technological developments into future-proof business models.

GFT Italy belongs to a multinational Group whose operating division GFT Solution is among the world’s leading IT service providers in the finance sector.

GFT is interested in the exploitation of its digital financial solutions; through BigDataStack, GFT is assessing to extend and improve its solutions, making it more attractive for these domains. The group has stood for technological expertise, innovative strength and premium quality for more than 25 years; BigDataStack offered an excellent opportunity to further improve GFT’s competencies and offering in the Big Data domain.

GFT has developed the financial demonstrator, also providing the needed datasets according to an agreement with one of its customers in the Insurance market. GFT has coordinated together with all the technical partners the integration of the final project platform.

The GFT Group is a global technology partner for future digital issues –covering everything from discovering innovation to developing and implementing sustainable business models.

Within the GFT Group, GFT stands for competent consulting and reliable development, implementation and maintenance of customized IT solutions.

The company is one of the world’s leading IT solutions providers in the financial services sector.

GFT has developed solutions in partnership with many experienced players in the financial services arena such as Backbase, Cloudera, Guidewire, Hicare, IBM, Informatica, or ServiceNow.

The range of services covers the areas of Claims Management, Customer Centricity, Digital Channels and IT Efficiency. BigDataStack project offered an excellent opportunity to further improve GFT’s competencies and offering in the Big Data domain.

A major focus area is a maintenance and further development of business-critical core processes. The division has many years of experience as a strategic IT partner for major financial institutes in this field. A further key area is the development of innovative solutions for the finance sector based on cutting-edge technological advances in the fields of Big Data, Mobility, Social Media and

Cloud Computing. With the aid of its Global Delivery Model, the GFT Solutions division can reliably supply its range of solutions to the core markets of Europe and America.

The exploitation activities envisioned by GFT are as follow.

<p>Exploitation plan</p>	<p>GFT is going to take the opportunity to expand its business to the trading market through the adoption of the financial demonstrator and further R&amp;D. In general, GFT aims to enhance its solutions in Big Data domain with the project’s results and main exploitation activities have been focused on the financial market. More precisely, during year 1, an internal dissemination activity of the BigDataStack solution has been performed by exploring, evaluating, and validating the potential valuable assets from BigDataStack outcomes (tangible or intangible). Year 2 and Year 3 aim was to share the project’s assets with GFT wide portfolio of customers from insurance companies to banks, credit cards circuits, merchants and telco operators, starting from the ones with a strict cooperation ongoing (e.g. SIA, NEXI). After the end of the project, GFT foresees to integrate the services provided by the BigDataStack platform to improve its solution and to offer better customized services to the clients.</p> <p>Another major achievement during the project lifetime, was the improvement of GFT knowledge in Big Data and Cloud technologies, also coordinating the design and development of the financial use case based on Insurance. The pilot mainly consists of an advanced recommender for custom products to Insurance clients, deployed on the BigDataStack testbed and covering both the pre-deployment and run-time phases of the BigDataStack architecture, helping to testing the capabilities of the following components: process modelling, process mapping, data toolkit, hyperparameters fine tuning, data quality assessment and adaptive visualizations.</p> <p>During the last three months of the project, GFT Spain entered the project as GFT linked third party, with the aim to support the final refinement and development of the model and analytics services of the Insurance use case, as well as to give a European dimension to GFT exploitation strategy.</p> <p>Thanks to BigDataStack, GFT Spain has had the opportunity to expand, explore and experiment new technology solutions related to the application of Artificial Intelligence and Machine learning solutions at Big Data and “cloud scale”.</p>
--------------------------	---

	<p>At the end of the project the aim is to expand the technological solution offering with the development of Machine Learning products at Big Data scale, integrable with third parties platforms and with high degree of automation in different sectors: in the insurance sector, as it has been demonstrated in the BigDataStack, but with plans to expand it to financial and retail sectors, both in EU and overseas.</p> <p>From a subjective innovation point of view, thanks to BigDataStack, GFT Spain took the opportunity to wider its background on Big Data and Machine Learning technologies and increased its knowledge of European research projects structure, with the aim to enlarge its participation to future projects based on Big Data &amp; Machine Learning technologies.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>GFT doesn't envision any potential barrier in the adoption of BigDataStack solutions within its portfolio.</p>

### 5.2.7. DANAOS Shipping Company LTD

DANAOS Shipping as an organization is envisaged in optimizing the scheduling of maintenance jobs for its fleet along with the purchase cycle of the spare parts needed. BigDataStack maritime use case is heading in that direction. Elaborating on big data analysis for orchestrating maintenance requirements and spare parts supply along with a large fleet with a high utilization rate where time for the dry dock is tight and maintenance plan constitutes a challenging task stands as strong motivation for DANAOS to be a part of this promising project. DANAOS was looking forward to a positive project result for further exploitation to achieve cost savings and operational excellence boosting up his reputation in the shipping industry. DANAOS main contribution to the project was to capitalize on shipping knowledge and provide insights and operational data for the development of the marine BigDataStack case.

DANAOS has offered the shipping domain knowledge along with all means to assist the provability of BigDataStack for this domain. This includes software and personnel experience as well. DANAOS One Platform is a platform that shipping companies place their orders to a list of suppliers allowing a bilateral transparent transaction with them. Voyage Estimator is another software solution that DANOS uses to estimate the total cost of a voyage. Both of these systems can assist the ordering of a spare part of the main engine and dynamically route a vessel to the closest port where the part is available.

The exploitation plan foreseen by DANAOS is as follow.

Exploitation plan	<p>DANAOS has contributed to BigDataStack project via integrating and testing the BigDataStack platform with existing systems that are required for the desired functionality. These existing systems are “DANAOS One” platform from which orders of spare parts can be placed to a list of preferred suppliers and “Voyage Estimator” which is a complete DANAOS product that estimates the total cost and time required for a vessel going from point A to point B at sea. Note here that these two components are property of DANAOS, hence cannot be used from other shipping companies, unless differently agreed.</p> <p>The exploitable result for DANAOS is twofold: (a) a system that can accurately perform predictive maintenance over large datasets, (b) a platform that given the requirements in CPUs, RAM, Disk space etc., can deploy any application distributed. Both results are of main interest for use only from DANAOS.</p> <p>After the completeness of the project, DANAOS intends to use it inhouse only as a predictive maintenance and purchasing tool and as a tool to distributed deploy other shipping domain – related applications.</p>
Potential risk, barriers or limitations to the exploitation and sustainability	<p>DANAOS doesn’t envision any barrier to the adoption of BigDataStack solutions, since during the project, BigDataStack has been integrated within existing DANAOS platforms.</p>

### 5.2.8. SingularLogic Romania Computer Applications SRL

SILO brings in the BigDataStack project its emerging expertise in 5G technologies and mainly contributes in the abstraction of data sources in order to enable data acquisition from heterogeneous sources, while also contributing in the engineering and development of the information-driven networking mechanisms.

The exploitation plan provided by SILO is next.

Exploitation plan	<p>Opportunities for SILO are lying in the area of data driven network engineering mechanisms and DevOps for Big Data as they are targeting at software defined networking technologies over containers and virtual machines for the enforcement of focused policies according to the data (real-time, near real-time and offline) and application needs. SILO team is pursuing a research agenda which strengthens its current expertise with the potential to extend it towards (a) High-complexity</p>
-------------------	---

	<p>Transactional Software Systems in the areas of Industry, Business, Finance, Insurance, Health and Public Administration, (b) Innovative communication and interconnecting technologies, supporting interoperability between Business Applications in commerce and public sector and providing solutions and tools for distributed collaboration, content management, and integration of services, (c) Software methodologies and tools in the areas of object-oriented software, enterprise modelling, requirements specification.</p> <p>The innovation potential of the Information Driven Networking includes the automatic service discovery, the efficient traffic management, the capabilities for service-to-service and origin-to-service security and access control, the functionalities about telemetry and distributed tracing.</p> <p>The Information Driven Networking has synergies with the Kuryr Cluster Network Operator and the Triple Monitoring Engine. A potential exploitation path and the related commercial opportunities of the Information Driven Networking may include an integrated product with Kuryr and the Triple Monitoring Engine working in collaboration with the RedHat and UPRC partners to automate the networking of microservices, narrow services/pods technical specificities, ease the installation and enforcement of network policies, and control multiple policies at different OSI layers.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>Driven Networking has been developed by a small development team with limited resources availability what can be a limitation for future exploitation.</p> <p>It has not a fair potential in this market as the equivalent competitors' component (e.g. Tigera Inc., OpenStack).</p> <p>Big players in this market have already launched an equivalent component.</p>

### 5.2.9. Athens Technology Center SA

ATC is an international software company. For more than 30 years, it provides innovative solutions for the Media, Banking and Retail Sectors, Utilities and Public Sector Organizations as well as horizontal solutions focusing on Content Management, Enterprise Software, Web.

Being an innovative IT company, ATC is actively involved both in R&D and commercial activities relevant to Big Data analytics and exhibits long-standing experience in state-of-the-art machine learning and quantitative analysis technologies to deliver valuable insight concerning both semantics and network-level properties that matter to the big data media professionals. Having commercial products addressing the data needs of media companies

but also large enterprises, ATC has a strong interest in advancing the relevant technologies and activities to remain competitive in the fast-moving sector of Big Data.

The exploitation plan envisioned by ATC is as follow.

Exploitation plan	<p>ATC has developed the Process Modelling Framework and the Visualization Environment within BigDataStack project.</p> <ul style="list-style-type: none"> <li>• Process Modelling Framework and Visualization Environment are available in BigDataStack repository. Process Modelling is closely collaborating with Process Mapping and Data Toolkit, and Visualization Environment is the component that, collaborating with other components, visualizes the output of BigDataStack solution. Both components have an important role in the flow, and it is better to be exploited as part of the whole approach and not on their own, so they will be available under the license of BigDataStack solution.</li> <li>• Both components will be maintained by ATC using their own resources and use it in future R&amp;D projects and will analyze how it could be included into its own commercial portfolio.</li> </ul> <p>During the first 18 months of the project, ATC made several internal presentations in the different departments of the company, in order to increase awareness of its participation in BigDataStack solution. Some informal presentations have also been made in its sales and partners network, while it has also mentioned, presented or made relevant material available in different events in which it participated. Through these presentations, ATC informed its network about its active involvement in big data area.</p> <p>The core of the exploitation actions of ATC is focused on the business of the integration and adaptation of BigDataStack developments in order to enhance ATC's portfolio of products and services. In all the events it participates during the last 18 months of the projects, ATC presents BigDataStack solutions, along with its involvement in it.</p> <p>ATC's future plans include a further investigation for possible new features and adaptations for already available ATC commercial products based on BigDataStack components and their potential use in future R&amp;D projects. Once the potentials will be located, new presentations, meetings and participation in events will be organized.</p>
-------------------	--

Potential risk, barriers or limitations to the exploitation and sustainability	Since both components are closely collaborating with other project components, it will mandatorily follow the progress of the other components. This can be a great advantage or a big risk and limitation for both components.
--	---

### 5.2.10. *LeanXcale SL*

LeanXcale is a start-up that commercializes an ultra-scalable operational database. LeanXcale has incorporated a couple of years ago. LeanXcale has been awarded in the EIT ICT Labs Idea Challenge start-up competition in the Future Cloud topic two years ago. Last year, the company was one of the 10 finalists in the South Summit European start-ups competition in the area of B2B.

Through BigDataStack project, LeanXcale will enhance the product through R&D with elasticity and integration with IBM object store and validate with the use cases of the project to later commercialize an evolved version of the product incorporating the new features resulting from the R&D performed in the project. Elasticity is a key capability to bring LeanXcale offering to the cloud. Currently, LeanXcale is only available for on-premise deployment. The goal of this project is to add elasticity to LeanXcale and integrate with IBM object store to bring it to the cloud. IBM object store has the same API as cloud storage services such as Amazon S3 and will help to make the solution deployable in the cloud.

The exploitation plan elaborated by LXS is as follow.

Exploitation plan	<p>LXS has developed the Adaptable Distributed Storage and Seamless components within the BigDataStack project.</p> <ul style="list-style-type: none"> <li>Adaptable Distributed Storage is an integral part of the LeanXcale datastore and has been developed based on the background of the solution. Therefore, it is under LeanXcale IPR. This component has been incorporated in the core of the LeanXcale datastore distribution and will be maintained by the development team of the organization. There have been planned some PoCs with potentials customers of the organization to validate its potential.</li> <li>Seamless component has been jointly developed with IBM. Both partners are participating in the EU H2020 project namely PolicyCloud, where partners will evolve this component. It is not expected for the coming future that the component is Open Sourced.</li> </ul>
Potential risk, barriers or	The solution is limited to use cases for which historical data does not need to be updated.

<p>limitations to the exploitation and sustainability</p>	<p>The components come to market much later than competitors Still mostly proprietary New projects which introduce table formats such as UDI and ICEBERG may reduce the novelty of this component especially since they cope with data transactions.</p>
---	--

### 5.2.11. UBITECH Limited

UBITECH is a highly innovative software house, systems integrator and technology provider, established to provide leading-edge intelligent technical solutions and consulting services to businesses, organizations and government to allow the efficient and effective secure access and communication with various heterogeneous information resources and services, anytime and anywhere. UBITECH enables real-time valid information processing and decision making, the realization of intelligent business environments, and B2B and B2C transactions by providing high added-value business-oriented and –based solutions.

UBITECH R&D team is highly motivated with the concepts introduced by BigDataStack in the direction to develop, integrate, deploy and evaluate innovative technologies, utilities, features and processes. Its role in the BigDataStack relates to the provision of cloud infrastructure and networking utilities. UBITECH team will also contribute towards the realization of information driven networking framework and BigDataStack analytics toolkit as a service on top of a bouquet of analytic algorithms.

Within BigDataStack, UBI has developed the Data Toolkit.

The exploitation plan developed by UBI is outlined as follow.

<p>Exploitation plan</p>	<p>The Data Toolkit has been developed with open source JS frameworks including VUE.js and the improvements since M18 have been integrated using REST APIs.</p> <p>The Data Toolkit has synergies with the Process Modelling and the Adaptive Visualization component. A potential exploitation path and the related commercial opportunities of the Data Toolkit may include an integrated product with the Process Modelling and the Adaptive Visualization component working in collaboration with the ATC partners to ease the automated machine learning and application setup, parameterisation and deployment operations.</p> <p>The new exploitation activities and enhancements of the Data Toolkit include the incorporation of more functionalities to ease user interaction and services configuration at the underlying BigDataStack testbed.</p> <p>UBI will ensure the sustainability of the Data Toolkit by further R&amp;D projects.</p>
--------------------------	---

<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>Developed by a small development team with limited resources availability</p> <p>It has not a fair potential in this market as the equivalent competitors' component (e.g. Google, Microsoft)</p> <p>Big players in this market have already launched an equivalent component</p> <p>The financial crisis does not leave space for non-critical addons</p> <p>COVID-19 crisis may not facilitate the addon of further features</p> <p>Specific tool which is tailored in cloud infrastructures</p>
---	---

### 5.2.12. TRUST-IT Services Limited

Founded in 2002, Trust-IT's central asset is a highly-skilled, dynamic, international team that performs quality ICT market research and result-oriented stakeholder communication and engagement strategies, working closely with SMEs, government and research institutions.

Trust-IT actively participates to EC-policy discussions around data through the Research Data Alliance, the European Open Science Cloud (EOSC) and is a member of ECSO the European cybersecurity organisation and works on cybersecurity & risk management and is a founding partner of the WISER "Wide-impact cybersecurity risk framework" [www.cyberwiser.eu](http://www.cyberwiser.eu) initiative (2015-2017) where it has both authored the concepts of the tools (CyberWISER Light, Essential & Plus) & delivered some of the technical implementation specifications to these products & services within the WISER Innovation Action project.

Trust-It coordinated key projects in the field of Data infrastructures and Cloud Computing such as RDA Europe 4.0 (the European Plug-into the Research Data Alliance [rd-alliance.org](http://rd-alliance.org)), CloudWATCH & CloudWATCH 2.0. ([www.cloudwatchhub.eu](http://www.cloudwatchhub.eu)) SLA-READY ([www.sla-ready.eu](http://www.sla-ready.eu)) & ICTFOOTPRINT.eu. It also leads the communication, outreach and dissemination as well as impact & exploitation activities for a range of different European funded projects in the field of Data Infrastructures, Cloud Computing, Cyber Security, ICT Standardisation, 5G, ICT services procurement including MAX (MATERIALS design at the eXascale), Full5G, SpeakNGI, SWForum and StandICT 2.0.

Trust-IT has gained high expertise in the ICT areas, covering both the technological and human interaction aspects and creating innovative business models & exploitation plans.

Trust-IT can mobilise a networked community around the stakeholders involved in BigDataStack and supports around structured and tailored marketing campaigns to ensure exploitation take-up that it will do for the roll-out of the workshops.

Our dedicated and direct cooperation and support in putting together coordinated workshops around the specific use cases are key to create a direct and solid impact to the project.

Creating one-pager readers digests or success stories from the deliverables and pulling out the impacts and results are key to pursue an effective exploitation strategy.

Knowledge and network gained in BigDataStack on Big Data research and innovation for European Industry will be leveraged upon in proposals, consortia and implementation of future R&D projects.

The exploitation plan elaborated by Trust-IT is as follow.

Exploitation plan	<p>Knowledge and network gained in BigDataStack on Big Data research and innovation for European Industry, will be leveraged upon in proposals, consortia and implementation of future R&amp;D projects.</p> <p>Working on BigDataStack, more specifically in Big Data domain, has allowed Trust-It to engage with the BDVA ecosystem and created valuable synergies with stakeholders from technology providers, consumers, research and policy makers. These synergies are a solid basis for the dissemination of BigDataStack results in future synergies. A few synergies a worth mentioning here:</p> <ul style="list-style-type: none"> <li>• Synergy created with I-BiDaaS, Track&amp;Know and Policy Cloud projects in a series of online events. Policy Cloud is currently working on taking up technologies developed in BigDataStack (seamless &amp; data skipping), already being disseminated via the Policy Cloud project.</li> <li>• Synergy created with flagship initiative INFINITECH through a joint awareness raising webinar. Key technologies developed in BigDataStack are being taken up by the INFINITECH project and disseminated as such. Connections have been made with the EBDVF task force on Fintech and Insurance tech, Trust-IT will build on these synergies in future projects and initiatives.</li> <li>• The synergy created through the co-authoring of an Open Access booklet with the BDVA, Trust-IT will leverage on this synergy in future projects.</li> <li>• The synergy created with the BDVe Coordination and Support Action led to the definition of a policy brief for policy makers. The recommendations defined on Policy for Data will be taken up by future Big Data R&amp;I projects such as Policy Cloud, when engaging with Policy Makers.</li> </ul>
-------------------	--

Additionally, BigDataStack has allowed Trust-IT to better know the standards involved in Big Data projects, since the open calls for StandICT project “*Supporting European Experts Presence in International Standardisation Activities in ICT*” were disseminated among the BigDataStack project partners. For the StandICT project, Trust-IT set the standards watch in place, monitoring the status of ICT standards at international level, on five priority areas of the Digital Single Market: 5G communications, cloud computing, cybersecurity, data technology, and IoT – Internet of Things. Currently Trust-IT is coordinating the StandICT 2.0 project. Trust-IT worked on the mapping of the standards used in the BigDataStack project and shared this with the StandICT team for a joint article. This knowledge gained here will be taken up in running and future Research and Innovation projects and initiatives in which Trust-IT is involved in e.g. Policy Cloud, StandICT 2.0.

Synergies between the StandICT and BigDataStack initiatives have been created in terms of standards identification, involvement of the StandICT Impact Creation Board Member Ray Walshe in the BigDataStack final event and joint exploration meetings with the Open Forum Europe. The knowledge gained from BigDataStack on the added value of standards use in Big Data research and Innovation as well as the Open Source default way of a standardised use of software will be shared and exploited in future Big Data related projects, such as Policy Cloud, where Trust-IT leads the work on standardisation.

BigDataStack has been involved in Open Source Community events such as the Open Source Forum, KubeCon and CloudNative Con, the flagship conference from Cloud Native Computing Foundation gathers adopters and technologists from leading open source and cloud native communities. Trust-IT will leverage on this identification for future R&I projects and initiatives.

In addition, in October 2020, the European Commission approved its new [Open Source Software Strategy 2020-2023](#), a part of the overarching [Digital Strategy Of The Commission](#) and contributing to the Digital Europe programme. The aim of the strategy is to increase the use of Open Source in both practical and strategic sectors, while recognising the importance of collaboration between governments, researchers, industry, and the public at large for building new,

	<p>innovative digital solutions that work across borders and towards technological sovereignty.</p> <p>Trust-IT leads the dissemination work in the EC funded project SWForum.eu, that aims to create a self-sustainable online forum that facilitates and encourages both researchers and practitioners as well as projects in software engineering, digital infrastructure, and cybersecurity to create intersections of expertise and a multidisciplinary approach to research and innovation. This openness and collaboration is very much in line with the <a href="#">European Commission's Goals For A Digital Europe</a> and, in this particular case, for Open Source Software.</p> <p>In BigDataStack Trust-IT lead the work on Communication and Dissemination, implementing and monitoring the strategies defined. BigDataStack assets, use cases and research stood at the core of the strategy, specifically targeting stakeholders from Research &amp; Academia, Policy Makers, Industry. Trust-It has used a mix of on-and offline collaterals, 1-page digests of project deliverables for reuse, videos for promotion and training as well as downloadable factsheets for end-users. The mix of dissemination and communication activities and outputs pulled out in BigDataStack can be used as knowledge base for future dissemination in R&amp;I projects, but also in the Horizons Results Booster, that encourages European funded projects to come together to identify a common portfolio of results and shows them how best to disseminate to end-users, with an eye on exploitation opportunities.</p> <p>Trust-IT is also involved in a number of initiatives behind supporting a European Open Science Cloud (EOSC) in Europe. It leads the communication activities on the EOSC Secretariat, EOSC-Hub, FAIRsFAIR, EOSC Pillar, EOSC Enhance, Reinforce, the Social Sciences and Humanities Open Cloud (SSHOC), BlueCloud and CS3MESH4EOSC projects, and is involved in ESCAPE. We support the building of the EOSC portal where a defined marketplace and service catalogue to help European researchers store, retrieve, access research data all in one place. Providing visibility for BigDataStack, to have visibility via the EOSC portal would be extremely important.</p>
<p>Potential risk, barriers or limitations to the</p>	<p>N/A</p>

exploitation and sustainability	
---------------------------------	--

### 5.2.13. *University of Piraeus Research Center*

UPRC, as a non-profit academic institution, intends to be involved in challenging, real-life problems that extend its research interests to new areas and thus advance and proliferate scientific knowledge. UPRC addresses research in the domain of software engineering and distributed computing and collaborates with IT companies with the goal of delivering research outcomes into the business world.

Participating in the BigDataStack project has been an opportunity for UPRC to interact with high-level research/industrial partners, being introduced to real-world use cases, and participating in the development of new technologies that can be exploited.

Moreover, dealing with big data is a challenging area at the forefront of technology and UPRC as a research centre is interested in this field of technology. BigDataStack has allowed UPRC to contribute with its expertise and to interact with other partners who have experience and knowledge that will be very useful to further push on in its different researches.

The exploitation plan developed by UPRC is as follow.

Exploitation plan	<p>UPRC has developed four components within BigDataStack project. Namely, Triple Monitoring, Application Dimensioning Workbench, Data Quality Assessment and Process Mapping. The exploitation plans for them are:</p> <ul style="list-style-type: none"> <li>• Triple Monitoring: UPRC plans to ensure the sustainability of this component by continuing supporting it through own-funding as well as through EU-funded research projects (e.g. MORPHEMIC) towards several enhancements. The current plan includes extensions for edge environments.</li> </ul> <p>The short-term plan refers to R&amp;D purposes in order to further enhance the framework. Based on that and the given collaborations with Greek IT companies we will try to commercialize it (as open-source, commercialization will reflect the support activities).</p> <p>Along with the QoS evaluation mechanism, and to this end we plan to explore potential routes with ATOS.</p> <ul style="list-style-type: none"> <li>• Application Dimensioning Workbench: UPRC plans to further enhance the workbench and extend both the benchmarking part</li> </ul>
-------------------	--

	<p>(with new benchmarks and targeted applications) and the forecasting part with additional scenarios beyond the provision of cloud resources.</p> <p>The goal is to primarily exploit it in new R&amp;D activities, with an emphasis on serverless architectures and provision of FaaS.</p> <p>For the overall workbench including the annotation of the application playbook, synergies can be identified with the Deployment Pattern Generation and Ranking.</p> <p>For the case of the benchmarking part of the workbench, this can be exploited as standalone, in order to offer stress testing services. Relevant actions in this case would be support plans to perform the actual benchmarking, as well as extension of the tool in order to adapt to a new data service/deployment platform/load injection process etc.</p> <ul style="list-style-type: none"><li>• <b>Data Quality Assessment:</b> UPRC will continue supporting it, currently applied to policy-related datasets with information obtained from public administration databases and from social media.</li></ul> <p>Mainly through R&amp;D activities, while commercialization will also be targeted as soon as results from additional sectors (as currently utilized under health- and policy- related sectors) verify its wide applicability to different domains.</p> <p>Some exploitation activities have been carried out in the scope of a new application domain: eHealth also with categorical data.</p> <ul style="list-style-type: none"><li>• <b>Process Mapping:</b> UPRC will continue our efforts in the process mapping approach by extending it for process maps (beyond single / atomic processes).</li></ul> <p>The next steps include further research building on top of the current prototype to extend it to reach TRL5 and then approach IT companies for its potential commercialization.</p> <p>The process mapping mechanism can be exploited with the process modelling framework, since it obtains a set of processes from the latter and enables their mapping to executable tasks.</p> <p>Some exploitation activities such Interviews with other stakeholders (specifically with policy makers) in order to foster the overall AutoML approach.</p>
--	--

<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>Triple Monitoring: The main barrier refers to the evaluation of the framework with more scenarios and data services. It needs to be evaluated to a greater extend (and to edge environments), which is happening now with new applications and configurations.</p> <p>Application Dimensioning Workbench: Load injection needs to be adapted to each service that needs to be benchmarked, which includes complexity and is a potential barrier to exploitation. However, the use of generic, multi-purpose baseline load definition tools such as Apache Jmeter is expected to minimize this risk.</p> <p>Moreover, the deployment and execution of the tests depends on the used deployment platform (e.g. Docker Swarm, Kubernetes etc), which is a limitation.</p> <p>Data Quality Assessment: The main barrier is the need for a respectable amount of data in order to train the model and discover intricate relationships between the values in a data set.</p> <p>Process Mapping: The main barrier is the consumption of resources required to perform various executions and conclude on the best-fit algorithm for different processes.</p>
---	--

#### 5.2.14. *Universidad Politecnica de Madrid*

UPM is a top technical university in Spain. UPM has a long entrepreneurship tradition with a 15-year old entrepreneurship program, ActuaUPM, which creates 20-25 start-ups every year from which 70% survive after 3 years. The Distributed Systems Lab (LSD) has a long technology transfer tradition having sold technology to Ericsson, Telefonica, and Bull among others. LeanXcale is a spinoff of UPM exploiting the results of several EU projects led by LSD-UPM (CumuloNimbo, CoherentPaaS and LeanBigData). The LSD counts with a cluster of 100 nodes for scalability testing of its solution and LeanXcale's.

The lab director, Prof. Marta Patiño, has been co-founder of the .COM joint institute between Santander bank and UPM. She is also co-founder of LeanXcale.

UPM participates in BigDataStack developing a CEP that will integrate with LeanXcale ultra-scalable database to deliver a big platform able to manage streaming data and data at rest in an integrated manner and UPM also expect to obtain a Combined infrastructure for data streaming and data at rest that will enable to deliver real-time analytics.

The exploitation plan for CEP is as follow.

Exploitation plan	The sustainability of CEP will be based on UPM funds and funding from
-------------------	---

	<p>national and regional projects.</p> <p>New features will be added to CEP as part of the research done at UPM in other research projects. Currently, UPM has been granted one national research project, one regional project and a European project where the CEP will be used.</p> <p>In addition, UPM has long experience in technology transfer signing contracts with companies for technology transfer and exploitation of research results. UPM has signed previously exploitation agreements with LeanXcale.</p> <p>UPM aims at exploiting the CEP components: CEP and the Triple Monitoring of the CEP.</p> <p>The exploitation route for those components is mainly to include them as part of the commercial solution of LeanXcale, with who UPM has previous commercial agreements, which enables LeanXcale UPM-LSD results in exchange to royalties paid to UPM. The component licensing will be proprietary.</p> <p>The commercialization roadmap for the CEP component will be undertaken by UPM.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>The main risk for CEP is the lack of funding for improving the TRL. Moreover, hiring people for upgrading the software is difficult for universities.</p> <p>Other barriers are:</p> <ul style="list-style-type: none"> <li>• Competing open source projects develop similar features</li> <li>• Immature implementation</li> </ul>

### 5.2.15. *University of Glasgow*

The University of Glasgow is a top university in the UK with around 3k academic staff and 30k students. It is research-focused and is a member of Universitas 21, the Russell Group and the Guild of European Research-Intensive Universities. Participation within BigDataStack is run by the Information Retrieval Group, part of the Information Data and Analysis Section of the School of Computing Science.

The Information Retrieval group is comprised of over 10 research staff and students working in different specialist areas, providing a broad skill base. We also maintain a dedicated compute cluster comprising over 30 machines with attached Big Data storage to support fast

on-site experimentation, in addition to having access to wider compute infrastructure provided by the University.

In addition to the Realization Engine (extended from the Deployment Recommendation Service), during the last 18 months, GLA has developed a new component not included in the previous D7.2, namely, Beta-RecSys Recommendations Development Framework.

The exploitation plan developed by GLA is as follow.

<p>Exploitation plan</p>	<p>The exploitation and sustainability plan for the components is:</p> <ul style="list-style-type: none"> <li>Realization Engine: The component is no longer dependant on other components of the BigDataStack platform. Instead, it supports interoperability with BigDataStack components via 'plugin' Operations.</li> </ul> <p>The sustainability plans for the service includes:</p> <ul style="list-style-type: none"> <li>Exploiting the service as part of the BigDataStack platform i.e. engage in joint exploitation activities.</li> <li>Using the component in training activities of GLA, primarily targeting undergraduate and post-graduate students and staff using the cluster infrastructure within the University. This may be expanded to include degree courses towards illustrating recommendations and information retrieval use cases, where the realization engine can significantly reduce the complexity of setting up such complex applications in lab environments.</li> <li>Developing a web site and video tutorials about the operation of the component/service.</li> <li>The component will be released as an independent open source project.</li> </ul> <p>In the short term, the component will be primarily used for academic purposes (R&amp;D, teaching). Other opportunities will be explored in case of follow up R&amp;D projects that can use the component (e.g. cross-over uses in the Flagship H2020 Infinittech project are being considered). The component will be released as an open source project for free. Subsequent commercialisation through platform training (e.g. via MOOC) may also be performed. A commercial exploitation will not be pursued prior to the open source release. However, the component is addressing the growing market of cloud infrastructures and services.</p> <p>GLA is offering the component both as a stand-alone service, as well as a component of the wider BigDataStack platform. We are interested in exploring joint exploitation activities with other</p>
--------------------------	---

	<p>partners.</p> <ul style="list-style-type: none"> <li> <b>Beta-RecSys Recommendations Development Framework:</b> Beta-RecSys will be exploited individually as it does not depend on any other BigDataStack component. The framework exhibits significant innovation potential as it can support the development of state-of-the-art recommender systems based on deep learning algorithms. Beta-RecSys has already been released with an open source license. It is available at: <a href="https://github.com/beta-team/beta-recsys">https://github.com/beta-team/beta-recsys</a>.         </li> </ul> <p>GLA is committed to sustaining the framework and to develop it further. To this end, the following activities are already planned:</p> <ul style="list-style-type: none"> <li>Use of the framework in other R&amp;D projects where GLA participates (e.g., H2020 INFINITECH).</li> <li>Pursuit of additional grants where the framework can be further developed.</li> <li>Use of the framework to support the practical part of University courses and professional training courses developed by GLA.</li> </ul> <p>There are already plans for using the framework in other H2020 projects (e.g., INFINITECH). Moreover, GLA aspires to offer the framework to industrial enterprises using open source products/services. To this end, it will track relevant downloads and will provide support to end-users if needed.</p> <p>In addition to exploiting the Beta-RecSys independently, GLA is interested in supporting the exploitation and take up of the use case systems (ATOS, GLA) where our technology is used.</p>
<p>Potential risk, barriers or limitations to the exploitation and sustainability</p>	<p>The primary risks to the exploitation of the Realization Engine are the speed of development in the cloud/cluster management space and the introduction of new disruptive technology. As cluster management systems (e.g. Docker Swarm and Openshift) become more mature, it is anticipated that some of the features of the Realization Engine may become native to those systems, reducing the platform added value. Additionally, the platform is only compatible with Openshift-based installations of Kubernetes at this time (although this is rectifiable). Finally, so far, the product is only TRL level 5, and as such as yet to be demonstrated in large-scale environments beyond BigDataStack.</p> <p>The potential barriers and risks for Beta-RecSys Recommendations Development Framework to penetrate the recommendation engine</p>

	<p>market include:</p> <ul style="list-style-type: none"><li>• The lack of financial resources to further develop and advance the product; and</li><li>• The emergence of similar frameworks with stronger support and community</li></ul>
--	--

## 5.3. BigDataStack joint exploitation

---

BigDataStack consortium identified three business models to exploit BigDataStack outcomes:

- **BigDataStack-as-a-Service:** BigDataStack commercialization Business Model to offer Big Data products and services.
- **Educational Business Model:** Applies to educational purpose and knowledge transfer in general. Academia partners such as **UPRC** and **GLA** have already put in place this BM and are offering educational services to the research community and students.
- **Consultancy Business Model:** Provides Big Data Consultancy Services in general, and for application of the BigDataStack Use cases to the specific sectors such as Shipping, Insurance and Retail. **Worldline** is currently providing this type of services to its customer involved within BigDataStack use case, to try to bring to a real use case the knowledge gained during the project. Moreover, **GFT** is exploring commercial opportunities with its customers, such as SulAmerica, HDI or NEXI, to provide consultancy services to adopt BigDataStack Insurance use case knowledge.

Within D7.2 we included the Business Models Canvas<sup>53</sup> (BMC) of each BM.

Next section focuses on the Commercialization Business Model (BigDataStack-as-a-Service) for joint exploitation and the joint exploitation strategy.

### 5.3.1. BigDataStack Commercialization Business Model

The business model for commercialization the BigDataStack outcomes (BigDataStack-as-a-Service) considers two different ways to provide BigDataStack key offering:

- **Offer turn-key big data management and analytics solutions.** BigDataStack consortium would extend the project use cases to commercial use cases and provide them to use case partners or customers from Retail, shipping and insurance sector. Additionally, other use cases will arise and will be developed to expand the offering to other sectors.
- **Offer BigDataStack tools or services to solution providers.** Instead of developing turn-key solutions, BigDataStack tools will be offered to other companies such as innovative SMEs working on Big Data to enable them to create their own turn-key solutions or integrate them into their solutions.

Here, we provide the updated version of the BMC for the Commercialization Business Model:

---

<sup>53</sup> <https://www.strategyzer.com/canvas/business-model-canvas>

Key Partners	Key Activities	Value Proposition	Customer Relationships	Customer Segments
<p>Infrastructure Providers</p> <p>Service providers (Cloud Providers)</p>	<p>Update and Improvement, update and maintenance of SW components version</p> <p>Support and maintenance services</p>	<p>Complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that addresses the emerging needs of data operations and applications</p> <p>Frontrunner data-driven architecture and system ensuring that infrastructure management will be fully efficient and optimized for data operations and data-intensive applications</p> <p>Use Cases assets and knowledge on the application of Big Data in Retail, Shipping and Insurance sectors</p>	<p>Webinars and training courses</p> <p>Incorporation of new use cases of different sectors</p> <p>Software Update</p> <p>Strategic partnerships with other Big Data providers (SMEs, Research Community, Industry)</p>	<p>SMEs and Big Industry</p> <p>Public Organizations</p> <p>Use Case Partners</p> <p>Customers on Retail, Shipping and Insurance Sectors</p> <p>Industrial Sectors with Big Data needs to improve their businesses performance (Industry 4.0, e-Health, etc.)</p> <p>Infrastructure and solution providers</p> <p>Application Providers</p>
	<p><b>Key Resources</b></p> <p>BigDataStack components and key offerings (i.e. main building blocks)</p> <p>Consortium experience and Know-how</p> <p>ICT Infrastructures</p>		<p><b>Channels</b></p> <p>BigDataStack project Website</p> <p>Project Dissemination (Conferences, events, scientific publications)</p> <p>Dissemination on BDVA and other EU Big Data initiatives</p> <p>Industrial events</p>	
<p><b>Cost Structure</b></p> <p>Marketing</p> <p>Maintenance of software and infrastructures</p> <p>Personal costs</p> <p>Server Costs</p> <p>Infrastructure providers costs</p> <p>Variable costs, depending on the scenario</p>			<p><b>Revenue Streams</b></p> <p>Pay-as-you-go fee for the offering of BigDataStack tools</p> <p>Agreements with customers for turn-key projects</p>	

Table 8. BigDataStack commercialization Business Model Canvas

### 5.3.2. Joint Exploitation Strategy

BigDataStack consortium is mainly industrial as it is composed of almost 80% Industrial partners. Industrial partners are mostly interested in commercialization of the project assets in a joint or isolated way, Use Case partners are interested in reuse Use Case pilots in their own business (DANAOS) or for providing it to their customers (GFT and Worldline), and

Academia partners are interested in disseminating scientific results and participate in R&D related to their research line.

The joint exploitation activities analysed by BigDataStack consortium can be classified considering the partners' profiles:

- Commercial exploitation of the BigDataStack framework under the commercialization BM previously proposed or the commercialization of group of components or standalone components
- Continue participating in events to promote BigDataStack outcomes: participation in events, conferences, workshops, seminars
- Publication of joint scientific articles or the use case outcomes in industrial journals

To know the partners' willingness to participate in these joint activities, the exploitation team has conducted a short questionnaire. The questionnaire outcomes are depicted below.

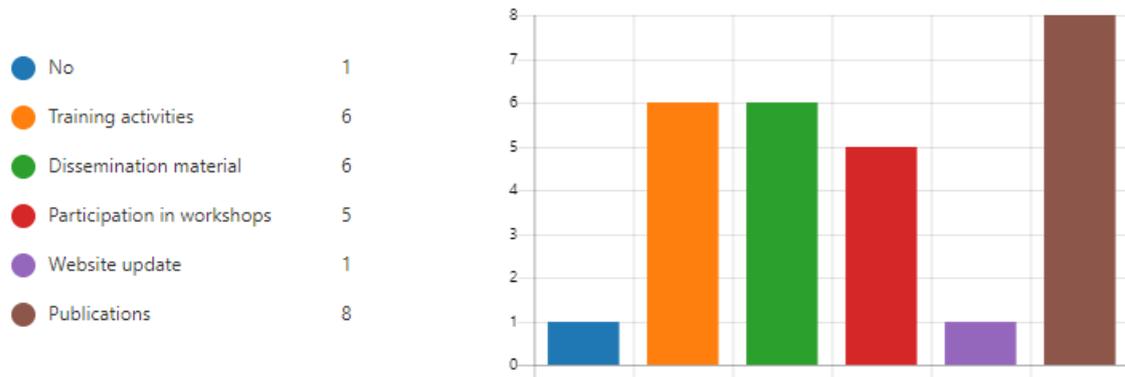
4. After the project, Would you be willing to join a BigDataStack support team to reply to requests from developers or potential customers?



5. In what type of activities are you willing to participate within the support team?



6. Would you be willing to participate in promotional activities of BigDataStack results after the project ends?



7. The European Commission expects the consortium partners to engage in joint exploitation for BigDataStack results if commercial opportunities arise after the project. For this, a Memorandum of Understanding (MoU) is usually signed by partners, in which aspects such as IPR, ownerships, etc. are covered. Would you sign an MoU for future joint exploitation?

Yes, I would prefer to sign a M...	5
No, I would prefer to sign a M...	5
Due to my organization's inter...	1



8. In case of a future commercialization/exploitation opportunity in which your organisation were not interested. Would your organisation sign a Third-party exploitation agreement with the consortium partners to allow them to exploit BigDataStack solution? (royalties, rewards, etc. will be negotiated among partners).

Yes	3
No	0
I don't know now	8



The outcomes of the questionnaire can be summarized as follow:

- Partners would participate in supporting actions for early-adopters
- Most partners would be willing to participate in promotional activities, mainly in scientific publications, dissemination material, and training activities
- Some industrial partners, due legal restrictions from their companies, cannot sign at this stage a Memorandum of Understanding (MoU) or an Exploitation agreement for future joint exploitation. Moreover, the rest of the partners would prefer to sign that kind of document once a commercial opportunity arises

Although the BigDataStack Consortium hasn't signed an Exploitation Agreement, all the partners have shown their willingness in facilitating future exploitation or commercialization of BigDataStack and study the possibility of a third-party agreement (to allow other organization to exploit its component) in case they weren't interested in participating in the joint exploitation.

Additionally, all partners commit to continue looking for joint opportunities to exploit BigDataStack outcomes in the near future, as they have already demonstrated and reported within Section 3.

## 6. Conclusions

This deliverable is the last one resulting from the work done within T7.1, and it presents the results of the exploitation activities carried out during the second period of the project, from M18 to M34.

Although COVID-19 pandemic situation, it hasn't been a limitation for the proper performance of exploitation activities. BigDataStack partners have actively worked on their exploitation strategies and in the activities proposed by the exploitation team, as evidenced by the achievements in exploiting BigDataStack outcomes as well as in ensuring sustainability.

During the first phase of the project, as duly reported in the previous deliverable, the work was focused on getting knowledge of the BigDataStack solution and of its business potential. We defined the value proposition, value chain, identification of SW components, and the market context of the project, such as competitors and trends, and several business models to exploit BigDataStack outcomes; moreover, partners introduced their preliminary exploitation interests. Conversely, the second period of the project, the work was focused to put in place the individual exploitation and commercialization strategies from partners, and the BigDataStack consortium made a significant effort to identify opportunities in order to jointly exploit the different BigDataStack software components.

Regarding the sustainability of the BigDataStack outcomes, the most remarkable achievements are the joining to the BDVA Innovation Marketplace, the Collaboration agreement signed with the EOSC-DIH, to continue promoting project results and identifying future exploitation opportunities, and of course the upstream contribution to the OS communities.

BigDataStack can be considered a successful R&D project since several exploitable assets developed within the project are already being commercialized, as is the case with IBM and RED HAT, and most of the partners are participating together in ongoing or future R&D projects to evolve the BigDataStack SW components to an advanced TRL nearer to commercialization.