

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>

## D7.2 – Exploitation Plan and Business Potential-V1

Work Package	WP7 – Communication, Exploitation, Standardisation, & Roadmapping
Lead Author (Org)	Ana Belén González (ATOS)
Contributing Author(s) (Org)	All partners
Due Date	01.07.2019
Date	01.07.2019
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	Date	Author	Notes
	11/06/2018	Ana Belén González (ATOS)	Elaboration of Individual Exploitation Template to collect partners' first exploitation interest. Sent to all partners to be filled in.
	18/02/2019	Ana Belén González (ATOS)	Elaboration of Components Template to collect information about the SW components that are going to be developed within the project. Sent to all partners to be filled in.
0.1	27/05/2019	Ana Belén González (ATOS)	ToC and initial draft
0.2	01/06/2019	Ana Belén González (ATOS)	Integration of Partners' Individual Exploitation Plans from all partners (Section 7)
0.3	03/06/2019	Ana Belén González (ATOS)	Integration of information about SW components from all partners (Section 6.2)
0.4	15/06/2019	Ana Belén González (ATOS)	Elaboration of the rest of the sections
0.5	28/06/2019	Ana Belén González (ATOS)	Executive Summary and Conclusions
0.6	30/06/2019	Yosef Moatti (IBM), Dimosthenis Kyriazis (UPRC)	Review of V0.5
1.0	01/07/2019	Ana Belén González (ATOS)	Final version

#### Disclaimer

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.

## Table of Contents

Table of Contents.....	3
List of tables .....	4
List of figures.....	6
Acronyms .....	6
1 Executive Summary.....	8
2 Introduction .....	9
2.1. Purpose and Scope.....	9
2.2. Structure of the document .....	9
3 Exploitation Objectives and Activities Plan.....	10
3.1. Impact Creation Board activities planned .....	11
4 Market Analysis.....	13
4.1. Market Overview .....	13
4.2. Stakeholder Analysis and Value Chain.....	14
4.3. Initial SWOT Analysis .....	16
4.4. BigDataStack Value Proposition.....	17
5 Potential Business Models for BigDataStack .....	19
5.1. BigDataStack commercialization Business Model .....	19
5.2. Educational Business Model .....	20
5.3. Consultancy Business Model .....	21
6 BigDataStack Exploitable Assets .....	23
6.1. Analysis of Exploitable Assets .....	25
6.1.1. SQL Data Skipping .....	25
6.1.2. Dynamic Orchestration.....	27
6.1.3. Network Policy support at Kuryr .....	33
6.1.4. Kuryr Cluster Network Operator .....	34
6.1.5. NvME-mdev Kernel driver .....	35
6.1.6. QoS Evaluation.....	36
6.1.7. Information Driven Networking .....	42
6.1.8. Process Modelling Framework .....	45
6.1.9. Visualization Environment.....	47
6.1.10. Seamless Analytics.....	49
6.1.11. Adaptable Distributed Storage .....	51
6.1.12. Data Toolkit .....	54
6.1.13. Data Quality Assessment.....	56
6.1.14. Application Dimensioning Workbench.....	59
6.1.15. Process Mapping .....	63
6.1.16. Complex Event Processing.....	66
6.1.17. Deployment Recommendation Service.....	68
6.2. Preliminary IPR and Licensing Analysis .....	70
6.2.1. Data Skipping .....	71
6.2.2. Dynamic Orchestration.....	71
6.2.3. Network Policy support at Kuryr, Kuryr Cluster Network Operator and NvME-	

mdev Kernel driver .....	71
6.2.4. QoS Evaluation.....	71
6.2.5. Information Driven Networking .....	72
6.2.6. Process Modelling Framework .....	72
6.2.7. Visualization Environment.....	72
6.2.8. Seamless Analytics.....	72
6.2.9. Adaptable Distributed Storage .....	73
6.2.10. Data Toolkit .....	73
6.2.11. Data Quality Assessment.....	73
6.2.12. Application Dimensioning Workbench.....	73
6.2.13. Process Mapping .....	73
6.2.14. Complex Event Processing.....	74
6.2.15. Deployment Recommendation Service.....	74
7 Exploitation Strategy.....	75
7.1. Individual Exploitation Plans.....	75
7.1.1. IBM ISRAEL – SCIENCE AND TECHNOLOGY LTD.....	75
7.1.2. NEC EUROPE LTD .....	76
7.1.3. RED HAT ISRAEL LTD .....	78
7.1.4. ATOS SPAIN SA.....	79
7.1.5. ATOS Worldline Iberia SA .....	80
7.1.6. GFT ITALIA SRL .....	82
7.1.7. DANAOS SHIPPING COMPANY LIMITED .....	83
7.1.8. SINGULARLOGIC ROMANIA COMPUTER APPLICATIONS SRL .....	85
7.1.9. ATHENS TECHNOLOGY CENTER SA .....	86
7.1.10. LEANXCALE SL .....	88
7.1.11. UBITECH LIMITED.....	88
7.1.12. TRUST-IT SERVICES LIMITED .....	90
7.1.13. UNIVERSITY OF PIRAEUS RESEARCH CENTER .....	91
7.1.14. UNIVERSIDAD POLITECNICA DE MADRID .....	93
7.1.15. UNIVERSITY OF GLASGOW.....	94
7.2. Joint Exploitation Strategy .....	95
8 Conclusions .....	97

## List of tables

TABLE 1. ICB CONFIRMED MEMBERS.....	11
TABLE 1. INITIAL BIGDATASTACK SWOT ANALYSIS .....	17
TABLE 2. BIGDATASTACK COMMERCIALIZATION BUSINESS MODEL CANVAS .....	20
TABLE 4. EDUCATIONAL BUSINESS MODEL CANVAS .....	21
TABLE 5. CONSULTANCY BUSINESS MODEL CANVAS .....	22
TABLE 6. SUMMARY LIST OF BIGDATASTACK SW COMPONENTS.....	24
TABLE 7. INVOLVEMENT OF SW COMPONENTS IN BIGDATASTACK KEY OFFERING .....	25
TABLE 8. SQL DATA SKIPPING DESCRIPTION .....	26
TABLE 9. SQL DATA SKIPPING COMPETITORS .....	26

TABLE 10. SQL DATA SKIPPING SWOT .....	27
TABLE 11. SQL DATA SKIPPING PRELIMINARY EXPLOITATION ROUTE .....	27
TABLE 12. DYNAMIC ORCHESTRATOR DESCRIPTION .....	29
TABLE 13. DYNAMIC ORCHESTRATOR COMPETITOR 1.....	30
TABLE 14. DYNAMIC ORCHESTRATOR COMPETITOR 2.....	31
TABLE 15. DYNAMIC ORCHESTRATOR SWOT ANALYSIS .....	32
TABLE 16. DYNAMIC ORCHESTRATOR PRELIMINARY EXPLOITATION ROUTE.....	32
TABLE 17. NETWORK POLICY SUPPORT AT KURYR DESCRIPTION.....	33
TABLE 18. KURYR CLUSTER NETWORK OPERATOR DESCRIPTION .....	34
TABLE 19. NVME-MDEV KERNEL DRIVER DESCRIPTION .....	35
TABLE 20. QoS EVALUATOR DESCRIPTION .....	38
TABLE 21. QoS COMPETITOR 1 .....	39
TABLE 22. QoS EVALUATOR COMPETITOR 2 .....	40
TABLE 23. QoS COMPETITOR 3.....	41
TABLE 24. QoS EVALUATOR SWOT ANALYSIS .....	41
TABLE 25. QoS EVALUATOR PRELIMINARY EXPLOITATION ROUTE.....	42
TABLE 26. INFORMATION DRIVEN NETWORKING.....	43
TABLE 27. INFORMATION DRIVEN NETWORKING COMPETITOR 1 .....	44
TABLE 28. INFORMATION DRIVEN NETWORKING COMPETITOR 2 .....	44
TABLE 29. INFORMATION DRIVEN NETWORKING SWOT ANALYSIS.....	45
TABLE 30. INFORMATION DRIVEN NETWORKING PRELIMINARY EXPLOITATION ROUTE .....	45
TABLE 31. PROCESS MODELLING FRAMEWORK DESCRIPTION.....	46
TABLE 32. PROCESS MODELLING FRAMEWORK SWOT ANALYSIS.....	46
TABLE 33. PROCESS MODELLING FRAMEWORK PRELIMINARY EXPLOITATION ROUTE .....	47
TABLE 34. ADAPTABLE VISUALIZATION .....	48
TABLE 35. ADAPTABLE VISUALIZATION SWOT ANALYSIS.....	49
TABLE 36. ADAPTABLE VISUALIZATION PRELIMINARY EXPLOITATION ROUTE .....	49
TABLE 37. SEAMLESS DESCRIPTION .....	50
TABLE 38. SEAMLESS SWOT ANALYSIS.....	50
TABLE 39. SEAMLESS PRELIMINARY EXPLOITATION ROUTE .....	51
TABLE 40. ADAPTABLE DISTRIBUTED STORAGE DESCRIPTION .....	53
TABLE 41. ADAPTABLE DISTRIBUTED STORAGE COMPETITOR .....	53
TABLE 42. ADAPTABLE DISTRIBUTED STORAGE SWOT ANALYSIS .....	54
TABLE 43. ADAPTABLE DISTRIBUTED STORAGE PRELIMINARY EXPLOITATION ROUTE.....	54
TABLE 44. DATA TOOLKIT DESCRIPTION .....	55
TABLE 45. DATA TOOLKIT COMPETITOR .....	55
TABLE 46. DATA TOOLKIT SWOT ANALYSIS.....	56
TABLE 47. DATA TOOLKIT PRELIMINARY EXPLOITATION ROUTE .....	56
TABLE 48. DATA CLEANING DESCRIPTION.....	57
TABLE 49. DATA CLEANING COMPETITOR .....	58
TABLE 50. DATA CLEANING SWOT ANALYSIS .....	58
TABLE 51. DATA CLEANING PRELIMINARY EXPLOITATION ROUTE.....	58
TABLE 52. APPLICATION DIMENSIONING WORKBENCH DESCRIPTION.....	60
TABLE 53. APPLICATION DIMENSIONING WORKBENCH COMPETITOR 1.....	60
TABLE 54. APPLICATION DIMENSIONING WORKBENCH COMPETITOR 2.....	61

TABLE 55. APPLICATION DIMENSIONING WORKBENCH COMPETITOR 3 .....	61
TABLE 56. APPLICATION DIMENSIONING WORKBENCH SWOT ANALYSIS .....	62
TABLE 57 APPLICATION DIMENSIONING WORKBENCH PRELIMINARY EXPLOITATION ROUTE.....	63
TABLE 58. PROCESS MAPPING DESCRIPTION.....	64
TABLE 59. PROCESS MAPPING COMPETITOR 1.....	64
TABLE 60. PROCESS MAPPING COMPETITOR 2.....	65
TABLE 61. PROCESS MAPPING SWOT ANALYSIS .....	65
TABLE 62. PROCESS MAPPING PRELIMINARY EXPLOITATION ROUTE .....	66
TABLE 63. COMPLEX EVENT PROCESSING DESCRIPTION .....	67
TABLE 64. COMPLEX EVENT PROCESSING COMPETITOR.....	67
TABLE 65. COMPLEX EVENT PROCESSING SWOT ANALYSIS .....	67
TABLE 66. COMPLEX EVENT PROCESSING PRELIMINARY EXPLOITATION ROUTE.....	68
TABLE 67. DEPLOYMENT RECOMMENDATION SERVICE DESCRIPTION .....	69
TABLE 68. DEPLOYMENT RECOMMENDATION SWOT ANALYSIS.....	69
TABLE 69. DEPLOYMENT RECOMMENDATION SERVICE PRELIMINARY EXPLOITATION ROUTE .....	70
TABLE 70. SQL DATA SKIPPING LIBRARIES AND LICENSES .....	71
TABLE 71. DYNAMIC ORCHESTRATOR LIBRARIES AND LICENSES.....	71
TABLE 72. QoS EVALUATOR LIBRARIES AND LICENSES .....	71
TABLE 73. INFORMATION DRIVEN NETWORKING LIBRARIES AND LICENSES.....	72
TABLE 74. PROCESS MODELLING FRAMEWORK LIBRARIES AND LICENSES .....	72
TABLE 75. VISUALIZATION ENVIRONMENT LIBRARIES AND LICENSES .....	72
TABLE 76. SEAMLESS LIBRARIES AND LICENSES .....	72
TABLE 77. DATA TOOLKIT LIBRARIES AND LICENSES .....	73
TABLE 78. DATA CLEANING LIBRARIES AND LICENSES.....	73
TABLE 79. APPLICATION DIMENSIONING WORKBENCH LIBRARIES AND LICENSES .....	73
TABLE 80. PROCESS MAPPING LIBRARIES AND LICENSES .....	73
TABLE 81. COMPLEX EVENT PROCESSING LIBRARIES AND LICENSES.....	74
TABLE 82. DEPLOYMENT RECOMMENDATION SERVICE LIBRARIES AND LICENCES .....	74
TABLE 83. SUMMARY OF PARTNERS INDIVIDUAL EXPLOITATION INTERESTS AND INTENTIONS .....	75

## List of figures

FIGURE 1. BIGDATASTACK VALUE CHAIN.....	16
FIGURE 2. BIGDATASTACK KEY OFFERING .....	23
FIGURE 3. OS LICENSES COMPATIBILITY .....	70

## Acronyms

AIOITI	Alliance for Internet of Things Innovation
API	Application Program Interface
ATC	Athens Technology Center SA
AWS	Amazon Web Services
B	Billions
BDVA	Big Data Value Association

BM	Business Model
BMC	Business Model Canvas
CPS	Cyber Physical Systems
DANAOS	DANAOS Shipping Company Limited
EOSC	European Open Science Cloud
GFT	GFT Italia SRL
GLA	University of Glasgow
HTTP	Hypertext Transfer Protocol
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
ms	Millisecond
NEC	NEC Europe LTD
OS	Open Source
PPP	Public and Private Partnership
QoS	Quality of Service
R&D	Research and Development
RHT	Red Hat Israel LTD
RL	Reinforced Learning
SILO	Singularlogic Romania Computer Application SRL
SLO	Service Level Objectives
SME	Small and Medium Enterprise
SW	Software
SWOT	Strengths, Weaknesses, Opportunities, and Threats
TRL	Technology Readiness Level
UBI	Ubitech Limited
UK	United Kingdom
UPM	University of Madrid
UPRC	University of Piraeus Research Center
US	United States
WP	Workpackage

## 1 Executive Summary

---

This deliverable provides the foundation to define an appropriate exploitation for the BigDataStack results. The basis to define a realistic exploitation is the identification of the exploitable assets of the project, technical features, competitors, licensing analysis, etc. all the aspects needed to position the asset in the market context. At this regard, BigDataStack has identified 17 software components / outcomes developed within the BigDataStack key offering proposed.

In addition to the identification of the exploitable assets, a market analysis has been carried out to define the market context of the project as well as the stakeholders, SWOT analysis and value proposition. To complement this, a preliminary study of three business models have been described, not just to exploit the BigDataStack key offerings but also to exploit the knowledge acquired during the project for educational and consultancy purposes.

Finally, an initial joint exploitation strategy has been defined, which will be completed once the consortium has decided the business model to exploit the BigDataStack.



## 2 Introduction

---

### 2.1. Purpose and Scope

This deliverable is the first deliverable of a set of two Exploitation and Business Potentials deliverables included in the BigDataStack Grant Agreement, and due in M18 and M36.

The objective of this deliverable is to report on the activities carried out into Task 7.1 from WP7, which includes Market analysis, Business Potentials and Exploitation of BigDataStack outcomes.

### 2.2. Structure of the document

This deliverable is structured in 5 main sections:

- Section 3. Exploitation Objectives and Activities: This section provides an overview of the exploitation activities to be carried out during the project lifetime and in which deliverables they are going to be reported.
- Section 4. Market Analysis: It provides information about the market context, stakeholders, SWOT analysis and value proposition is provided in this section.
- Section 5. Potential Business Models for BigDataStack: Three business models have been identified so far, and they are briefly described in this section.
- Section 6. BigDataStack Exploitable Assets. This section provides wide information about all the SW components developed by the partners and included in the BigDataStack Key offering. This information encompasses technical aspects, competitors, SWOT analysis of the components, IPR and licensing analysis, and a preliminary exploitation route for the components.
- Section 7. Exploitation Strategy. This section shows the partners' preliminary exploitation interests and intentions at this stage of the project.

### 3 Exploitation Objectives and Activities Plan

---

The goal of BigDataStack project is to deliver 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>1</sup>.

BigDataStack has defined the following specific Exploitation Objectives:

- Perform a market analysis to assess the market context of the BigDataStack results
- Explore Business Scenarios to exploit BigDataStack outcomes and define the best-suited Business Models for them
- Identify all the software components developed within the project and study their features to find exploitation routes
- Help partners to elaborate their Individual Exploitation Plans. This plan is elaborated in the first period of the project and it is improved during the project lifetime.
- Work closely with the Impact Creation Board to get feedback from an industrial point of view to define a Roadmap adoption for project outcomes.

The tasks that will contribute to achieve those goals are within WP7, which consists of 4 different tasks, two of them regarding Exploitation and adoption of project outcomes:

- Task 7.1 Market Analysis, Business scenarios and Exploitation [M1-M36]
- Task 7.4 Roadmapping and Impact on adoption [M25-M36]

Two deliverables will report the activities carried out within the tasks previously mentioned, and the information included in those deliverables is:

---

<sup>1</sup> DoA, BigDataStack Consortium, 2017

### D7.1 Exploitation plan and Business Potential V1 -M18

- Market context: Market trends, competitors, stakeholders, SWOT Analysis, in the Big Data Ecosystem. Definition of BigDataStack Value Proposition
- Business Potentials and Business Models Scenarios
- Identification and Analysis of Exploitable Assets: Technical description, competitors, Preliminary IPR and License analysis
- Partners' Exploitation Plans
- Initial joint exploitation and sustainability strategy

### D7.3 Exploitation Plan and Business Potential V2 -M34

- General Market Context Update and Market analysis of the BigDataStack key offerings and software components.
- Extended Analysis of Business Models identified in first period
- Identification of Exploitation Routes for software components or cluster of them
- Joint Exploitation Plan and Sustainability Plan. Internal Workshop among Partners about Joint Exploitation
- Improvement and Update of Partners' exploitation plans
- Survey to the Impact Creation Board (ICB) to get their feedback in order to define a realistic Roadmap adoption strategy
- Definition of a Roadmap adoption plan

## 3.1. Impact Creation Board activities planned

The BigDataStack consortium set up an external Impact Creation Board (ICB), mainly composed of industry experts, to provide knowledge, guidance and advice to the consortium on how to exploit the project outcomes.

The list of ICB members is:

Jörn Altmann	Seoul University
Sue Dale	Tech UK
Rene Heek	NTT Data
Yücel Karabulut	Oracle US
Adam Kocoloski	Cloudant
Karsten Oberle	Nokia Labs

Table 1. ICB confirmed members

Once the project has delivered the first developments and this deliverable has been elaborated, the activities planned for the second period with the ICB are:

- Share D7.2 with the ICB to update their market context, exploitable assets, business models identified, etc. This information will help ICB to know the project outcomes and the market context for them.
- Elaborate and circulate a questionnaire among ICB members regarding exploitation aspects as well as technical requirements considered for the development of project results. With this action, the BigDataStack consortium aims at receiving feedback about how to promote the use of project results and keeping the project objectives aligned with research community and industrial demands. Different meeting calls will

be organized to discuss about the questionnaire outcomes between the ICB and the BigDataStack Consortium.

In addition to the previous activities, the ICB will support the uptake of project outcomes in social and networked media, as well as help the BigDataStack Consortium to explore synergies with others research communities and R&D projects researching in the same domain.

## 4 Market Analysis

---

### 4.1. Market Overview

The amount of data in the world is expected to reach 44 zettabytes by 2020<sup>2</sup> and Big data management and data analytics continue to be the most dynamic and fastest growing segment of the IT industry. Global Data Market Size 2017-2019<sup>3</sup>, already predicted that European Data market will exceed \$4 B in 2019.

This increase of data has led to a need of technologies to process those data and get valuable insights from them, and IDC Forecast Revenues for Big Data and Business Analytics solutions<sup>4</sup> will reach \$189,1 B worldwide in 2019 with a double-digit annual growth through 2022.

This data explosion has fostered a big data technology industry with new start-ups and established vendors, both providing new innovative big data solutions to deal with data.

“Big Data 100 list”<sup>5</sup> has identified vendors that are developing innovative big data solutions and services that help business face with big data challenges. These vendors have been firstly divided into four categories:

- *Business Analytics*: Companies delivering solutions from simple-to-use reporting and visualization tools to advanced analytics and data science platforms for dealing with the most complex data analysis tasks. MarketsandMarkets forecast that analytics software and business intelligence market would reach \$26.78 B by 2020.
- *Data Science and Machine Learning*: Companies developing big data software that use analytics and other techniques. According to Statistics Market Research Consulting, the global market for data science platforms will reach \$128,21 B by 2022.
- *Big Data Systems and Platforms*: Companies with on-premise and cloud systems to build and operate big data systems.
- *Data Management and Integration*: Companies offering from next-generation database software to advanced data integration and preparation tools.

It seems to be that collecting, cleaning, integrating, and managing data are the biggest challenges for the big data ecosystem so making software tools to perform these tasks are of critical importance. BigDataStack delivers tools to facilitate these tasks, and BigDataStack key offering can be included into several of the aforementioned categories (e.g. the LeanXcale database in the category of “*Data Management and Integration*” as well as in the category of “*Big Data Systems & Platforms*”, the data skipping approach in the category of “*Business Analytics*”, etc).

In this context, leading Big Data industry players such as IBM, HP, Google, SAP, Cloudera, and Oracle, are investing in R&D, for the development of unified big data solutions to provide

---

<sup>2</sup> <https://www.emc.com/leadership/digital-universe/2014iiview/executive-summary.htm>

<sup>3</sup> [https://www.onaudience.com/resources/wp-content/uploads/2018/09/OnAudience.com\\_Global\\_Data\\_Market\\_Size\\_2017-2019-1.pdf](https://www.onaudience.com/resources/wp-content/uploads/2018/09/OnAudience.com_Global_Data_Market_Size_2017-2019-1.pdf)

<sup>4</sup>

[https://www.idc.com/getdoc.jsp?containerId=prUS44998419&utm\\_medium=rss\\_feed&utm\\_source=Alert&utm\\_campaign=rss\\_syndication](https://www.idc.com/getdoc.jsp?containerId=prUS44998419&utm_medium=rss_feed&utm_source=Alert&utm_campaign=rss_syndication)

<sup>5</sup> <https://www.crn.com/news/storage/the-big-data-100-2019>

improved analytics and integrated management of data. Moreover, these players are focusing on strategies of merging and acquisition of emerging companies / start-ups to diversify their product portfolio with big data technologies. For instance, in Microsoft acquired Revolution Analytics and IBM acquired Cloudant and Cleversafe.

Regarding the European Big Data landscape, the Big Data Value Strategic Research and Innovation Agenda<sup>6</sup> developed by BDVA<sup>7</sup> has analysed and identified the technical priorities and objectives to be addressed to achieve a more competitive Big Data landscape in Europe.

Among those BDVA's technical priorities and objectives, we have identified four that are aligned with the solutions delivered by BigDataStack project:

- Optimisation of architectures for real-time analytics of data at rest and in motion, enabling data-driven decision-making, especially in heterogeneous datasets.
- Solutions leading towards the use of big data value technology for increased productivity, optimised production, more efficient logistics, etc.
- Solutions to challenge with data management lifecycle and governance, data-as-a-service model and paradigm, data-in-motion and at rest, efficient mechanisms for storage and processing, analytics frameworks & processing, data visualization, etc.
- Tools with usability and open source approach, to address the lack of Big Data skills.

Big Data platforms, services and tools are creating new opportunities for industries in different sectors. Next, are depicted the sectors that make the most from big data analytics:

- E-Health/Medicine: This sector is using big data and analytics to improve healthcare management. Amount of data from different sources: wearables, patients' platforms, etc. are gathered to provide a better healthcare service at less cost. But big data analytics is not only applied to healthcare but also to help to identify trends and patterns to treat patients' disease.
- Retail: Big data and analytics provide the insights needed to track consumers behaviours and give competitive advantage to retail companies, which anticipate consumers' needs.
- Banking: Big Data and analytics are being used in the banking sector for predictive analytics, natural language processing to help customers in transactions and fight against fraud.
- Transportation: Big data and analytics help transportation companies to anticipate unexpected transportation problems and optimize the management of the transportation fleet.

## 4.2. Stakeholder Analysis and Value Chain

BigDataStack key offering provides solutions and adds value to a wide range of stakeholders working in different sectors and to end-users working in Big Data domains. BigDataStack

---

<sup>6</sup> [http://bdva.eu/sites/default/files/EuropeanBigDataValuePartnership\\_SRIA\\_\\_v3\\_0.pdf#overlay-context=downloads%3Fq%3Ddownloads](http://bdva.eu/sites/default/files/EuropeanBigDataValuePartnership_SRIA__v3_0.pdf#overlay-context=downloads%3Fq%3Ddownloads)

<sup>7</sup> <http://bdva.eu>

outcomes will provide tools and services to facilitate the job to a series of stakeholders and enable them to provide to end-users improvement services using Big Data tools.

We can make a first classification of the stakeholders as follow:

- Stakeholders working in Big Data domains that provide products and services, whom BigDataStack key offerings can facilitate their job (**Supply Side**):
  - Infrastructure providers: BigDataStack solutions provides the means (data-driven Infrastructure/cluster management) to facilitate big data needs.
  - Data Providers: Data-as-a-service solution provided by BigDataStack project allowing them to offer data quality assessment, modelled, stored and analysed data.
  - Application Providers: BigDataStack solutions allowing them to perform data-intensive applications with specific performance and guarantees.
  - Data Practitioners: BigDataStack Data Toolkit and visualization environment allowing them to develop their own Big data algorithms and offer them through BigDataStack deployments
- Stakeholders that hire the services or products delivered by the Supply Side, and provide solutions to end-users, acting as a second-step entities and taking advantage of the Data-as-a-service model or data-intensive application provisioning. These stakeholders are: Infrastructure Brokers, Data aggregators & Data (re-)sellers and Marketplace owners.
- 
- End-users which demand Big Data applications for their businesses (**Demand Side**):
  - Citizens: BigDataStack will allow them to use Big data solutions with levels of quality.
  - SMEs and big industries: BigDataStack will be used to facilitate their solutions deployments, data needs, and allowing them to expand their business portfolio.
  - Public Organizations: Using BigDataStack for their internal data needs purpose optimizing the operations across the entire data path.
  - Entrepreneurs: Allowing them to deploy and offer data intensive applications with Open Source solutions.
  - Decision Makers: Drive business decisions based on insights provided by using BigDataStack solutions

The relation among the different stakeholders is depicted in the Value Chain figure below:

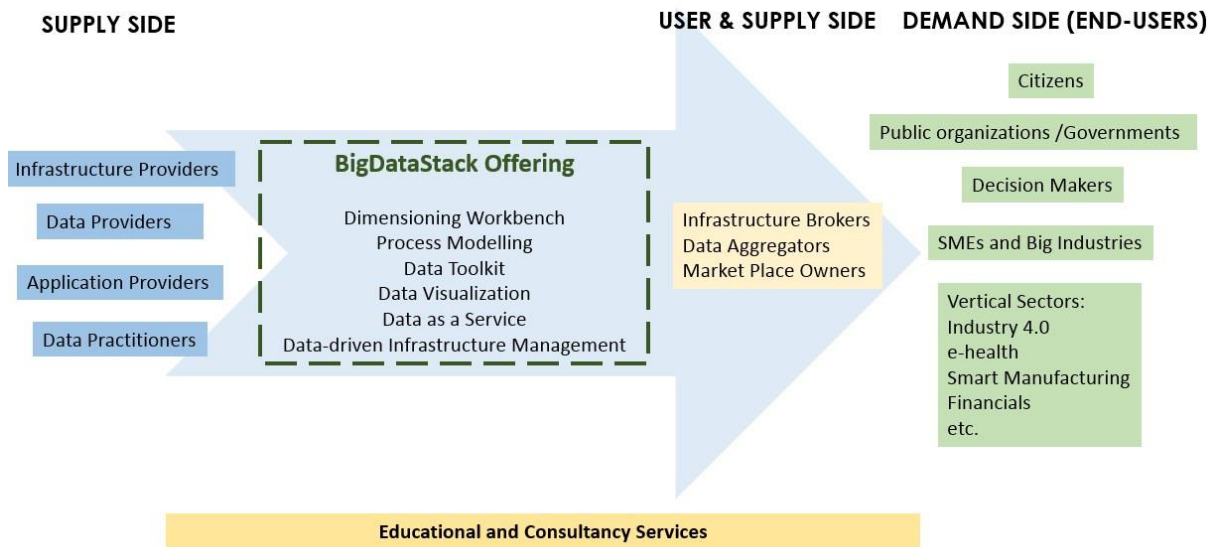


Figure 1. BigDataStack Value Chain

### 4.3. Initial SWOT Analysis

SWOT Analysis<sup>8</sup> is a useful technique for understanding the Strengths and Weaknesses of a technology/product, and for identifying both its Opportunities and the Threats that it faces.

Strengths and weakness are frequently internally-related, while opportunities and threats commonly focus on the external environment. The name is an acronym for the four parameters this technique examines:

- Strengths: characteristics of the business or project that give it a relative advantage.
- Weaknesses: characteristics of the business that place the business or project at a relative disadvantage .
- Opportunities: elements in the environment that the business or project could exploit
- Threats: elements in the environment that could cause trouble for the business or project.

For BigDataStack outcomes and considering the environment described in the Market context, the initial SWOT is as follow:

	Helpful	Harmful
	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
Internal	Performant end-to-end data-oriented solution	Integration complexity into an overall solution
	Partners' expertise in Big Data Technologies with a verified background and baseline products	Many proprietary APIs for data acquisition

<sup>8</sup> [https://www.mindtools.com/pages/article/newTMC\\_05.htm](https://www.mindtools.com/pages/article/newTMC_05.htm)



	European Open Source Initiative as a specific means towards open-source exploitation	Established infrastructures management systems vendors
	Strategic leading position of BigDataStack partners in target communities and projects (ATOS in BDVA, TRUST-IT in EOSC, etc.)	
	<b>OPPORTUNITIES</b>	<b>THREATS</b>
<b>External</b>	Increasing number of data operations and data-intensive applications	3 <sup>rd</sup> Party closed source components used by application providers
	Wide IoT/CPS deployments generating big amounts of data to process from different sources	Lack of established standards for data access and representation
	Stakeholders seeking new ways to enhance competitiveness and sustainability and requiring implementation of Big Data solutions to their businesses	Legal Frameworks governing data operations
	Business Models requiring Data as a Service, to clean, model, store and analyze data	Emergence of competitive solutions from SMEs and Big Vendors
	Open-Source projects on the rise	

Table 2. Initial BigDataStack SWOT Analysis

#### 4.4. BigDataStack Value Proposition

The value proposition of the BigDataStack project is to deliver “a complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that address the emerging needs of data operations and applications”, additionally BigDataStack introduces the paradigm of a “new frontrunner data-driven architecture and system ensuring that infrastructure management will be fully efficient and optimized for data operations and data-intensive applications”<sup>9</sup>.

The BigDataStack approach puts emphasis in current Big Data challenges:

- Address the need for analytics on both data in flight and at rest and optimize the deployment and execution of big data applications and operations through **data-driven management of the underlying resources**.

<sup>9</sup> BigDataStack Grant Agreement

- Provide a **seamless data analytics framework** to analyse data across multiple data stores and locations by eliminating the need to perform application-level adaptations to address (and query) the various heterogeneous underlying stores.
- Address the complete data path operations, providing **Data-as-a-Service** to address the complete data path (including data quality assessment, efficient distributed storage, optimized performance through data skipping, and data analytics)
- Facilitate the usability and extensibility by delivering a fully exploitable and open source solution through a **Data Toolkit** that enables new analytics (and analytics pipelines) to be ingested and executed in the underlying infrastructure.
- Deliver an innovative **Process Modelling** framework, accompanied with process analytics / mining and process mapping techniques to address the needs of business analysts.

## 5 Potential Business Models for BigDataStack

---

At this stage of the project, three business models have been identified to exploit the BigDataStack outcomes for different targets and purpose and not just for software components developed but also for the know-how acquired within the project.

These business models are:

- **BigDataStack-as-a-Service:** BigDataStack commercialization model to offer Big Data products and services.
- **Educational Business Model:** Applies to educational purpose and knowledge transfer in general.
- **Consultancy Business Model:** Provides Big Data Consultancy Services in general, and for application of Big Data to specific sectors such as Shipping, Financing and Consumers.

To explain the identified business models, we will use the Business Model Canvas Methodology<sup>10</sup>, that will help us to easily understand business models and define their business components.

Next, we briefly describe the identified business models, which will be evolved during the second phase of the project. It is planned to hold an Exploitation Workshop among partners to discuss all the business aspects of the different business models and select the best-suited business models according to the partners' profiles and preferences.

These business models will be more extensively explained in subsequent deliverable.

### 5.1. BigDataStack commercialization Business Model

The business model for the BigDataStack key offering is planned to be in two different ways:

- **Offer turn-key big data management and analytics solutions.** These solutions will be offered to use case partners as well as other customers to which BigDataStack outcomes can be applied. This Business Model is targeted to Demand Side Stakeholder.
- **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 in their solutions. This Business model is targeted to Supply Side stakeholders.

Next, we briefly describe the Business Model Canvas for the BigDataStack commercialization business models.

---

<sup>10</sup> <https://www.strategyzer.com/canvas/business-model-canvas>

<p><b>Key Partners</b> Infrastructure Providers  Service providers (Cloud Providers)</p>	<p><b>Key Activities</b> Operate the complete BigDataStack components to optimize infrastructure management  Provide the different BigDataStack key offering in an isolated way  Support and maintenance services</p>	<p><b>Value Proposition</b> Complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that address the emerging needs of data operations and applications  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><b>Customer Relationships</b> Webinars and training courses  Incorporation of new use cases of different sectors  Software Update  Strategic partnerships with other Big Data providers (SMEs, Research Community, Industry)</p>	<p><b>Customer Segments</b> Citizens  SMEs and Big Industry  Public Organizations  Decision Makers  Industrial Sectors with Big Data needs to improve their businesses performance (Industry 4.0, e-Health, etc.)  Infrastructure and solution providers  Application Providers  Data providers  Data practitioners</p>
<p><b>Cost Structure</b> Marketing Maintenance of software and infrastructures Personal costs Server Costs Infrastructure providers costs Variable costs, depending on the scenario</p>		<p><b>Revenue Streams</b> Pay-as-you-go fee Explore other payment agreement with specific customers</p>		

Table 3. BigDataStack commercialization Business Model Canvas

## 5.2. Educational Business Model

This business model applies to educational purpose and knowledge transfer of the BigDataStack outcomes and is mainly targeted to Academia and Research Communities.

BigDataStack partners, such as UPRC, UPM and GLA, can leverage the experience and know-how acquired within the project to deliver educational services.

<b>Key Partners</b> Research Communities  Technology & Infrastructure Providers	<b>Key Activities</b> Share knowledge and experience acquired during the project  Continue with R&D in areas related to BigDataStack topics	<b>Value Proposition</b> Complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that address the emerging needs of data operations and applications  Frontrunner data-driven architecture and system ensuring that infrastructure management will be fully efficient and optimized for data operations and data-intensive applications	<b>Customer Relationships</b> Community building  Webinars and training courses  Support and maintenance  Knowledge Transfer	<b>Customer Segments</b> Research Communities  Educational Institutions  Innovative SMEs developing big data and intensive applications
	<b>Key Resources</b> Consortium experience and Know-how  ICT Infrastructures		<b>Channels</b> BigDataStack project Website  Project Dissemination (Conferences, events, scientific publications)  Other related EU research projects and initiatives (BDVA)	
<b>Cost Structure</b> Promotion costs Maintenance of the platform and infrastructures Personal costs Server Costs Infrastructure providers costs			<b>Revenue Streams</b> Payment of educational services such as: Conference and events Educational programs and courses Research lines	

Table 4. Educational Business Model Canvas

### 5.3. Consultancy Business Model

This business model is targeting industry in general to provide guidance and know-how in the implementation of Big Data technologies to their businesses, as well as SMEs working in Big Data domains on how they can improve their commercial solutions with the integration of BigDataStack technologies or services provided in the BigDataStack commercialization model.

<b>Key Partners</b> Technology & Infrastructure Providers	<b>Key Activities</b> Share knowledge and experience acquired during the project  Consultancy of BigDataStack specific key offering	<b>Value Proposition</b>	<b>Customer Relationships</b> Support and maintenance  Knowledge Transfer	<b>Customer Segments</b> Innovative SMEs developing big data and intensive applications  Industry in all sectors
--	--	--------------------------	--	---

	<p><b>Key Resources</b>          Consortium experience and Know-how</p> <p>ICT Infrastructures</p>		<p><b>Channels</b>          BigDataStack project Website</p> <p>Project Dissemination (Conferences, events, scientific publications)</p> <p>Other related EU research projects and initiatives (BDVA)</p>	
<p><b>Cost Structure</b>          Promotion costs          Maintenance of the platform and infrastructures          Personal costs          Server Costs          Infrastructure providers costs</p>		<p><b>Revenue Streams</b>          Payment of services such as:          Conference and events          Educational programs and courses          Research lines</p>		

Table 5. Consultancy Business Model Canvas

## 6 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 is depicted in the next figure.

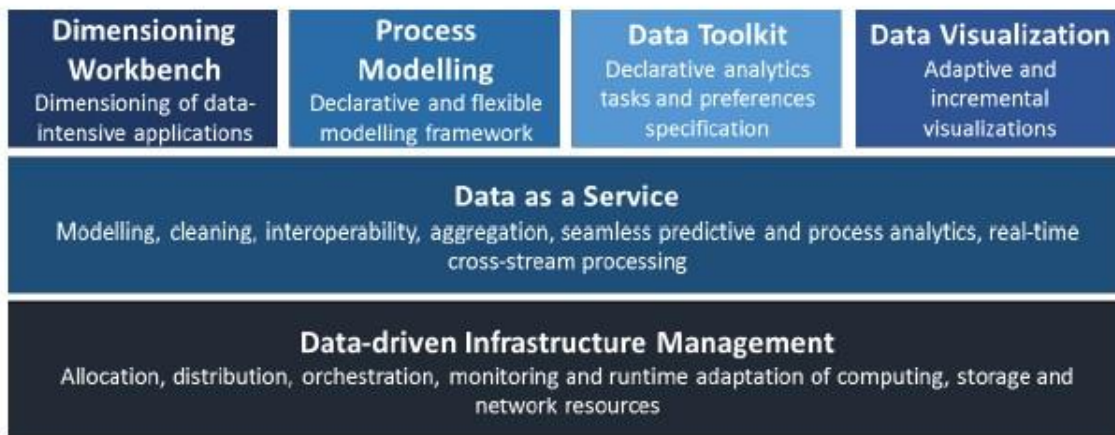


Figure 2. BigDataStack Key offering<sup>11</sup>

In this section we describe not just these 6 blocks that made up the BigDataStack offering, but also all the subcomponents developed by the different partners.

All partners have filled in a Components’ Template elaborated by the Exploitation Team to collect relevant information about the SW components: technical description, Licenses analysis, competitors, SWOT analysis (strengths, weaknesses, opportunities and threats), and exploitation possibilities.

This preliminary information on SW project outcomes collected in the first period of the project will be analysed in depth to explore realistic exploitation routes as stand-alone software for the SW components or cluster of them.

Below, a summary list of SW components identified in the first period and TRL expected at the end of the project.

	BigDataStack Assets	TRL1	TRL2	TRL3	TRL4	TRL5	TRL6	TRL7	TRL8	TRL9
IBM	Data Skipping									
NEC	Dynamic Orchestration									
RHT	Network Policy support at Kuryr									
	Kuryr Cluster Network Operator									
	NvME-mdev Kernel driver									
ATOS	QoS Evaluation									

<sup>11</sup> D2.4 Conceptual Model and Reference Architecture





GLA	Deployment						
	Recommendation Service						X

Table 7. Involvement of SW components in BigDataStack Key offering

## 6.1. Analysis of Exploitable Assets

### 6.1.1. SQL Data Skipping

- Main partner involved: IBM
- Description of the component

Name of the Component: Data Skipping	
Short description of the component	SQL queries are a widespread technique to analyse datasets. In BigDataStack, data skipping for SQL queries has been further researched and developed. This technology is relevant when the dataset resides in an Object Store. This research has already been contributed to the IBM SQL Query service for now as a closed beta.
Input of the component	An SQL query against a given data set which is stored in an Object Store.
Output of the component	The result of the query
Describe the target customer segments/user groups of the component	Since our technology is already in advanced stage of integration with the IBM SQL Query service, the customers of this service are our natural customers.
Key features and benefits of the component	This technology reduces the size of the data scanned for answering the SQL query. Thus it a) it reduces the price to be paid by the customer b) accelerates the query
Initial TRL at the beginning of the project	IBM was already working at integrating with the IBM SQL Query Service, however this was for the basic data skipping technology. BigDataStack permitted to enhance in a much non-trivial way the basic technology. These enhancements have been integrated within the IBM SQL Query Service.
TRL to be reached at the end of the project	IBM hopes to move from beta to full productization (TRL8/TRL9)
Essential information for potential users (requirements, use restrictions, etc.)	Prior to using this technology, the dataset has to be indexed as explained in the documentation of the IBM SQL Query Service.

<b>Standards involved in the development of the component</b>	None
<b>What BigDataStack Use Case is the component related with?</b>	For now, the Data Skipping technology has been tested on <i>Real time Shipping</i> , however the <i>Connected Consumer</i> seems to be an even better use case for it. It is also planned to apply it in the third use case (insurance).
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	No

Table 8. SQL Data Skipping description

- **Competitors and trends in the market**

<b>Competitor1: Amazon Athena<sup>12</sup></b>	
<b>Short description</b>	Amazon Athena
<b>Developed by</b>	Amazon
<b>Type of License: Open Source or proprietary</b>	Proprietary product
<b>Business Models (free, license, pay per use, etc.)</b>	Pay per use (function of the volume of the scanned data)
<b>Explain what the bonus (features, status, business model, etc.) of the component in relation to the competitor</b>	Our skipping technology has the potential to scan less data for a given data set and query, thus to be more attractive to user
<b>Explain what the minus of your component in relation to the competitor is</b>	Maybe this component is not as speedy as Athena

Table 9. SQL Data Skipping competitors

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
<b>Internal</b>	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Most effective technology in term of skipping performance	In development phase yet
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	Reinforce the features that differentiates from the competitors, and promote it among the OS communities	We come to market much later than competitors (e.g., AWS Athena)

<sup>12</sup> <https://aws.amazon.com/athena/>

Table 10. SQL Data Skipping SWOT

- Preliminary Exploitation route

Has the component synergies with other component regarding exploitation?	This component is stand-alone and does not depends on other components.
Is it expected an exploitation strategy for the component alone or along with another one?	Alone
Has the component synergies with other component regarding exploitation?	This component can also be used within the Seamless component.
Is IBM interested in exploiting the component? How?	Yes. This technology naturally finds its place in the IBM SQL Query Service and maybe other IBM services/products.

Table 11. SQL Data Skipping preliminary exploitation route

## 6.1.2. Dynamic Orchestration

- Main Partner involved: NEC
- Description of the component

Name of the Component: Dynamic Orchestrator	
Short description of the component	The Dynamic Orchestrator triggers the redeployment of applications during runtime to ensure they comply with their Service Level Objectives (SLOs). For this task it is being used a Reinforced Learning-based approach which can operate efficiently, with a light overhead for the overall system.
Input of the component	System and application metrics, SLOs specification and metric, deployment configurations
Output of the component	Redeployment changes to be performed (if any)
Describe the target customer segments/user groups of the component	This component can be used by all developers that utilize BigDataStack.
Key features and benefits of the component	The Dynamic Orchestrator triggers redeployment mechanisms during runtime in order to adapt applications and services to the changing context and ensure they comply with their requirements.

	<p>Features: flexible, learns on its own, fast reaction</p> <p>Benefits: ensures applications keep up with their requirements in changing environments</p>
<b>Initial TRL at the beginning of the project</b>	TRL1
<b>TRL to be reached at the end of the project</b>	TRL7
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	The dynamic orchestrator is implemented in a logic based on machine learning. Because of this, it needs time to learn about its environment (application/service and deployment environment) in order to learn how to behave correctly, i.e. the behavior when the application has just started might not be the optimal for a short period of time.
<b>Standards involved in the development of the component</b>	The development of the component is aligned with the latest recommendations from the BDVA and AIOTI <sup>13</sup> .
<b>What BigDataStack Use Case is the component related with?</b>	This component is related to all three use cases to be implemented within BigDataStack. However, until now It has been focused on the use case of <i>Connected Consumer</i> , in which you have a varying load of inputs to be processed by the application and because of this, it is necessary to monitor that SLOs are met during runtime and if not, trigger the corresponding deployment changes to improve the application's performance.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	It is expected that the results of this component will innovate in data-driven infrastructure management, which are closely related to IoT and Big Data solutions. Because of this, the results of this component will contribute to the Data Management, Data Processing Architectures, Data Analytics, and Data Visualization and User Interaction activities within BigData PPP as well as the evolution of BDVAs Strategic Research and Innovation Agenda. In addition, the

<sup>13</sup> <https://aioti.eu/>

	<p>outcomes of the component will provide a significant evolution for data management in IoT, contributing to AIOTI standardization strategies. Finally, it is planned on using the Dynamic Orchestrator in FogFlow<sup>14</sup> to orchestrate data flows from edge nodes to data centers, using and complementing NGSI 9 &amp; 10 specifications.</p>
--	---

Table 12. Dynamic Orchestrator description

- **Competitors and trends in the market**

<b>Competitor 1: Kubernetes Autoscaler<sup>15</sup></b>	
<b>Short description</b>	This is a feature of Kubernetes, which increases or decreases the number of pods used by a running application according to the current usage of pods.
<b>Developed by</b>	Kubernetes
<b>Type of License: Open Source or proprietary</b>	Open Source
<b>Business Models (free, license, pay per use, etc.)</b>	Free
<b>The bonus (features, status, business model, etc.) of the component in relation to the competitor</b>	<p>While the autoscaler in Kubernetes uses the current usage of pods in order to scale in or out, the Dynamic Orchestrator component uses more system metrics and more importantly, SLOs metrics, in order to decide when it is necessary to change the current deployment. This component's approach provides a higher abstraction for developers, making it easier to ensure SLOs and application requirements.</p> <p>Moreover, the redeployment action spectrum of the Dynamic Orchestrator includes more actions than just scaling in or out, which makes this component more flexible.</p>
<b>The minus of the component in relation to the competitor</b>	Because we are using a learning algorithm instead of a heuristic as the one used by Kubernetes, the decisions taken when the application has just started its execution might not be the optimal ones. However, as

<sup>14</sup> <https://fogflow.readthedocs.io/en/latest/>

<sup>15</sup> <https://github.com/kubernetes/autoscaler/tree/master/cluster-autoscaler>

	time passes by, these decisions become more and more accurate, and thanks to this learning approach, the Dynamic Orchestrator is flexible to work with different application requirements and SLOs.
--	---

Table 13. Dynamic Orchestrator competitor 1

<b>Competitor 2: AWS Autoscaling<sup>16</sup></b>	
<b>Short description</b>	AWS Auto Scaling monitors applications and automatically adjusts their capacity to maintain steady, predictable performance at the lowest possible cost. It also provides an interface to build scaling plans to scale processing nodes, database indexes and tables.
<b>Developed by</b>	Amazon
<b>Type of License: Open Source or proprietary</b>	Proprietary
<b>Business Models (free, license, pay per use, etc.)</b>	Free to use when using AWS (which are pay per use)
<b>the bonus (features, status, business model, etc.) of the component in relation to the competitor</b>	The Dynamic Orchestrator uses more system metrics to decide when to scale, and more importantly, it also uses SLOs metrics in order to decide when it is necessary to change the current deployment. Our approach provides a higher abstraction for developers, making it easier to ensure SLOs and application requirements. Moreover, the redeployment action spectrum of the Dynamic Orchestrator includes more actions than just scaling in or out, which makes our component more flexible.
<b>the minus of your component in relation to the competitor</b>	Again, AWS Autoscaling uses heuristics for deciding when to scale, which means it will always have the same performance. However, this adds a burden for developers because they have to define the scaling plan for their application or service. In addition, even though the Dynamic Orchestrator does not provide database scaling, the database scaling is provided by LeanXcale in BigDataStack.

<sup>16</sup> <https://aws.amazon.com/autoscaling/>

Table 14. Dynamic Orchestrator Competitor 2

- **SWOT Analysis**

	Helpful	Harmful
<b>Internal</b>	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	<p>The application of RL to orchestrate applications and data services is a novel approach which, in our knowledge, has not been researched by any other party yet. RL provides more flexibility in terms of SLOs and requirements to ensure, as well as a more customized performance for each application that run on BigDataStack.</p> <p>The application of this AI approach is possible thanks to the already existing knowledge in AI and Orchestration techniques we count with at NEC, which is a competitive advantage of our component.</p>	<p>The use of RL for autoscaling applications and services is a new approach and because of this it has to be properly tested with different use cases and in several situations. In addition, since we are dealing with a learning approach, the performance of the Dynamic Orchestrator will not be very high when the application has just started and then it will improve over time. This is a challenge we are researching and for which we are considering different options in order to provide a high performance from the beginning.</p>
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	<p>Cloud computing is currently widely adopted and used for most applications and services nowadays. In this scenario, big data applications are particularly challenging because the amounts of data that need to be processed vary greatly from application to application and also with time. In addition, these applications usually need low response times in order to provide a useful service, such as Augmented Reality applications. In this scenario, developers face the decision of overprovisioning their applications, which means the running costs of the applications will be high or risking not to provide a satisfying user experience. Thanks to the Dynamic Orchestration however, developers that use BigDataStack can enjoy the best of both worlds: costs are kept as low as possible while providing a satisfying experience for all users.</p>	<p>The first threat is related to other competitors that are already working on dynamically orchestrating applications and services, such as the ones named in section 3. However, the Dynamic Orchestrator, as well as its competitors, and not sold as an independent service but as part of a platform, and because of this we believe that combined with all the other features offered by the BigDataStack, as well as FogFlow, our product will be superior than our competitors’.</p> <p>Another threat is the progress of new technologies that might decrease the costs of cloud resources, creating the opportunity for all applications to be able to overprovision for a low cost. However, recent trends indicate that even though technologies are advancing fast, data is growing exponentially and surpassing technological advances, proof of this</p>

	are trends such as Edge and Fog computing which have emerged as a response to limitations of the Cloud computing paradigm. Because of this, we foresee that managing resources in Cloud computing will continue to be an important topic in the future.
--	---

Table 15. Dynamic Orchestrator SWOT Analysis

- **Preliminary Exploitation route**

<b>Has the component synergies with other component regarding exploitation?</b>	<p>The Dynamic Orchestrator interacts with the QoS, the 3ME and the ADS-Ranker, however the Dynamic Orchestrator has been designed so it can interact on its own with different systems.</p> <p>The QoS and the 3ME preprocess system metrics and SLOs metrics and send them to the Dynamic Orchestrator, which uses it to determine the current status and decide if a change in deployment is needed and if so, what change should it be. If a change is needed, a message is sent to the ADS-Ranker, which will check the current available deployment configurations and trigger the redeployment.</p>
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	<p>It is planned on using the Dynamic Orchestrator along with FogFlow, our open source fog execution framework. The Dynamic Orchestrator will optimize the deployment of big data processing flows from edge nodes to data centers and the Cloud. In this scenario, our component can be deployed on its own as part of FogFlow, for which we have a different setting for system metrics, which are already given by FogFlow and changes in the deployment are also made by FogFlow.</p>
<b>Is NEC interested in exploiting the component? How?</b>	<p>As stated in the previous point, we plan to use the Dynamic Orchestrator for dynamically adapting services during runtime on FogFlow. FogFlow is open source and currently a Full GE of FIWARE<sup>17</sup>.</p>

Table 16. Dynamic Orchestrator preliminary exploitation route

<sup>17</sup> <https://www.fiware.org/>



### 6.1.3. Network Policy support at Kuryr

- Main partner involved: RHT
- Description of the component

Name of the Component: Network Policy support at Kuryr	
Short description of the component	Network Policy support at Kuryr
Input of the component	Network Policy Kubernetes Object
Output of the component	Network policy being enforced through OpenStack security groups.
Describe the target customer segments/user groups of the component	Customers willing to use kuryr to avoid double encapsulation problem that have fine grain network access policies to their different applications components
Key features and benefits of the component	It does not apply; network policy support is the feature.
Initial TRL at the beginning of the project	-
TRL to be reached at the end of the project	Already merged upstream, TRL7, expected to be TRL9 at the end of the project.
Essential information for potential users (requirements, use restrictions, etc.)	N/A
Standards involved in the development of the component	Open Source, OpenStack upstream community.
What BigDataStack Use Case is the component related with?	Network Traffic shaping/controlling.
Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?	It is already part of Kuryr OpenStack project <sup>18</sup> .

Table 17. Network Policy support at Kuryr description

- Competitors and trends in the market

All the functionalities (features, operators and drivers) are open sourced and developed **upstream** within their related project communities. This means with cooperation with other companies (i.e., providing feedback, reviews, requirement).

<sup>18</sup> <https://wiki.openstack.org/wiki/Kuryr>

## 6.1.4. Kuryr Cluster Network Operator

- Main partner involved: RHT
- Description of the component

Name of the Component: Kuryr Cluster Network Operator	
Short description of the component	Kuryr Cluster Network Operator
Input of the component	Cluster and Operator config yamls ( <a href="https://github.com/openshift/cluster-network-operator#configuration-objects">https://github.com/openshift/cluster-network-operator#configuration-objects</a> )
Output of the component	Kuryr SDN installed and configured on the OpenShift cluster.
Describe the target customer segments/user groups of the component	Customers willing to use kuryr to avoid double encapsulation problem when running OpenShift on top of OpenStack
Key features and benefits of the component	Easy deployment and management of KuryrSDN on OpenShift 4.X cluster on top of OpenStack VMs
Initial TRL at the beginning of the project	-
TRL to be reached at the end of the project	TRL 9
Essential information for potential users (requirements, use restrictions, etc.)	N/A
Standards involved in the development of the component	Open Source, modifications made to different upstream projects, such as Cluster API, Cluster Network Operator, OpenShift/Installer, Gophercloud, Terraform, etc.
What BigDataStack Use Case is the component related with?	Infrastructure set up: Cluster Management
Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?	It is being developed as part of the upstream Open Source Community around Kubernetes/OpenShift

Table 18. Kuryr Cluster Network Operator description

- Competitors and trends in the market

All the functionalities (features, operators and drivers) are open sourced and developed **upstream** within their related project communities. This means with cooperation with other companies (i.e., providing feedback, reviews, requirement).

### 6.1.5.NvME-mdev Kernel driver

- Main partner involved: RHT
- Description of the component

Name of the Component: NVMe-mdev kernel driver	
Short description of the component developed by your organization	NVMe-mdev kernel driver to have fast storage virtualization
Input of the component	Users use out of the box NVMe drivers
Output of the component	Virtual NVMe driver provided to the VMs
Describe the target customer segments/user groups of your component	Kernel module which implements a virtual NVMe device which VMs can use as a drop-in replacement for PCI assignment
Key features and benefits of your component	Allows sharing of NVMe device between server VMs. Simple configuration. Roughly the same performance as SPDK (95% of native)
Initial TRL at the beginning of the project	TRL1
TRL to be reached at the end of the project	TRL7
Essential information for potential users (requirements, use restrictions, etc.)	Being proposed upstream for the Linux kernel. Unclear if it will be accepted at its current form.
Standards involved in the development of your component	Following Linux kernel upstream mechanisms
What BigDataStack Use Case is your component related with?	All components accessing local disk on the VMs may benefit from it
Do you envision that BigDataStack results or your components can influence in some Standardization foundation or initiative?	If merged on the Linux kernel it will be available on every Linux distribution with a recent kernel version

Table 19. NVMe-mdev kernel driver description

- Competitors and trends in the market

All the functionalities (features, operators and drivers) are open sourced and developed **upstream** within their related project communities. This means with cooperation with other companies (i.e., providing feedback, reviews, requirement)

- **Preliminary Exploitation routes for components developed by RHT**

After the feature/operator/driver are merged upstream, Red Hat exploitation strategy is to make them part of its portfolio by making a product out of the upstream project. To do that, extra QA tasks will be performed as well as an evaluation of the possible interest and use cases where they may be required by our customers.

### 6.1.6. QoS Evaluation

- **Main partner involved: ATOS**
- **Description of the component**

<b>Name of the Component: QoS Evaluator</b>	
<b>Short description of the component</b>	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.
<b>Input of the component</b>	The input of the system is twofold: <ul style="list-style-type: none"> <li>• Constrains or SLO (Service Level Objective) from the Dynamic Orchestrator component.</li> <li>• Metrics collected by the Triple Monitoring from a monitoring tool, used for evaluating the SLO fulfilment.</li> </ul>
<b>Output of the component</b>	QoS violations are published in a queue system where the Triple Monitoring and the Dynamic Orchestrator components are subscribed; the former to compute and publish violation-based metrics; the latter to make decisions as to how to adapt the

	deployment of the application or service to improve its performance.
<b>Describe the target customer segments/user groups of the component</b>	Developers who want to add QoS evaluation capability to Big Data software applications and systems.
<b>Key features and benefits of the component</b>	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).
<b>Initial TRL at the beginning of the project</b>	TRL7
<b>TRL to be reached at the end of the project</b>	TRL8
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	The QoS Evaluator component deploys as a micro-service in a single Docker container. For this Project, the QoS Evaluator has been successfully integrated with the Prometheus-based monitoring system to query performance metrics and to provide QoS violation metrics in return (to be recorded and published together). The component provides two main interfaces to developers, who need to adapt in order to integrate with their software: The application Observer, to subscribe to violation notifications; and the Monitoring adapter to feed the QoS Evaluator with the metrics to evaluate.
<b>Standards involved in the development of the component</b>	The QoS model is loosely inspired in the WS-Agreement (Web Services Agreement

	Specification) specification from the Open Grid Forum (OGF) <sup>19</sup>
<b>What BigDataStack Use Case is the component related with?</b>	The three BigDataStack use cases will make use of the QoS evaluation, as all of them need the Data-Driven Infrastructure Management capability of BigDataStack to ensure specific performance constraints in the operation of their analytics applications and processes.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	No

Table 20. QoS Evaluator description

- **Competitors and trends in the market**

<b>Competitor 1: Prometheus Alerting<sup>20</sup></b>	
<b>Short description</b>	Prometheus is a system monitoring and alerting toolkit. The main component is the Prometheus server, which scrapes and stores time series data (metrics); it also generates alerts based on metrics and sends it to the Alert Manager, which sends out notifications via methods such as email, PagerDuty or HipChat.
<b>Developed by</b>	Prometheus community <sup>21</sup> under the Cloud Native Computing Foundation (as a hosted CNCF project since 2016) and The Linux Foundation.
<b>Type of License: Open Source or proprietary</b>	Open Source (Apache v2 <sup>22</sup> )
<b>Business Models (free, license, pay per use, etc.)</b>	Free
<b>The bonus (features, status, business model, etc.) of the component in relation to competitors</b>	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 provide and API to create and manage the life cycle of the alerts, including

<sup>19</sup> Open Grid Forum (2011): “Web Services Agreement Specification (WS-Agreement),” Oct. 10, 2011.

<http://ogf.org/documents/GFD.192.pdf><https://www.ogf.org/ogf/doku.php>

<sup>20</sup> Prometheus Alerting – Overview <https://prometheus.io/docs/alerting/overview/>

<sup>21</sup> Prometheus Community: <https://prometheus.io/community/>

<sup>22</sup> Apache 2: <https://github.com/prometheus/prometheus/blob/master/LICENSE>

	<p>the specification of their lifespan and/or conditions for activation, deactivation and depreciation.</p> <p>Also, penalization information related to the violation may be attached and computed (i.e. economic costs).</p>
<p><b>The minus of the component in relation to the competitor</b></p>	<p>Alerting rules in Prometheus allow you to define alert conditions based on PromQL (Prometheus Query Language), a powerful and convenient DSL (Domain Specific Language) to query metrics and express logical conditions upon them. In the case of QoS Evaluator, this language is a lower level C-like logical expressions language<sup>23</sup>.</p> <p>Tightly integrated so packaged and deployed by default with Prometheus; therefore, it is easier to deploy, operate and use if you use Prometheus as your monitoring solution. In the case of QoS Evaluator, it has to be deployed and operated as an independent microservice.</p> <p>Backed by a large open source community and developed under important organizations such as the Cloud Native Computing Foundation (CNCF) and The Linux Foundation.</p>

Table 21. QoS competitor 1

Competitor 2: Instrumental Intelligent Alerts <sup>24</sup>	
<p><b>Short description</b></p>	<p>Instrumental provides automatic data collection from diverse IT infrastructure components and well-known applications and data services, real-time visualizations, and intelligent alerts. Instrumental offers high performance monitoring and alerting capabilities, being able to process over a million data points every second. Therefore, alerts are triggered in real-time (i.e., in the order or milliseconds) and so the Intelligent Alerts capability fits well to real-time applications and services</p>
<p><b>Developed by</b></p>	<p>Instrumental<sup>25</sup></p>

<sup>23</sup> Govaluate: <https://github.com/Knetic/govaluate/blob/master/MANUAL.md>

<sup>24</sup> Instrumental Intelligent Alerts – Overview: <https://instrumentalapp.com/features/alerts>

<sup>25</sup> [instrumentalapp.com](https://instrumentalapp.com)

<b>Type of License: Open Source or proprietary</b>	Proprietary
<b>Business Models (free, license, pay per use, etc.)</b>	Pay per use: Free for up to 500 metrics, then \$0.10/month per metric.
<b>The bonus (features, status, business model, etc.) of the component in relation to competitors</b>	<p>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.</p> <p>Intelligent Alerts are tightly tied to the Instrumental monitoring solution and can't be distributed not used separately</p>
<b>The minus of the component in relation to the competitor</b>	<p>Anturis (including Alerts &amp; Notifications) is a mature monitoring and alerting solution, with important customer already depending on its capabilities in production-grade business applications.</p> <p>Plenty of integrations with IT infrastructure components and well-known application to get performance metrics from them.</p> <p>The Query Language available in Instrumental to evaluate stored metrics is a DSL (Domain-Specific Language) which lets you to transform, aggregate and time-shift data – in real-time.</p>

Table 22. QoS Evaluator Competitor 2

<b>Competitor 3: Anturis Alerts &amp; Notifications<sup>26</sup></b>	
<b>Short description</b>	Anturis is a cloud-based (SaaS) monitoring service for IT infrastructure (servers and applications) and websites. It provides a flexible capability to generate alerts and send notifications (e.g. through e-mail, SMS, and voice). Some of the alerting features supported are hierarchy organization, severity levels, warning and error thresholds, roles and responsibilities in incident management.
<b>Developed by</b>	Anturis <sup>27</sup>
<b>Type of License: Open Source or proprietary</b>	Proprietary
<b>Business Models (free, license, pay per use, etc.)</b>	Free plan, or License (from \$9.50 monthly)

<sup>26</sup> Anturis Features – Alerts & Notifications: <https://anturis.com/features/>

<sup>27</sup> [anturis.com](https://anturis.com)



<p><b>The bonus (features, status, business model, etc.) of your component in relation to competitors</b></p>	<p>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.</p> <p>Alerts &amp; Notifications are tightly tied to the Anturis monitoring solution and can't be distributed not used separately.</p> <p>Anturis Alerting rules are quite simple, based on thresholds, in comparison to the complex logic than be specified in QoS Evaluator rules.</p>
<p><b>The minus of your component in relation to the competitor</b></p>	<p>Anturis (including Alerts &amp; Notifications) is a mature monitoring and alerting solution, with important customer already depending on its capabilities in production-grade business applications.</p>

Table 23. QoS Competitor 3

- **SWOT Analysis**

		Helpful	Harmful
		STRENGTHS	WEAKNESSES
Internal	<p>SLA driven approach to application and infrastructure evaluation and alerting.</p> <p>Specially designed for the needs and constrains of cloud, fog and edge computing applications.</p>	<p>Few available integrations with notification frameworks and monitoring systems.</p> <p>Not backed by a community of developers and/or users.</p> <p>Distributed, deployed and operated separately from other open source monitoring and alerting solutions.</p>	
	External	<p><b>OPPORTUNITIES</b></p>	<p><b>THREATS</b></p>
<p>Increasing interest in ensuring QoS in big data analytics applications.</p>		<p>Popular open-source and commercial monitoring and alerting solutions provide similar components.</p>	

Table 24. QoS Evaluator SWOT Analysis

- **Preliminary Exploitation route**

<p><b>Has the component synergies with other component regarding exploitation?</b></p>	<p><u>Triple Monitoring (TM)</u>: This Prometheus-based component provides the QoS Evaluator with performance metrics at three levels (application, data and infrastructure level). Being Prometheus a popular open source application and infrastructure monitoring</p>
--	--

	<p>system, and the QoS Evaluator a natural fit for the alerting capability of any monitoring system, virtually any place where the TM may be commercialized would be a good place to introduce the QoS Evaluation.</p> <p><u>Dynamic Orchestrator</u> (DD): This component requires the QoS Evaluator to continuously evaluate the performance of orchestrated applications and therefore of its own work adapting the underlying resources and infrastructure to provide application QoS. Thus, any commercialized DD-based solution will need to include the QoS Evaluator.</p>
<p><b>Is it expected an exploitation strategy for the component alone or along with another one?</b></p>	<p>Yes. This component is distributable and deployable in isolation of the rest of components of BigDataStack; therefore, there is a chance to the exploited separately.</p> <p>However, <u>Triple Monitoring</u> and QoS Evaluator together comprise a complete and powerful monitoring and QoS evaluation solution to be commercialized together.</p> <p>Moreover, QoS Evaluator is key in any solution based on the <u>Dynamic Orchestrator</u>; therefore, any commercialization of exploitation strategy including the latter would need to include and take the former into consideration.</p>
<p><b>Is ATOS interested in exploiting the component? How?</b></p>	<p>Yes, transferring this asset to the consulting teams from other business units of the company.</p>

Table 25. QoS Evaluator Preliminary Exploitation route

### 6.1.7. Information Driven Networking

- Main partner involved: SILO
- Description of the component

Name of the Component 1: Information Driven Networking	
<p><b>Short description of the component</b></p>	<p>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 needs. It supports a set of</p>

	mechanisms operating at services layer to understand the virtual hosts, URLs and other HTTP headers and at network layer to understand the workloads in storage services, DNS and a plethora of other services that do not use HTTP.
<b>Input of the component</b>	Types of policies to be enforced at service and network layer of networking operations among containers and virtual machines.
<b>Output of the component</b>	Specific policy enforcement according to security constraints, real-time requirements (within ms), or as data workloads are changing in the cloud environment.
<b>Describe the target customer segments/user groups of the component</b>	Network operators Network engineers Cloud providers
<b>Key features and benefits of the component</b>	Information driven networking is fully parameterized and can be deployed in any cloud environment.
<b>Initial TRL at the beginning of the project</b>	TRL3
<b>TRL to be reached at the end of the project</b>	TRL5
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	N/A
<b>Standards involved in the development of the component</b>	OSI standard
<b>What BigDataStack Use Case is the component related with?</b>	All use cases, as it is a backend cloud environment mechanism.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	No

Table 26. Information Driven Networking

- **Competitors and trends in the market**

<b>Competitor 1: Calico<sup>28</sup></b>	
<b>Short description</b>	Calico is a Pure Layer 3 Approach to Virtual Networking for Highly Scalable Data Centres.
<b>Developed by</b>	Tigera, Inc.

<sup>28</sup> <https://www.tigera.io/tigera-calico/>

<b>Type of License: Open Source or proprietary</b>	Open Source
<b>Business Models (free, license, pay per use, etc.)</b>	Free
<b>the bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	The Information Driven Networking component can be deployed over any cloud native environment.
<b>the minus of your component in relation to the competitor</b>	It is produced by a small development team and has not a fair potential compared with big players and their resources availability in this market.

Table 27. Information Driven Networking Competitor 1

<b>Competitor 2: Kuryr</b>	
<b>Short description</b>	Kuryr is a bridge between containers frameworks networking models to OpenStack networking abstraction layer.
<b>Developed by</b>	OpenStack
<b>Type of License: Open Source or proprietary</b>	Open Source, Apache License
<b>Business Models (free, license, pay per use, etc.)</b>	Free
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	The Information Driven Networking component can be deployed over any cloud native environment.
<b>The minus of your component in relation to the competitor</b>	It is produced by a small development team and has not a fair potential compared with big players and their resources availability in this market.

Table 28. Information Driven Networking competitor 2

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
<b>Internal</b>	<b>STRENGTHS</b> Simple for use and highly parameterized Scalable, as it is built on a fully distributed architecture Offers micro-segmentation capabilities able to support secure and resilient operations	<b>WEAKNESSES</b> Developed by a small development team with limited resources availability It has not a fair potential in this market as the equivalent competitors' component (e.g. Tigera Inc., OpenStack)
	<b>EXTERNAL</b>	<b>OPPORTUNITIES</b>

Diverse network engineering capabilities	Specific tool which is tailored in cloud infrastructures
Multiple policies enablement at different OSI layers	Big players in this market have already launched an equivalent component

Table 29. Information Driven Networking SWOT Analysis

- **Preliminary Exploitation route**

<b>Has the component synergies with other component regarding exploitation?</b>	It has synergies with the Triple Monitoring Engine component.
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	It is envisioned an exploitation strategy along with the Triple Monitoring Engine component.
<b>Is SILO interested in exploiting the component? How?</b>	We intend to integrate it within a product providing simple, scalable and secure network engineering mechanisms in the support of user defined network policies.

Table 30. Information Driven Networking Preliminary Exploitation route

## 6.1.8. Process Modelling Framework

- **Main partner involved: ATC**
- **Description of the component**

<b>Name of the Component: Process Modelling Framework</b>	
<b>Short description of the component</b>	The Process Modelling Tool provides an interface to business users to model their business processes and workflows.
<b>Input of the component</b>	<ul style="list-style-type: none"> <li>• A list of nodes (reflecting processes) and possible attributes per node</li> <li>• A list of connection restrictions</li> <li>• End-to-end graph / workflow objectives (e.g. accuracy, overall time, etc)</li> </ul>
<b>Output of the component</b>	A process workflow in JSON format
<b>Describe the target customer segments/user groups of the component</b>	<ul style="list-style-type: none"> <li>• Business Analysts in different industries</li> <li>• Data Scientists</li> </ul>
<b>Key features and benefits of the component</b>	It provides an easy way for business analysts to model their processes
<b>Initial TRL at the beginning of the project</b>	-
<b>TRL to be reached at the end of the project</b>	TRL5

<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	<ul style="list-style-type: none"> <li>• There are restrictions in the connection of the nodes</li> <li>• The tool is dockerized</li> <li>• It requires a username/password through the Visualisation tool</li> </ul>
<b>Standards involved in the development of the component</b>	N/A
<b>What BigDataStack Use Case is the component related with?</b>	All use cases (currently Shipping Management and Connected Consumer)
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	N/A

Table 31. Process Modelling Framework description

- **Competitors and trends in the market**

Not identified yet.

- **SWOT Analysis**

		Helpful	Harmful
Intern		<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Adaptable to every use-case		It is recommended to follow possible modifications of Data Toolkit.
External		<b>OPPORTUNITIES</b>	<b>THREATS</b>
	Easy-to-use Interface for business users to model their business processes and workflows		With regard to the collaborating components (e.g. Data Toolkit), there are dependencies from underlying technologies/tools.

Table 32. Process Modelling Framework SWOT Analysis

- **Preliminary Exploitation route**

<b>Has your component synergies with other component regarding exploitation?</b>	Not at the moment.
<b>Do you envision an exploitation strategy for your component alone or along with another one?</b>	The exploitation of the Process Modelling Framework is tightly coupled with the exploitation of the whole BigDataStack platform.
<b>Is ATC interested in exploiting the component? How?</b>	The Process Modelling Framework will be part of BigDataStack solution. ATC is planning to use the knowledge acquired through the development of this component into its products.

Table 33. Process Modelling Framework preliminary exploitation route

### 6.1.9. Visualization Environment

- Main partner involved: ATC
- Description of the component

Name of the Component: Adaptable Visualization	
Short description of the component	Adaptable Visualization
Input of the component	Process Modeller (Embedded) Data Toolkit (Embedded) REST API from BenchMarking RabbitMQ Service Dimensioning Workbench Triple Monitoring Engine (Embedded) REST API Data cleaning
Output of the component	Web Application
Describe the target customer segments/user groups of the component	Application owners, Infrastructure providers
Key features and benefits of the component	<p>Adaptable Visualization has a dual purpose:</p> <ol style="list-style-type: none"> <li>supporting the visualization of data analytics for the applications deployed in BigDataStack,</li> <li>provide a visual application performance monitoring dashboard of the data operations and the applications both during benchmarking and during their operation.</li> </ol> <p>Importantly, the dashboard will be able to monitor the application deployed over the infrastructure. The application and infrastructure performance monitoring dashboard will be adaptable, since it will enable to build a custom dashboard that can include charts with the KPIs chosen by the user out of a set of chart catalogue and a set of transformation operator that can aggregate and correlate the received metrics in any manner.</p>
Initial TRL at the beginning of the project	-
TRL to be reached at the end of the project	TRL5

<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	User has to be authorized and authenticated towards Adaptable Visualization. The component is fully deployed based on Docker container software service
<b>Standards involved in the development of your component</b>	JWT, Docker Compose, Kubernetes
<b>What BigDataStack Use Case is your component related with?</b>	DANAOS and ATOS Wordline UCs (until now).
<b>Do you envision that BigDataStack results or your components can influence in some Standardization foundation or initiative?</b>	N/A

Table 34. Adaptable Visualization

- **Competitors and trends in the market**

Not identified yet

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
<b>Internal</b>	<p>Adapt all the different technologies used for the different components, which are to be embedded into Adaptable Visualization Component</p> <p>Provide meaningful and relevant experiences to user on UX level</p> <p>Ensure the securely transmitting information between parties.</p> <p>Can work/deploy on the entire application front-end service at the entry level</p>	<p>Use of some components embedded within Adaptable Visualization Component. There will be a try to consume them as services.</p>
	<b>OPPORTUNITIES</b>	<b>THREATS</b>
<b>External</b>	<p>The user-friendly interface of a frontrunner infrastructure management system that drives decisions according to data aspects.</p> <p>The fully scalability, runtime adaptable potential and the performance on big data operations and data-intensive</p>	<p>With regard to the integrated components, there are dependencies from underlying technologies/tools in terms of deployment or load injection.</p>



	applications.	
--	---------------	--

Table 35. Adaptable Visualization SWOT Analysis

- **Preliminary Exploitation Route**

<b>Has the component synergies with other component regarding exploitation?</b>	Not at the moment
<b>Do you envision an exploitation strategy for your component alone or along with another one?</b>	The exploitation of Adaptable Visualization Component is tightly coupled with the exploitation of the whole BigDataStack platform.
<b>Is your organization interested in exploiting the component? How?</b>	The Adaptable Visualization component will be part of BigDataStack solution. ATC is planning to use the knowledge acquired through the development of this component to its products.

Table 36. Adaptable Visualization preliminary exploitation route

### 6.1.10. Seamless Analytics

- **Main partner involved: LXS**
- **Description of the component**

<b>Name of the Component: Seamless Analytics Framework</b>	
<b>Short description of the component</b>	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 in both the LXS DB and the Object Store). The component is responsible for retrieving the data that are split across datastores, join the results and return the response back to the user.
<b>Input of the component</b>	A standard SQL query
<b>Output of the component</b>	The resultset of the execution of the query, containing the aggregated answer of the SQL statement after it was redirected to both the LXS DB and the Object Store.

<b>Describe the target customer segments/user groups of the component</b>	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.
<b>Key features and benefits of the component</b>	Novel aggregation of these 2 data stores, exploiting the unique and diverse characteristics of each one of those, without truly compromising some requirements for the benefit of others.
<b>Initial TRL at the beginning of the project</b>	-
<b>TRL to be reached at the end of the project</b>	Prototype working for big data use cases (TRL4)
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	N/A
<b>Standards involved in the development of the component</b>	JDBC specification as the connector with the component.
<b>What BigDataStack Use Case is the component related with?</b>	This component could fit any of the use cases. However, the bigger the dataset, the more relevant is this component.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	Not for now

Table 37. Seamless description

- **Competitors and trends in the market**

Not identified yet

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
<b>Internal</b>	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Novel integration of two different data stores	Currently linked with the specific underlying data stores available in the project
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	First of its kind technology that eliminates the need to perform changes on application level to deal with different underlying data stores	We have made sure that the advantages of each separate data store are kept when used together with the seamless component

Table 38. Seamless SWOT Analysis

- **Preliminary Exploitation route**

<b>Has the component synergies with other component regarding exploitation?</b>	This is a common LXS/IBM component
<b>Is it expected an exploitation strategy for your component alone or along with another one? If so, explain which one, and if not, why.</b>	Still too soon to know
<b>Is LXS interested in exploiting the component? How?</b>	Still too soon to know, but an exploitation LXS/IBM could in theory be possible.

Table 39. Seamless preliminary exploitation route

### 6.1.11. Adaptable Distributed Storage

- **Main partner involved: LXS**
- **Description of the component**

<b>Name of the Component: Adaptable Distributed Storage</b>	
<b>Short description of the component</b>	<p>The Adaptable Distributed Storage component of BigDataStack consists of a novel mechanism that allows the data storage layer to be adapted to diverse workloads during runtime. It allows LXS datastore to partition its datasets to smaller fragments and split, move or merge those fragments effectivity among the available data nodes in order to achieve the balance of the load, both in terms of incoming workloads and stored data load. As scaling in/out the nodes of a datastore is not an easy task, databases generally either sacrificing the provision of transactional semantics in order to be able to scale effectively on the run-time or have to suffer of long periods of downtime or decreased performance during the scaling process. The Adaptable Distributed Storage on the other hand will allow the LXS datastore to scale effectively during runtime, without downtimes or increased performance overheads, while on the same time, ensuring transactional semantics. This will allow its storage to become truly elastic.</p>

<b>Input of the component</b>	Information regarding how a dataset can be fragmented (split points) that can be provided by the system administrator.
<b>Output of the component</b>	This component produces no output. The output could be an optimized allocation of the resources of the storage layer.
<b>Describe the target customer segments/user groups of the component</b>	Any enterprise that deals with Big Data 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 insurance is crucial.
<b>Key features and benefits of the component</b>	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.
<b>Initial TRL at the beginning of the project</b>	At the beginning of the project, the basic principles that this component will have to provide had been observed, while the basic design of the component had been started. We can say that it was in TRL 2.
<b>TRL to be reached at the end of the project</b>	At the end of the project, this component will have to be validated in a simulated environment and integrated with the whole BigDataStack platform. That could make its TRL to be 5.
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	A domain expert may have to define the split points for the data to be fragmented, by defining additional indexes on specific columns.
<b>Standards involved in the development of the component</b>	none
<b>What BigDataStack Use Case is the component related with?</b>	This component will not be demonstrated in M18. As a result, for its validation we use the TPCC dataset. At this point, the Use Case(s) that his component will be related with, has not been identified yet. Most probably, a Use Case with an increased data volume will be a candidate.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	no

Table 40. Adaptable Distributed Storage description

- Competitors and trends in the market

Competitor 1: Nuodb <sup>29</sup>	
<b>Short description</b>	Nuodb's distributed SQL database combines the elastic scale and continuous availability of the cloud with the transactional consistency and durability that databases of record demand.
<b>Developed by</b>	Nuodb
<b>Type of License: Open Source or proprietary</b>	Proprietary product
<b>Business Models (free, license, pay per use, etc.)</b>	Not identified
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor is</b>	The Adaptable Distributed Storage can scale out to numerous nodes, with no decreased performance. Moreover, the novel algorithm for resource allocation can allow the data to be fragmented/split and moved across data nodes, thus providing an optimal deployment that is expected to reduce the overall cost for resources.
<b>The minus of your component in relation to the competitor</b>	As the component is under the development phase, it is too soon to know.

Table 41. Adaptable Distributed Storage competitor

- SWOT Analysis

	Helpful	Harmful
Internal	<b>STRENGTHS</b> Elastic scalability while ensuring data consistency and transactional semantics	<b>WEAKNESSES</b> We come to market much later than competitors
	<b>OPPORTUNITIES</b> An enterprise can reduce its operational cost by better allocate 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.	<b>THREATS</b> Not identified yet
External		

<sup>29</sup> <https://www.nuodb.com/>

Table 42. Adaptable Distributed Storage SWOT Analysis

- **Exploitation preliminary route**

<b>Has the component synergies with other component regarding exploitation?</b>	This component is stand-alone and does not depends on other components. It is too soon to know synergies with the data infrastructure components can be useful, in order to request data resources where the overall workloads require more resources.
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	Alone. This technology naturally finds its place in the LXS data store.
<b>Is LXS interested in exploiting the component? How?</b>	Too soon to know.

Table 43. Adaptable Distributed Storage preliminary exploitation route

### 6.1.12. Data Toolkit

- **Main partner involved: UBI**
- **Description of the component**

<b>Name of the Component: Data Toolkit</b>	
<b>Short description of the component</b>	The Data Toolkit is the component which takes care to design an end-to-end Big Data application graph and create a common serialization format in order that it is feasible to execute valid analytics pipelines. It supports Spark and Spark ML library, but it may be also extended to support other machine learning frameworks (namely indicatively Tensorflow, Scikit-learn).
<b>Input of the component</b>	High level graph from Process Modelling component.
<b>Output of the component</b>	Fine grained, enriched and grounded execution graph.
<b>Describe the target customer segments/user groups of the component</b>	<ul style="list-style-type: none"> <li>- Data scientists</li> <li>- Data analysts</li> <li>- Data engineers</li> </ul>
<b>Key features and benefits of the component</b>	<ul style="list-style-type: none"> <li>- Automation</li> <li>- Online graph composition constraints enforcement</li> </ul>
<b>Initial TRL at the beginning of the project</b>	TRL3
<b>TRL to be reached at the end of the project</b>	TRL5

Essential information for potential users (requirements, use restrictions, etc.)	N/A
Standards involved in the development of the component	N/A
What BigDataStack Use Case is the component related with?	All Use Cases, currently DANAOS & ATOS WRL. However, it is not limited to specific UCs as it adopts a generalized approach which can be applicable to other cases, too.
Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?	No

Table 44. Data Toolkit Description

- Competitors and trends in the market

Competitor: Kuberflow <sup>30</sup>	
Short description	Kubeflow is the Machine Learning Toolkit for Kubernetes
Developed by	Google Developers
Type of License: Open Source or proprietary	Open Source, Apache 2.0 License
Business Models (free, license, pay per use, etc.)	Free
The bonus (features, status, business model, etc.) of your component in relation to the competitor	Data Toolkit can be deployed over any type of cloud computing operating system, including OpenStack, Microsoft Azure, etc.
the minus of your component in relation to the competitor	It is produced by a small development team and has not a fair potential compared with big players and their resources availability in this market.

Table 45. Data Toolkit Competitor

- SWOT Analysis

	Helpful	Harmful
	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
Internal	Automation	Developed by a small development team with limited resources availability
	Online graph validation & deployment	
	Easy to use	It has not a fair potential in this market as the equivalent competitors'

<sup>30</sup> <https://github.com/kubeflow/kubeflow>

	Easy to parameterize it  Horizontal component with huge potential in multiple verticals	component (e.g. Google, Microsoft)
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	End-to-end Big Data orchestration apps enablement  Several and diverse vertical markets can use it	Big players in this market have already launched an equivalent component  Financial crisis does not leave space for non-critical addons  Specific tool which is tailored in cloud infrastructures

Table 46. Data Toolkit SWOT Analysis

- **Preliminary Exploitation Route**

<b>Has the component synergies with other component regarding exploitation?</b>	It has synergies with Process Modeling and Adaptive Visualization component.
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	UBI envisions an exploitation strategy along with Process Modelling and Adaptive Visualization component as an integrated product for automated machine learning functionalities.
<b>Is UBI interested in exploiting the component?</b>	UBI intends to integrate it within a product providing simplified analytics with powerful, no-code or with code according to users' experience, automated machine learning capabilities.

Table 47. Data Toolkit preliminary Exploitation Route

### 6.1.13. Data Quality Assessment

- **Main partner involved: UPRC**
- **Description of the component**

<b>Name of the Component: Data Quality Assessment</b>	
<b>Short description of the component</b>	A domain-agnostic data assessment and improvement framework, that can identify valid records in a relational database and establish data veracity.
<b>Input of the component</b>	Datasets



<b>Output of the component</b>	Evaluated quality of the datasets on a per-tuple level
<b>Describe the target customer segments/user groups of the component</b>	Data quality assessment 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.
<b>Key features and benefits of the component</b>	The main contribution of the data quality assessment and improvement module is that is able to provide a domain-agnostic framework for automating the data cleaning.
<b>Initial TRL at the beginning of the project</b>	-
<b>TRL to be reached at the end of the project</b>	TRL5
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	The data quality assessment and improvement framework consume a respectable amount of data in order to train its model and discover intricate relationships between the values in a data set. The use of GPU can also facilitate this process by allowing the algorithm to leverage the multicore architecture of such hardware.
<b>Standards involved in the development of the component</b>	N/A
<b>What BigDataStack Use Case is the component related with?</b>	Data Cleaning component is targeting all the BigDataStack use cases.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	N/A

Table 48. Data Cleaning description

- **Competitors and trends in the market**

<b>Competitor: HoloClean<sup>31</sup></b>	
<b>Short description</b>	HoloClean
<b>Developed by</b>	Stanford University and University of Waterloo
<b>Type of License: Open Source or proprietary</b>	Apache 2.0

<sup>31</sup> <https://github.com/HoloClean/holoclean>

<b>Business Models (free, license, pay per use, etc.)</b>	free
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	Domain agnostic
<b>The minus of your component in relation to the competitor</b>	Not enough community engagement and recognition

Table 49. Data Cleaning Competitor

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
<b>Internal</b>	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Domain-agnostic process Automation of tedious tasks	Historical data required for training the deep learning approach of the component in order to assess the data quality (i.e. cold start)
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	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  Increase of Big Data application in the market	Valid and corrupted datasets are needed to train the model. The latter are usually hard to find

Table 50. Data Cleaning SWOT Analysis

- **Preliminary Exploitation Route**

<b>Has the component synergies with other component regarding exploitation?</b>	N/A
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	The data assessment and improvement framework will be an open source library that our aim is to continue to expand and use in the projects to come. Being open source will invite fellow developers to work on it and extend its features.
<b>Is UPRC interested in exploiting the component? How?</b>	Our organization aims to reap the research results of such a project and continue to use the outcomes in future projects and software designs as a separate library.

Table 51. Data Cleaning preliminary exploitation route

## 6.1.14. Application Dimensioning Workbench

- Main partner involved: UPRC
- Description of the component

<b>Name of the Component: Application Dimensioning Workbench</b>	
<b>Short description of the component</b>	Application Dimensioning Workbench
<b>Input of the component</b>	Application playbook created in Data toolkit and annotated with candidate deployment patterns in Pattern Generation
<b>Output of the component</b>	Annotated playbook with expected QoS levels for each pattern
<b>Describe the target customer segments/user groups of the component</b>	Primarily data service owners (such as LXS, IBM and UPM in BigDS) and/or application owners
<b>Key features and benefits of the component</b>	The component has two purposes: a) initially benchmark the target service via easily configured and automated parameter sweep tests, thus gathering the necessary performance data b) train prediction models that are able to regress for cases that have not been met before
<b>Initial TRL at the beginning of the project</b>	3 for the model creation part, 0 for the benchmarking part (the benchmarking part did not exist at the beginning of the project)
<b>TRL to be reached at the end of the project</b>	5 for both
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	Load injection needs to be adapted to each service that needs to be benchmarked. Also service structure should follow the format defined by the Data toolkit. Deployment and execution of the tests depends on the used deployment platform (e.g. Docker Swarm, Kubernetes etc)
<b>Standards involved in the development of the component</b>	YML, Docker Compose v3, Kubernetes specification
<b>What BigDataStack Use Case is the component related with?</b>	The component is primarily related to the generic data services offered by the project. Indicative workloads for these have been drawn at this point for the DANAOS and ATOS Wordline UCs.
<b>Is it expected that BigDataStack results or the components can influence in</b>	N/A

<b>some Standardization foundation or initiative?</b>	
---	--

Table 52. Application Dimensioning Workbench description

- **Competitors and trends in the market**

<b>Competitor 1: CBTool<sup>32</sup></b>	
<b>Short description</b>	Cloudbench Tool (CBTOOL) is a multi-benchmark harness that automates infrastructure as a service (IaaS) cloud stress and scalability testing through running controlled experiments.
<b>Developed by</b>	IBM and community
<b>Type of License: Open Source or proprietary</b>	Open Source
<b>Business Models (free, license, pay per use, etc.)</b>	Apache License v2.0
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	Workload in CBTool is completely designed by user/contributor, thus cannot be easily abstracted and generalized as in the case of ADW. Not clear if parameter sweep experiments can be defined and workloads seem to be statically defined and not altered based on current need
<b>The minus of your component in relation to the competitor</b>	CBTool has already a number of deployment modes available

Table 53. Application Dimensioning Workbench competitor 1

<b>Competitor 2: CloudPerfect Benchmarking Suite<sup>33</sup></b>	
<b>Short description</b>	The Benchmarking Suite is an all-in-one solution for benchmarking cloud services simulating different typical application behaviours and comparing the results on different cloud providers.
<b>Developed by</b>	FP7 ARTIST and H2020 Cloudperfect projects, Engineering
<b>Type of License: Open Source or proprietary</b>	Apache License v2.0
<b>Business Models (free, license, pay per use, etc.)</b>	Free, potential exploitation channel through QoE, a portal that would use the benchmark regularly and then commercialize the benchmark scores against the Cloud providers
<b>The bonus (features, status, business model, etc.) of your</b>	Benchsuite is designed for running the same tests in a repetitive process against different Cloud

<sup>32</sup> <https://github.com/ibmcb/cbtool>

<sup>33</sup> <https://benchmarking-suite.readthedocs.io/en/latest/>

<b>component in relation to the competitor</b>	providers. Thus, workload adaptation is difficult as well as parameter sweep definition of a test, in contrast to ADW. Furthermore, it does not include generic deployments that can benchmark a service graph but targets at measuring individual resources performance on the used benchmarks.
<b>The minus of your component in relation to the competitor</b>	N/A

Table 54. Application Dimensioning Workbench competitor 2

<b>Competitor 3: CloudSuite<sup>34</sup></b>	
<b>Short description</b>	CloudSuite is a benchmark suite for cloud services. The third release consists of eight applications that have been selected based on their popularity in today's datacenters. The benchmarks are based on real-world software stacks and represent real-world setups.
<b>Developed by</b>	Parallel Systems Architecture Lab, EPFL
<b>Type of License: Open Source or proprietary</b>	CloudSuite 3.0 License (Open source)
<b>Business Models (free, license, pay per use, etc.)</b>	Free, no mentioning of support plans or assignment for extensions
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	Static setups based on specific stacks, measuring in fact the underlying resources and not the software parameterization (and specifically parameterization of data services like in the case of ADW)
<b>The minus of your component in relation to the competitor is</b>	ADW is based on the specific data services used in BigDS and thus is less generic than the setups included in CloudSuite. However, it can be extended with relevant dockerized setups

Table 55. Application Dimensioning Workbench competitor 3

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
<b>Internal</b>	Decoupling between the framework and the underlying load injection process	The aforementioned decoupling in the STRENGTHS section implies that adaptation drivers should exist for new cases of deployment targets as well as

<sup>34</sup> <https://cloudsuite.ch/>

	<p>Can work/deploy on the entire application service graph and load generate at the entry level</p> <p>Decoupling between the framework and the underlying deployment process</p> <p>Automation of stress test/benchmark implementation in a parameter sweep fashion</p> <p>Investigation of application/data service workload features and enable further reasoning on the performance data</p>	<p>underlying benchmarks used. However, given that we follow a dockerized approach, this inclusion is expected to be performed with little effort per case.</p> <p>Need to adapt launch execution and parameter passing based on each individual service graph (however this is also linked with the commission-based development in Opportunities)</p>
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	<p>Ability to understand which parameters (e.g. data service configuration parameters) affect performance metrics can lead to instrumentation of a specific deployment based on the tailored needs.</p> <p>Commission based Development of the adaptable drivers per needed business case can be viewed as an exploitation opportunity after the project end.</p> <p>Ability to regulate the size of the experiment and thus act also as stress testing framework</p>	<p>Dependencies from underlying technologies/tools in terms of deployment or load injection. However, the fact that these are pluggable enable us to simply discard cases that at some point may be deemed as obsolete or having lost their wide adoption.</p>

Table 56. Application Dimensioning Workbench SWOT Analysis

- **Preliminary Exploitation Route**

<b>Has the component synergies with other component regarding exploitation?</b>	Not at the moment
<b>Is it expected an exploitation strategy for the component alone or along with another one?</b>	Both cases can apply. For the overall ADW including the annotation of the application playbook, synergies can be identified with Pattern Generation as well as with the BigDS platform.

	<p>For the case of the Benchmarking part of the component, 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. Furthermore, given that load generation should be adapted to the specific UC, implementation of interviews and design of the respective workload aspects based on the interview outcome could also be a service offered to clients.</p>
<p><b>Is UPRC interested in exploiting the component? How?</b></p>	<p>Except for the scope mentioned in the previous row regarding standalone exploitation of the tool, we expect also to use the tool as baseline in new R&amp;D projects as well as exploit publications performed for its design and results.</p> <p>Another form of exploitation is its publication as an open source project with satisfactory download/usage statistics.</p>

Table 57 Application Dimensioning Workbench preliminary exploitation route

### 6.1.15. Process Mapping

- Main partner involved: UPRC
- Description of the component

Name of the Component: Process Mapping	
Short description of the component	Process Mapping
Input of the component	A step in a process that corresponds to a Machine learning task (e.g., clustering).
Output of the component	A specific machine learning (ML) algorithm that is automatically mapped to the process (e.g., k-means).
Describe the target customer segments/user groups of the component	Any organization/company/institution/individual that uses machine learning
Key features and benefits of the component	<ul style="list-style-type: none"> <li>• Automatic algorithm selection for ML tasks</li> <li>• The component follows a meta learning approach, thus it improves its performance as it is applied on more and more datasets</li> </ul>

	<ul style="list-style-type: none"> <li>Reduce time and effort spent for finding the best performing algorithm for an ML task</li> </ul>
<b>Initial TRL at the beginning of the project</b>	-
<b>TRL to be reached at the end of the project</b>	TRL5
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	The component will automatically select an algorithm from a pre-specified set of algorithms.
<b>Standards involved in the development of your component</b>	N/A
<b>What BigDataStack Use Case is your component related with?</b>	Can be applied in all BigDataStack use-cases.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	Probably not.

Table 58. Process Mapping description

- Competitors and trends in the market**

<b>Competitor 1: Auto-WEKA<sup>35</sup></b>	
<b>Short description</b>	Auto-WEKA considers the problem of simultaneously selecting a learning algorithm and setting its hyperparameters, going beyond previous methods that address these issues in isolation.
<b>Developed by</b>	Artificially Intelligent Manufacturing center (AIM) at the University of Wyoming.
<b>Type of License: Open Source or proprietary</b>	Open source
<b>Business Models (free, license, pay per use, etc.)</b>	Released under the GNU General Public License version 3.
<b>The bonus (features, status, business model, etc.) of your component in relation to the competitor</b>	<ul style="list-style-type: none"> <li>Focus on unsupervised learning.</li> </ul>
<b>The minus of your component in relation to the competitor</b>	<ul style="list-style-type: none"> <li>Our component does not focus on supervised learning yet.</li> <li>Our component does not focus on hyperparameter tuning yet.</li> </ul>

Table 59. Process Mapping competitor 1

<sup>35</sup> <https://www.cs.ubc.ca/labs/beta/Projects/autoweka/>



<b>Competitor 2: Auto-SKLEARN<sup>36</sup></b>	
<b>Short description</b>	Auto-sklearn provides out-of-the-box supervised machine learning. Built around the scikit-learn machine learning library, auto-sklearn automatically searches for the right learning algorithm for a new machine learning dataset and optimizes its hyperparameters.
<b>Developed by</b>	AutoML Freiburg
<b>Type of License: Open Source or proprietary</b>	Open source
<b>Business Models (free, license, pay per use, etc.)</b>	Free
<b>The bonus (features, status, business model, etc.) of the component in relation to the competitor</b>	Focus on unsupervised learning.
<b>The minus of the component in relation to the competitor</b>	The component does not focus on supervised learning yet. The component does not focus on hyperparameter tuning yet.

Table 60. Process Mapping competitor 2

- **SWOT Analysis**

	<b>Helpful</b>	<b>Harmful</b>
<b>Internal</b>	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Special focus on unsupervised learning (clustering), which is not targeted by others (mostly focusing on supervised learning)	It is a research prototype, so not supported as well as a commercial product
<b>External</b>	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	<p>Could be used to build an assisting tool for data scientists or machine learning practitioners</p> <p>Could be used for training activities of students working on big data analytics</p> <p>Reducing the time needed to find a suitable machine learning algorithm for a specific task is both saving resources and money for a company</p>	Similar functionality being provided by one of the competitors

Table 61. Process Mapping SWOT Analysis

<sup>36</sup> <https://automl.github.io/auto-sklearn/master/>

- **Preliminary Exploitation Route**

Has the component synergies with other component regarding exploitation?	With Process modeling framework.
Is it expected an exploitation strategy for your component alone or along with another one?	Probably alone but will depend on other components' exploitation actions.
Is UPRC interested in exploiting the component?	Not identified yet

Table 62. Process Mapping preliminary exploitation route

### 6.1.16. Complex Event Processing

- **Main partner involved: UPM**
- **Description of the component**

Name of the Component: <b>Complex Event Processing</b>	
Short description of the component	The Complex Event Processing (CEP) is in charge of processing flows of data on the fly without storing the data.
Input of the component	Data streams
Output of the component	Processes streams based on different and complex rules
Describe the target customer segments/user groups of the component	Any application that needs to process continuous (infinite) flows of data without storing them.
Key features and benefits of the component	It runs in a distributed system made out of heterogeneous devices (servers, laptops, Raspberry Pi...). The CEP takes into consideration the resources of the device in order to deploy and run the queries and where data is generated to minimize network usage and latency.
Initial TRL at the beginning of the project	TRL2
TRL to be reached at the end of the project	TRL3
Essential information for potential users (requirements, use restrictions, etc.)	N/A
Standards involved in the development of the component	No
What BigDataStack Use Case is the component related with?	Shipping management use case

Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?	No
--	----

Table 63. Complex Event Processing description

- Competitors and trends in the market

Competitor: Flink <sup>37</sup>	
Short description	Distributed Data Streaming Engine
Developed by	Software Artisans
Type of License: Open Source or proprietary	Open Source License Apache
Business Models (free, license, pay per use, etc.)	Not identified
The bonus (features, status, business model, etc.) of your component in relation to the competitor	Deployment in a heterogeneous device with very limited resources. The run time takes into consideration those resources and can move queries dynamically to other nodes to improve performance. Dynamic reconfiguration of the queries is performed.
The minus of your component in relation to the competitor	TRL low, very small number of PMs to increase the TRL

Table 64. Complex Event Processing Competitor

- SWOT Analysis

	Helpful	Harmful
Internal	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Dynamic reconfiguration of the query deployment Resource aware deployment	Immature implementation
External	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	Adoption by project partners	Competing open source projects

Table 65. Complex Event Processing SWOT Analysis

<sup>37</sup> <https://flink.apache.org/>

- **Preliminary Exploitation Route**

<b>Has the component synergies with other component regarding exploitation?</b>	LeanXscale Database
<b>Is it expected an exploitation strategy for your component alone or along with another one?</b>	It can be exploited as a standalone component or combined with the LeanXscale database.
<b>Is UPM organization interested in exploiting the component? How?</b>	Yes. Through exploitation agreements

Table 66. Complex Event Processing preliminary exploitation route

### 6.1.17. Deployment Recommendation Service

- **Main partner involved: GLA**
- **Description of the component**

<b>Name of the Component: Deployment Recommendation Service</b>	
<b>Short description of the component</b>	This component is designed to produce a recommended deployment configuration from amongst a series of options (candidate deployment patterns). In effect, it aims to estimate how resources allocated to each service will impact factors the user cares about, such as quality of service or cost, and uses that information to decide how much resourcing an application needs to perform effectively and efficiently.
<b>Input of the component</b>	Candidate Deployment Patterns with Benchmarking Information
<b>Output of the component</b>	One or more recommended candidate deployment patterns
<b>Describe the target customer segments/user groups of the component</b>	Application Engineers
<b>Key features and benefits of the component</b>	Allows for automated specification of application resources in cases where the user does not know the performance profile of their application.
<b>Initial TRL at the beginning of the project</b>	TRL1
<b>TRL to be reached at the end of the project</b>	TRL5
<b>Essential information for potential users (requirements, use restrictions, etc.)</b>	Service is dependent on components higher in the stack, namely: Dimensioning

	Workbench, Pattern Generation, Data Tool-kit and Process Modeller.
<b>Standards involved in the development of the component</b>	None
<b>What BigDataStack Use Case is the component related with?</b>	All use-cases that may need deployment assistance. It is envisaged that the <i>Connected Consumer</i> use case will be the most benefited by the technology.
<b>Is it expected that BigDataStack results or the components can influence in some Standardization foundation or initiative?</b>	Unclear

Table 67. Deployment Recommendation Service description

- **Competitors and trends in the market**

Not competitors identified yet for this type of technology

- **SWOT Analysis**

	Helpful	Harmful
Internal	<b>STRENGTHS</b>	<b>WEAKNESSES</b>
	Application of existing technology to a new domain  Helps decrease expertise needed for cloud deployment applications	If successful should be relatively easy to reproduce
External	<b>OPPORTUNITIES</b>	<b>THREATS</b>
	Technology is not available on the market	Technology may provide limited benefit in scenarios with a large number of heterogeneous applications (i.e. attaining generalizability across a wide range of application types will be difficult)  Technology is dependent on available training data from sample applications

Table 68. Deployment Recommendation SWOT Analysis

- **Preliminary Exploitation Routes**

<b>Has the component synergies with other component regarding exploitation? Please, identify which one</b>	Service is dependent on components higher in the stack, namely: Dimensioning Workbench, Pattern Generation, Data Tool-kit and Process Modeller, and hence would need to be provided as a package.
--	---

Is it expected an exploitation strategy for your component alone or along with another one? If so, explain which one, and if not, why.	In conjunction with the wider platform
Is GLA interested in exploiting the component? How?	Not commercially

Table 69. Deployment Recommendation Service preliminary exploitation route

## 6.2. Preliminary IPR and Licensing Analysis

In this phase of the project, it has been elaborated an initial list of the subcomponents’ and libraries’ licenses used to develop the SW components, and it has been identified components’ owners.

To define an appropriate license for each component, it will be used the next figure of OS License compatibilities, which help us to choose the compatible license for a component software resulting of combining several SW components with different OS licenses.

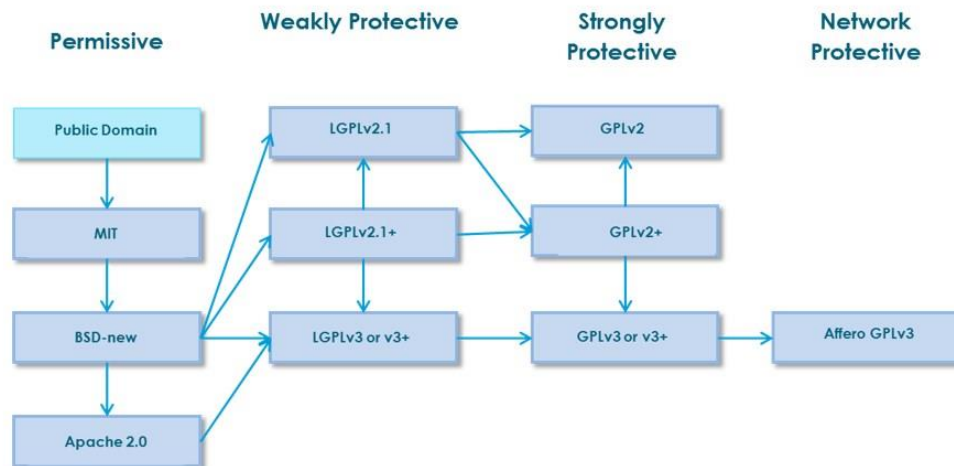


Figure 3. OS Licenses compatibility

It is worth mentioning two success stories with the components developed within BigDataStack project by two of the partners:

- IBM filled a patent around end of March 2019, for the novelty of its “Extensible Data Skipping” component
- LeanXcale is preparing the documentation to fill a patent for its component “Adaptable Distributed Storage” in next 6 months.

### 6.2.1. Data Skipping

Component Name	Libraries	License
SQL Data Skipping		TBD
	Apache Spark	Apache License 2.0 <sup>38</sup>

Table 70. SQL Data Skipping libraries and licenses

### 6.2.2. Dynamic Orchestration

Component Name	Libraries	License
Dynamic Orchestrator		GPL-Compatible
	RabbitMQ	Mozilla Public License <sup>39</sup>
	Python 3.4.0	GPL-Compatible <sup>40</sup>
	pika	BSD 3-Clause “New” or “Revised” <sup>41</sup>
	NumPy	BSD
	pyyaml	MIT License <sup>42</sup>

Table 71. Dynamic Orchestrator libraries and licenses

### 6.2.3. Network Policy support at Kuryr, Kuryr Cluster Network Operator and NvME-mdev Kernel driver

We are contributing to upstream projects, such as Kuryr-Openstack or Kuryr-Kubernetes, so we work under the preestablish license for them. Most of those licenses are Apache 2.0 or GPL-2.0 for linux kernel, for instance.

### 6.2.4. QoS Evaluation

Component Name	Libraries	License
QoS Evaluator Owner: Atos		Apache License 2.0
	Go	BSD
	viper	MIT
	gorilla mux	BSD 3-clause
	bbolt	MIT
	mgo	BSD 2-clause
	mgo/bson	BSD 2-clause
	govaluate	MIT

Table 72. QoS Evaluator libraries and Licenses

<sup>38</sup> <https://www.apache.org/licenses/LICENSE-2.0>

<sup>39</sup> <https://www.mozilla.org/en-US/MPL/>

<sup>40</sup> [http://gplv3.fsf.org/wiki/index.php/Compatible\\_licenses](http://gplv3.fsf.org/wiki/index.php/Compatible_licenses)

<sup>41</sup> <https://opensource.org/licenses/BSD-3-Clause>

<sup>42</sup> <https://opensource.org/licenses/MIT>

## 6.2.5. Information Driven Networking

Component Name	Libraries	License
	<b>Apache License 2</b>	
	<b>Drools</b>	Apache License 2.0
	<b>Prometheus</b>	Apache License 2.0

Table 73. Information Driven Networking libraries and Licenses

## 6.2.6. Process Modelling Framework

Component Name	Libraries	License
<b>Process Modelling Tool</b>	<b>MIT</b>	
	Vue.js	MIT
	Rete.js	MIT

Table 74. Process Modelling Framework libraries and licenses

## 6.2.7. Visualization Environment

Component Name	Libraries	License
<b>Adaptable Visualization</b>	<b>Apache License 2.0</b>	
	React JS Redux	MIT License
	Fuse	All Rights Reserved within pixels
	Material UI	MIT License
	Tailwind CSS	MIT License
	NodeJS	MIT License
	Rabbit MQ	Mozilla Public Licence
	Socket IO	MIT License
	React Charts JS 2	MIT License
	JWT authentication	MIT License
Docker Container	Apache license 2.0	

Table 75. Visualization Environment libraries and Licenses

## 6.2.8. Seamless Analytics

Component Name	Libraries	License
<b>Seamless Analytics Framework</b>	<b>TBD</b>	
	Thrift JDBC connector to Apache Spark	Apache License 2.0
	Apache Spark	Apache License 2.0
	Apache Derby	Apache License 2.0

Table 76. Seamless libraries and licenses



## 6.2.9. Adaptable Distributed Storage

This component has been entirely developed by LXS, without using any external libraries or other components.

### 6.2.10. Data Toolkit

Component Name	Libraries	License
Data Toolkit		MIT License
	ReactJS	MIT License

Table 77. Data Toolkit libraries and licenses

### 6.2.11. Data Quality Assessment

Component Name	Libraries	License
Data Quality Assessment		LGPL License
	TensorFlow	Apache 2.0
	Pytorch	BDS 2
	Pandas	BSD 3-Clause
	Numpy	BSD
	JayDeBeAPI	LGPL License
	Scikit-Learn	New BSD License

Table 78. Data Cleaning libraries and licenses

### 6.2.12. Application Dimensioning Workbench

Component Name	Libraries	License
Application Dimensioning Workbench		Apache License 2.0
	Node.js	MIT License
	Node-red tool, node-red-dashboard, node-red-sqlite	Apache License 2.0
	Baseline benchmarks such as Ycsb, jmeter etc	Apache License 2.0

Table 79. Application Dimensioning Workbench libraries and licenses

### 6.2.13. Process Mapping

Component Name	Libraries	License
Process Mapping		GPL v3 License
	WEKA <a href="https://www.cs.waikato.ac.nz/ml/weka/">https://www.cs.waikato.ac.nz/ml/weka/</a>	GPLv3 License <sup>43</sup>

Table 80. Process mapping libraries and licenses

<sup>43</sup> <https://www.gnu.org/licenses/gpl-3.0.html>

### 6.2.14. Complex Event Processing

Component Name	Libraries	License
Complex Event Processing		<b>TBD</b>
	Zookeeper	Apache License 2.0

Table 81. Complex Event Processing libraries and licenses

### 6.2.15. Deployment Recommendation Service

Component Name	Libraries	License
Deployment Recommendation Service		<b>Apache License 2.0</b>
	Apache Flink	Apache License 2.0
	Apache log4j	Apache License 2.0
	Google GSON	Apache License 2.0
	Fasterxml Jackson	Apache License 2.0
	Apache HttpComponents	Apache License 2.0

Table 82. Deployment Recommendation Service libraries and licences

## 7 Exploitation Strategy

### 7.1. Individual Exploitation Plans

BigDataStack partners have provided their preliminary exploitation interest and intentions for the first period of the project. These plans will evolve during the project lifetime as the technical results develop.

In the next section, all the partners' exploitation plans are presented, but a summary of the individual exploitation intentions shown by the partners for the first period of the project is depicted below.

Partners Profile	Initial Exploitation Interest
Industrial	<ul style="list-style-type: none"> <li>• Improve their current portfolio with Big Data features resulting from the R&amp;D performed within the project and provide new services</li> <li>• Further research with real use cases within the project</li> <li>• Enhance their current products with future AI features</li> <li>• Strengthens their current expertise with the potential to extend it towards High-complexity Transactional Software Systems in the areas of Industry, Business, Finance, Insurance, Health and Public Administration</li> <li>• Extend their current strategy for research and innovation reinforcing its internal capacity, experience and technological know-how in new emerging technologies addressing specific user needs</li> </ul>
Academia	<ul style="list-style-type: none"> <li>• Share knowledge among educational programs</li> <li>• Explore technology transfer with other institutions</li> <li>• Consultancy and specialized scientific expertise services</li> <li>• Licensing project results to a spin-off</li> <li>• Explore new future Research Lines</li> </ul>

Table 83. Summary of Partners individual exploitation interests and intentions

#### 7.1.1. IBM ISRAEL – SCIENCE AND TECHNOLOGY LTD

IBM Haifa Research Lab (HRL) is a center of competence in cloud computing and analytics, with a special focus on BigData analytics over Object Storage. HRL works closely with other IBM research and development labs, delivering innovated technologies which are integrated into IBM products and services.

HRL has a leading role in the Storage aspects of the BigDataStack project in general and in particular in the Object storage aspects. Participating into BigDataStack is an opportunity for the organization to interact with high level research/industrial partners, being introduced to real world use cases (which has been proven to be very fruitful for their research) and

participate in the development of new technologies that can be exploited by IBM products and services.

HRL has a well-established Cloud computing department which participated and is participating in multiple European projects related to cloud technologies (COSMOS, IOStack, etc.) and history of innovations developed in these projects and beyond which were integrated in IBM products and services.

IBM are responsible for the project coordination of BigDataStack project and they also lead the work package 4 and the big data layout task (4.7) for which the data skipping, and data partitioning aspects are central.

### **Exploitation Strategy**

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 will be a natural candidate for contribution to the IBM SQL Query service<sup>44</sup>, which their group in HRL is in close collaboration with its developers and product management. Optimized data layout is critical to 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 is to integrate the developed technology into the IBM SQL Query service (and possibly other IBM services). At this phase of the project, IBM have 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.

IBM early stage individual exploitation plan is:

- Year 1: Enhance the data skipping technology
- Year 2: Apply data partitioning and data skipping to use cases of the project, measure the resulting file skipping percentages and infer algorithm strength
- Year 3: reach publishable research level. At this stage both data skipping and data partitioning should be mature enough to be used in production. IBM SQL service Product team (and possible other IBM services) are targeted for technology contribution and integration.

### **7.1.2.NEC EUROPE LTD**

Under the brand statement, "Orchestrating a brighter world," NEC Group is focusing on Solutions for Society businesses that utilize the strengths of ICT to create the social value of safety, security, efficiency and equality that is necessary for people to live more prosperous lives. We firmly believe that social and business values increase when things and contexts are digitally connected, digitalized knowledge becomes common assets, and artificial intelligence (AI) yields new findings. NEC Laboratories Europe is in charge of carrying out advanced

---

<sup>44</sup> <https://www.ibm.com/cloud/blog/announcements/introducing-ibm-cloud-sql-query>

research addressing technology and business trends on ICT platforms and social solutions in Europe and worldwide.

They collaborate in the BigDataStack project with the development of the Dynamic Orchestrator, a controller that manages adaptation mechanisms during runtime in order to adapt to the changing context and comply with Big Data applications and services requirements.

In the IoT group at NEC Laboratories Europe, we have already been researching the topic of adaptive systems. We have presented a number of papers and patents about it, such as “Network-integrated edge computing orchestrator for application placement” [V. Karagiannis, A. Papageorgiou]<sup>45</sup>, “Towards adaptive actors for scalable iot applications at the edge” [J. Fuerst, M. Fadel Argerich, K. Chen and E. Kovacs]<sup>46</sup> and “Elastic Services for Edge Computing [J. Fuerst, M. Fadel Argerich, B. Cheng and A. Papageorgiou]”<sup>47</sup> (to be presented at CNSM’18). Because of this, NEC has the necessary knowledge and skills to design and implement the Dynamic Orchestrator for BigDataStack project.

NEC is in charge of the design and implementation of the Dynamic Orchestrator, part of WP3 of the BigDataStack project. They are currently researching about the application of a RL-based algorithm to orchestrate applications and data services. Their goal is to design a RL-based approach which can operate efficiently, with a light overhead for the overall system. Moreover, they aim to implement this and test it with different use cases to evaluate its performance.

### Exploitation Strategy

The implementation of a Dynamic Orchestrator is a subject in which NEC is very interested; They are currently exploring the application of AI to dynamically orchestrate tasks, as well as adapting its behaviour for NEC Business Unit. In particular, they are looking into the implementation of a Reinforcement Learning (RL) based logic orchestrator for FogFlow, a distributed execution framework to dynamically orchestrate IoT services over cloud and edges.

Their preliminary plan for the first version of the Dynamic Orchestrator is to keep it open for the consortium only.

NEC’s early stage individual exploitation plan is:

- Year 1: Design and implementation of a prototype for the Dynamic Orchestrator. Integration of the prototype to the BigDataStack for testing, early evaluation of its performance.
- Year 2: Further development of Dynamic Orchestrator, inclusion of more adaptive actions and improvements to the RL algorithm implementation. Testing and evaluation of its performance for the BigDataStack use cases. By using the knowledge acquired on this development, design and development of orchestrator for our own organization framework FogFlow.

---

<sup>45</sup> [https://www.researchgate.net/publication/322515292\\_Network-integrated\\_Edge\\_Computing\\_Orchestrator\\_for\\_Application\\_Placement](https://www.researchgate.net/publication/322515292_Network-integrated_Edge_Computing_Orchestrator_for_Application_Placement)

<sup>46</sup> [https://www.researchgate.net/publication/327238732\\_Towards\\_Adaptive\\_Actors\\_for\\_Scalable\\_IoT\\_Applications\\_at\\_the\\_Edge](https://www.researchgate.net/publication/327238732_Towards_Adaptive_Actors_for_Scalable_IoT_Applications_at_the_Edge)

<sup>47</sup> [https://www.researchgate.net/publication/331651737\\_Elastic\\_Services\\_for\\_Edge\\_Computing](https://www.researchgate.net/publication/331651737_Elastic_Services_for_Edge_Computing)

- Year 3: Completion of development of the Dynamic Orchestrator. Extensive testing and bug fixing of the component as part of the BigDataStack for different use cases and in different contexts, in order to make the orchestrator robust and efficient. Objective evaluation about its performance for a variety of use cases.

### 7.1.3. RED HAT ISRAEL LTD

Red Hat is the premier Linux and open source provider and is the recognized leader in enterprise solutions that take full advantage of the quality and performance provided by using the open source model. With Red Hat, enterprise hardware and software vendors have a standard platform on which to certify their technology. Red Hat Israel was founded in 2008, when Red Hat acquired Qumranet, an Israeli startup the founder and developer of KVM. Red Hat offers a wide range of consulting and engineering services to make enterprise open source deployments successful from complete Linux migration to client-directed engineering to custom software development and has broad expertise in open source technology. Red Hat is a significant contributor and maintainer of major open source software including Linux, GNU, and Apache Web server. Several Red Hat engineers are prominent open source developers and members of the open source community. Red Hat, rated as CIO Insight Magazine's Most Valued Vendor for the second consecutive year, Red Hat maintains the highest value and reliability rankings among its customers, and is the most recognized Linux brand in the world. Red Hat serves global enterprises through technology and services made possible by the open source model.

As the leader of the oVirt project [43], and a major contributor to the Open Stack project [41], both of which being mainly implemented by Red Hat and specifically Red Hat Israel, Red Hat pushes the envelope in open source standards and specifically on virtualization and cloud technologies. The Red Hat virtualization team develops and maintains the QEMU/KVM hypervisor which is the engine behind most of the public clouds running today.

Based on this expertise Red Hat will be mainly contributing to WP3, by providing the hybrid infrastructure management efforts in BigDataStack enabling various deployment on the infrastructure (i.e. containers, small-footprint VMs, and full-fledged VMs) and will also participate in WP4 by providing a distributed data storage from an infrastructure point of view, thus identifying and deploying different infrastructure resources (e.g. through containers, small-footprint or full-fledged VMs) according to the storage requirements. Finally, Red Hat will participate in WP7, by realizing the European Open Source Initiative that will aim at high impact of the BigDataStack outcomes.

#### Exploitation Strategy

As the world largest open source company, RHT will push 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 BigDataStack results and contributions will be upstreamed to relevant upstream projects and RHT's saleable products (e.g. OpenShift, RHOS and more).

## 7.1.4. ATOS SPAIN SA

After the acquisition of Bull and Siemens Solutions and Services by ATOS, the new ATOS is an international information technology services company, delivering hi-tech transactional services, consulting, systems' integration and managed services. ATOS is focused on business technology that powers progress and helps organizations to create their firm of the future. It is the Worldwide Information Technology Partner for the Olympic Games and is quoted on the Paris Eurolist Market. Atos operates under the brands Atos, Atos Consulting & Technology Services, Atos Worldline and Atos WorldGrid. Atos annual revenues are of €8,7 billion, and it has over 78,500 business technologists worldwide in 42 countries. The partnership with Siemens and the acquisition of Bull position ATOS as one of the main European players in Big Data and HPC solutions.

ATOS Research & Innovation (ARI), node of ATOS R&D&I located in Spain & Turkey is a worldwide reference for innovation for the whole ATOS group. ARI performs projects usually within an international scope that combine advanced technological developments with economic exploitation. ARI work is focused on several R&D areas: financial, public sector, health, industry, energy & utilities, Telecom, IT and media. ARI plays the role of source of innovative ideas to be used by Atos sales and technical staff. AATOS is member of many international research initiatives and communities. Gartner ranks ATOS as visionary. Rather than combining a bunch of disparate solutions, we have built ATOS CODEX – an open, powerful and secure analytics platform. So, ATOS' customers can focus on the essential: get actionable insights for business value.

Moreover, ATOS participates in other Big Data projects: (i) QROWD: a H2020 ICT-14 project that is using data analytics for data integration in the smart cities and smart mobility domains, (ii) TOREADOR, a H2020 ICT-16 project from the 2015 call dealing with the automatic provision of big data frameworks and machine learning libraries (ii) VELA\_ScO: a FP7 project dealing with simulations in the engineering domain focusing on Big Data in combination with HPC, (iv) PHEME, a FP7 project which aims verify rumours in social networks combining big data analytics with advanced linguistic and visual methods, and (iiiv) LEANBIGDATA, a FP7 project which targets at building an ultra-scalable and ultra-efficient integrated big data platform.

BigDataStack developments and expected results are very much aligned with Atos overall company strategy, commercial products and research developments. Atos approach it is that new available sources of data and what we call the Economy of Data will bring an impact comparable to third industrial revolution enabling a new and more powerful role for IT, enabling deeper understanding of customers and access to them. Atos's view is that these enable the creation of new business models, not solely consisting in data reselling, but able to enable completely new market offerings.

Atos Codex is the Atos brand for provision of data analytics solutions to customers. It is a fully integrated and cross market end-to-end analytics solution that enables organizations to maximize the value of their data quickly and cost efficiently. Atos Codex provides data analytics end-to-end along the complete IT value chain. The services range from digital transformation strategy & consulting, use case business modeling, data science expertise, and agile analytics deployment, to ongoing evolution management. Specifically, Atos Codex agile analytics deployment will be very much benefited by the specific research conducted in

BigDataStack. Atos R&D capacities with regards to QoS and Cloud and Edge cluster management address long term view for the realization of aforementioned challenges not only in Retail industry, but to develop innovation in Big Data and Cloud.

Within BigDataStack project, Atos will lead WP3, while actively contributing to this WP offering its background in infrastructure and distributed management in cloud / edge environments, as well as in automated deployment and re-configurability through scalability for data analytics clusters. Furthermore, Atos will contribute to the dimensioning workbench in WP5, providing their expertise on analysis of applications and characterization of their resource needs, whereas in WP6, will lead the Use Cases Scenarios as for the analysis, the definition, the implementation and the evaluation of the use cases.

Finally, Atos will participate in WP7, in dissemination and exploitation tasks, and will be involved, as a main actor, in the activities related to the market analysis, business scenarios, and exploitation plan definition.

### **Exploitation Strategy**

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

As Big Data and cloud providers, ATOS is very interested on following the results of BigDataStack project. We are currently in the process of promoting the results of BigDataStack project internally in ATOS, discussing with our global ATOS CODEX management how to proceed to test and validate it, and potentially use it in the scope of the deployment of analytics solutions in ATOS CODEX.

ATOS' early stage individual exploitation plan is:

- Year 1: Internally, ATOS is raising expectation about BigDataStack project, in order to show preliminary results
- Year 2: The Innovation Hub team of ARI will promote the QoS Evaluator generated by ATOS within BigDataStack among the ARI team and information about the project will be spread through all Business Units of ATOS.
- Year 3: ATOS will promote the results of BigDataStack project internally in ATOS, discussing with our global ATOS CODEX management how to proceed to test and validate it, and potentially use it in the scope of the deployment of analytics solutions in ATOS CODEX

### **7.1.5. ATOS 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.

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 that is the one providing the needed dataset for the implementation of our use-case.



ATOS Worldline participation is due to the real need of its partner (EROSKI) to provide personalized experience to their users. In this context, our partner on the project, one of the top retail grocery companies in Spain, wants to be able to offer tailored recommendations to their customers based on both their previous purchases in the store and on other customer's purchases. The machine learning algorithms developed during the implementation of the retail use-case of BigDataStack are expected to be delivered to our partner so that they can use them in their current systems (ecommerce, loyalty portal, loyalty app, ...).

### **Exploitation Strategy**

BigDataStack will allow ATOS Worldline both to get a solid background on BigData technologies and strengthen its relationship with one of its customers. Furthermore, the knowledge and experienced obtained by us during the project can help ATOS Worldline to participate in other Bigdata projects for another customers.

AtoS Worldline is in charge of implementing one of the use-cases that will help to test the capabilities of BigDataStack. AtoS Worldline participates in the project along with a partner that is the one providing the needed dataset for the implementation of our use-case

ATOS Worldline role in the project is to help to test the capabilities of BigDataStack by implementing a use-case as well as by providing one of the datasets to the project. Thus, no component of the platform itself is going to be developed by AWL.

However, the following assets produced during the implementation of the use-case are expected to be exploited later:

- Specific components produced for the preprocessing of the dataset during the implementation of the use case
- Specific analytic algorithms produced for the recommender system

The recommender system to be implemented consists of two separated modules. During the data preprocessing the system is preparing the data (i.e. cleaning, filtering and normalizing it) so that the model can be run. During the execution of the model the system runs the analytical algorithms to calculate which products can be recommended to the customers.

All of the components produced for the prototypes are expected to be exportable so that our partner could use them in the implementation of a recommender system running over a commercial Big Data platform.

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.

AWL's early stage individual exploitation plan is:

- Year 1:
  - First prototype of the analytic algorithm for recommending products based on other user's behavior (cross-selling) and on the feedback received from the consumers of the recommendations.
  - First prototype of the data pre-processing module that will tackle the needed normalization of data for cross-selling and will be implemented with CEP.

- Year 2
  - First prototype of the analytic algorithm for recommending products to customers based on their history of purchases, seasonality of products, etc. among other factors
  - Second prototype of the preprocessing module integrated with WP5 BigDataStack components:
    - Process Modelling Framework
    - Process Mapping
    - Data toolkit
    - Application dimensioning framework
- Year 3: First prototype of the analytic algorithm for recommending promotions to customers based on the recommendation of products calculated by the recommender and on the feedback received from the consumers of the recommendations.

### 7.1.6. 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 will 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 offers an excellent opportunity to further improve GFT's competencies and offering in the Big Data domain.

GFT will develop the financial demonstrator. GFT will coordinate 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 a number of 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 offers an excellent opportunity to further improve GFT's competencies and offering in the Big Data domain

A major focus area is the 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 the America.

### **Exploitation Strategy**

GFT will assess the opportunity to expand its business to the trading market through the development of the financial demonstrator and further R&D. In general, GFT plans to enhance its solutions in Big Data domain with the project's results.

GFT main exploitation activities will be focused on the financial market.

GFT early stage individual exploitation plan is:

- Year 1: An internal dissemination activity of the BigDataStack solution will be performed, by exploring, evaluating, and validating the potential valuable assets from BigDataStack outcomes (tangible or intangible).
- Year 2 and Year 3: It will share the project's assets with its wide portfolio of customers.
- 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.

GFT will exploit the BigDataStack results through its wide net of customers, from insurance companies to banks, credit cards circuits, merchants and telco operators.

### **7.1.7.DANAOS SHIPPING COMPANY LIMITED**

DANAOS Shipping is the ship manager of DANAOS Corporation a leading international owner of 60 containerships. DANAOS charters out containerships to a geographically diverse group of liner companies, including most of the largest ones globally. DANAOS has established its reputation by providing first class operation services. DANAOS has been actively engaged in both bona fide and applied operation Research by his prestigious research and innovation hub, DANAOS Research Center (DRC). DANAOS has been participating in several EU projects such as FP7 Mosaic, FP7 INCASS, FP7 SHOPERA, H2020 FIBRESIP & HOLISHIP. DANAOS is a laureate member of FRANZ EDELMAN academy and winner of the homonymous award in 2012 (the highest worldwide distinction in applied operation Research).

DANAOS Shipping as an organization is envisaged in optimizing scheduling of maintenance jobs for its fleet along with the purchase cycle of the spare parts needed. BigDATASTACK maritime use case is heading at that direction. Elaborating on big data analysis for orchestrating maintenance requirements and spare parts supply along a large fleet with a high utilization rate where time for 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 is looking forward to a positive project result for further exploitation in order to achieve cost savings and operational excellence boosting up his reputation in the shipping industry. DANAOS main contribution to the project is to capitalize on shipping knowledge and provide insights and operational data for the development of the marine BIGDATASTACK

case. Tech. applications applied to DANAOS procedures in supply and technical supervision of vessels will be integrated in the solution provided.

DANAOS offers 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 Danios 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. Furthermore, regarding the preventive maintenance algorithms, DANAOS has a large dataset of main engine and vessel data coming on a peri-minute-basis from sensors on-board that can be used. Furthermore, mechanical and naval engineers working in DANAOS can provide detailed knowledge on the main engine damages for which preventive maintenance will be performed.

### **Exploitation Strategy**

There seems to be an opportunity for DANAOS to deploy and test a working prototype of a complete system that can monitor the state of a vessel and via the use of appropriate analytics inform the end user for upcoming malfunctions on the main engine of the vessel. Furthermore, through its integration with other working systems, the end user will be able to order the appropriate spare part in the ideal port with the least deviation from optimal vessel routing plan. Optimization in maintenance and ordering plan will lead to significant cost savings.

DANAOS is not developing any part of the system. they contribute to BigDataStack via offering a big dataset with actual values as produced from sensors on-board and stored in an in-house database as part of a proof of concept. Furthermore, DANAOS contributes to this 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. 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.

DANOS early stage individual exploitation plan is:

- Year 1: Assist and contribute with recommendations and all means possible for the completeness of a proof of concept. Present the first outcomes of the project within the organization and highlight its potential for use as an assisting tool.
- Year 2 and 3: Same as year 1. Once the system is ready re-assure that works properly and fulfils its objectives towards predictive maintenance and dynamic routing.
- 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.

## 7.1.8. SINGULARLOGIC ROMANIA COMPUTER APPLICATIONS SRL

SingularLogic activities focus on the development and distribution of business software applications, design and implementation of Integrated IT Solutions, including provision and support of well-established international IT products. SILO supports the following services: (a) Integrated IT Solutions for large enterprises of the private and public sector, (b) Vertical IT solutions for the Retail Sector including supply chain management and Point of Sales solutions, (c) Integration and customization of business software applications (ERP, CRM, Accounting, HRM, etc.), and recently (d) Integration and support of voice-enabled products that use speech recognition technologies.

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 BigDataStack cloud infrastructure has been made available, so in the next period, a number of network monitoring and engineering methods will be developed to support the mechanisms which enforce policies according to application or data requirements. This will enable to efficiently handle ingress and egress traffic based on specific constraints (i.e. throughput, delay, packet loss, etc.) set by the end users. The latter will be realized through containerized services.

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.

### **Exploitation Strategy**

SILO has been significantly investing in technical equipment and training, by providing adequate computing/middleware infrastructure and training in cutting-edge technologies. SILO has set up and maintains private cloud infrastructures located in Athens and Bucharest, which host its cloud-based products (including reporting/data science/analytics modules). These include the Galaxy Enterprise Suite ([www.slgalaxy.eu](http://www.slgalaxy.eu)) and 4Doctors both targeting at cloud-based data intensive operations.

SILO team is pursuing a research agenda which strengthens its current expertise with the potential to extend it towards (a) High-complexity 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.

New components will be developed in the context of BigDataStack project extending the current state of the art in information driven networking, software defined networking technologies and mechanisms to enforce policies according to data / application needs.

SILO's early stage individual exploitation plan is to build a novel framework which enforces network policies according to the requirements tailored to the data or application needs by

taking into consideration the constraints imposed by the information traveling in the BigDataStack cloud environment.

### 7.1.9.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 Applications, Human Capital Resource Management and eLearning, and Mobile Applications. The activities of the Company span among several countries in EU, Eastern Europe and CIS countries, as well as the Balkans.

The ATC Innovation Lab carries more than 25 years of expertise in Research and Development. It focuses on innovation aspects, which are often overshadowed by research concerns, and on turning promising ideas into concrete and robust products, in a cost and time-efficient manner. It discovers or conceptualizes Innovation first, then turn it into working systems through intense and continuous involvement in cutting-edge research projects, focusing primarily in areas that can offer the next big advance to its commercial offerings, discovering, at the same time, new domains and creating the next company targets.

ATC is a partner in the NESSI ETP ([www.nessi-europe.com](http://www.nessi-europe.com)) and a full member in the Big Data Value Association ([www.bdva.eu](http://www.bdva.eu)).

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 in order to remain competitive in the fast-moving sector of Big Data.

ATC demonstrates long experience in service and platform development and integration for various domains and vertical business sectors related to the analysis of large-scale data sets. Relevant experience lies both in the combination of different modules and applications in the context of platform wide functionalities, as well as the deployment of secure and privacy friendly systems and services using various technologies. Also, ATC has experience in specifications, technical validation as well as user interaction and visualisation and it is very familiar with business analysis and market uptake innovation tasks, driven communication as well as promotion and dissemination tasks.

ATC develops horizontal B2B solutions with a special focus on content and workflow management, customized to the needs of business sectors such as the media, banking and public sector organizations.

ATC leads Innovation Management supporting innovation strategy and exploitation activities of BigDataStack project in WP1 and WP5, by leading into the delivery of the BigDataStack Process Modelling Framework. ATC contributes also in WP6, regarding the integration of their respective components in the overall integrated BigDataStack reference implementation and focuses on DANAOS use case definition, implementation and evaluation, being the integrator of the real-time ship management scenario.

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 in order to remain competitive in the fast-moving sector of Big Data, while it wants also to enhance specific media products (i.e. NewsAsset: [www.newasset.com](http://www.newasset.com), TruthNest: [www.truthnest.com](http://www.truthnest.com)) with advanced visualization and process modelling functionalities.

### **Exploitation Strategy**

ATC, based on their expertise and commercial offerings in interfaces and business processes optimization, is responsible for the BigDataStack Process Modelling Framework - which allows the specification of processes in a declarative and flexible way- and the adaptable visualization environment – which 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 both during benchmarking and during their operation.

ATC's commercial products already handle massive amounts of data and can significantly benefit through the incorporation of advanced big data technologies. Specific examples are NewsAsset ([www.newasset.com](http://www.newasset.com)) and TruthNest ([www.truthnest.com](http://www.truthnest.com)), ATC's products for organizing the content and workflow of media organizations. Such applications make extensive use of media data, therefore making the BigDataStack outcomes excellent add-ons for improving the management and analysis of such data. Since it is difficult to develop a single tool that will cover the needs of the different sectors in which ATC is active, ATC also envisages to further customize and adapt the BigDataStack outcomes to specific business processes in various sectors. More specific, ATC has been working very closely with the financial and banking sector for developing customized tools to their needs and requirements. ATC will pick up ExaFin tools and exploit them with its current networks of customers in this sector, by proposing new business solutions to them based on outcomes and results of this project. Apart from the domestic market in Greece, ATC has a specific focus on the eastern European market. As the BigDataStack can also be exploited in other markets, ATC sees the opportunity of expanding both its market share in established markets, as well as its reach to other markets, either by offering an enhanced own product portfolio or by marketing independent tools that can be incorporated in other companies' software suites.

The core of the exploitation actions of ATC will be focused on the business of the integration and adaptation of BigDataStack developments in order to enhance ATC's portfolio of products and services. Also, BigDataStack outcomes will be supported by strong sales and consultants' teams, which will grow gradually and increase the awareness of the BigDataStack offerings.

ATC's early stage individual exploitation plan is:

- Year1: During the first year, ATC will increase awareness of the BigDataStack solutions especially in ATC Sales Department.
- Year 2: During the second year, ATC will get the sales networks of ATC and its partners familiar with BigDataStack approach, generating also interest from larger partners, through Internet, partners' networks, ATC Sales Department and Existing Customer Base. The whole approach will be presented as an introduction of new innovative technologies and solutions, understanding of new capabilities and market advantages. There will be the first attempts of customization of solutions to partners' needs and an increased awareness of target market.

- Year 3: During the third year ATC will expand the tentative selling base. There will be a further investigation for possible new features and adaptations. Support services for training personnel will be available.

### 7.1.10. LEANXCALE SL

LeanXcale is a startup that commercializes an ultra-scalable operational database. LeanXcale was incorporated a couple of years ago. LeanXcale has been awarded in the EIT ICT Labs Idea Challenge startup competition in the Future Cloud topic two years ago. Last year, the company was one of the 10 finalists in the South Summit European startup competition in the area of B2B. This year it has been selected as one of the 100 most innovative startups by Red Herring.

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.

#### Exploitation Strategy

LeanXcale bases its exploitation strategy in the incorporation of the features and components developed within the project (Elasticity for LeanXcale database, Integration of LeanXcale database with IBM Object Store and Integration of LeanXcale database with UPM CEP) in its commercial solutions.

LeanXcale plans a proprietary license for their SW developments.

LeanXcale's early stage individual exploitation plan is based on two main activities:

- Adopt the developed technology as part of the current product portfolio (KiVi and LeanXcale database).
- Explore partnerships with partners:
  - With IBM to have an integrated solution for real-time analytics (LeanXcale database+IBM object store).
  - With UPM to have an integrated solution for a big data platform combining streaming data and data at rest.
  - With Atos and GFT as sales channels partner.
  - With the use case providers as potential customers.

### 7.1.11. 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 with 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.

The BigDataStack cloud infrastructure has been made available, set up and running. In the next period, a number of operating instances on the support of attributes definition for the user-defined analytics constraints/requirements through the BigDataStack toolkit and functionalities regarding the information driven networking will be deployed on the BigDataStack environment. The latter will be realized through containerized services.

Opportunities for UBITECH are lying in the area of data intensive applications and the need to efficiently orchestrate diverse network resources, user and applications needs. The experience gained in the context of BigDataStack project will facilitate to deliver sophisticated technologies for traffic engineering and network management for demanding data operations.

### **Exploitation Strategy**

UBITECH has deployed and maintains in its firewall protected computer room a highly-available experimental Cloud Infrastructure which includes 60 cores at 2.4GHz (with virtualization capabilities) supported by a Network Attached Storage (NAS) infrastructure having an effective capacity of 25 Tb of data. Currently, UBITECH is deploying KVM-based Hypervisor on top Metal-As-A-Service framework (OpenStack). Currently, an instance of this environment has been made available for the purposes of BigDataStack testbed. A set of services regarding the BigDataStack toolkit and information driven networking will be exploited in the direction to support data scientists and practitioners to handle, set up and automate intensive data analytics tasks and demanding Big Data applications.

UBITECH team is pursuing a research agenda with an emphasis on delivering tangible assets and reusable components for EU's data economy. This is the reason why the team is highly interested in these assets in order to extend its current strategy for research and innovation reinforcing its internal capacity, experience and technological know-how in new emerging technologies addressing specific user needs.

New components will be developed in the context of BigDataStack extending the current state of the art in information driven networking and data tools targeting the seamless conduct of advanced data analytics tasks.

UBITECH's early stage individual exploitation plan is

Year 1: The exploitation activities will be spearheaded by a SaaS platform, which will provide intensive data analysis services for real-time ship management, intelligent multi-channel banking and improved consumers' shopping experience through connected services. Given the time frame of the project, these services will be available by end of Y1, and thus have a competitive advantage in know-how, and a sizeable head-start for establishing a successful

business activity on data analysis services for the real-world use cases coupled within the BigDataStack offerings.

Year 2 and Year 3: UBITECH aims to conduct in close collaboration with the consortium partners extensive market monitoring in order to set directions to the required actions and development activities towards the exploitation of the first BigDataStack results. Adjustments will likewise propel use case partners, data scientists and practitioners to early adopt the BigDataStack solutions.

### 7.1.12. 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) high Level expert group and is a member of ECSO the European cybersecurity organization and works on cybersecurity & risk management and is a founding partner of the WISER "Wide-impact cyber security 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) & delivers some of the technical implementation specifications to these products & services within the WISER Innovation Action project.

Trust-It is the coordinator of key projects in the field of Data infrastructures and Cloud Computing such as RDA Europe 2.0 (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 CLARUS, WISER, SLA-Ready, 5GENSURE, HELIX NEBULA & PICSE.

Trust-IT is involved in a number of initiatives behind supporting a European Open Science Cloud (EOSC) in Europe. It leads the communication activities on the EOSCPilot & EOSC-Hub projects and Trust-IT's CEO is the chair of the 2nd EOSC HLEG group until end of Dec 2018. 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 components to have visibility via the EOSC portal would be extremely important.

#### Exploitation Strategy

Trust-IT is committed to drive forward new ideas, pinpoint market opportunities, create multi-stakeholder partnership and build international networks and communities.

With over 15 years of experience in the analysis of complex ICT systems and digital community development, Trust-IT sees that it plays a central part in the BigDataStack stakeholder engagement strategy and project impact and exploitation activities, building on its competences for the creation of a solid foundation of collaborative knowledge and diversity

enhancement, made up of research institutions, industry & SMEs public authorities and, finally, citizens – especially around the topics of Data infrastructures, ICT Standards, cloud computing, next generation internet topics,

Trust-IT can mobilize 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.

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.

It will leverage its STANDICT.eu network (an initiative to attract European individuals working on ICT standards efforts) to see if any standardization work within BigDataStack could be carried out in one of the open calls around Big Data, Private data.

### 7.1.13. 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 is 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 for front of technology and UPRC as a research centre is interested in this field of technology. BigDataStack allows 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

UPRC contributes to many technological projects on national and international levels. The experience gained during all those projects, the diverse team of UPRC (in terms of expertise), composed of passionate professors, students, PHD candidates and the technical environment are UPRC's main means for achieving the objectives of BigDataStack.

UPRC participation in BigDataStack project is in the context of its strategic plans in the areas of: (i) Education: the BigDataStack results will be proliferated among the attendants of the University activities, mainly among postgraduate and continuing education programs due to the advanced nature of the topics, (ii) Technology transfer to the Greek IT industry, offering technology transfer services to companies and public bodies through joint projects, and (iii) Technology promotion in the Greek industry as part of an effort to increase the adoption of BigDataStack technologies. Moreover, UPRC has close and strong collaboration with commercial, industrial and public organizations providing specialized scientific expertise and innovation to improve and enhance products and services.

#### **Exploitation Strategy**

UPRC will aim at exploitation the following assets:

- Data monitoring engine
- Data quality assessment & improvement framework

- Predictive process mining mechanism
- Process mapping mechanism
- Application dimensioning workbench with an emphasis on data-intensive applications and processes

For each asset identified, UPRC plans to exploit in the next ways:

- Data quality assessment & improvement

The analysis of data and, consequently, big data passes through several distinct phases, with each of them posing different challenges. In contrast to the much more researched modelling phase, the cleaning step is often seen as a sore point, even though without it the modelling phase is of little use. For example, consider the case of medical data, where a central processing unit collects information from many sensors, to come to a conclusion. If the data are too noisy, or even flawed because of a damaged source, the conclusions might be associated with a high level of uncertainty or even reduced to garbage.

- Process mining and mapping

Modern information systems maintain detailed trails of the business processes they support, including records of key process execution events, such as the creation of a case or the execution of a task within an ongoing case. Process mining techniques allow analysts to extract insights about the actual performance of a process from collections of such event records.

Current data mining and machine learning tools are characterized by a plethora of algorithms but a lack of guidelines to select the right method according to the nature of the problem under analysis. Since real-world applications are generally time-sensitive, engineers tend to use only a few available algorithms for data analysis, hoping that the set of assumptions embedded in these algorithms will match the characteristics of the data. Such practices in data mining and the application of machine learning has spurred the research community to investigate whether learning from data is made of a single operational layer or whether there are in fact several operational layers that can be exploited to produce an increase in performance over time. The latter alternative implies that it should be possible to learn about the learning process itself, and that a system could learn to profit from previous experience to generate additional knowledge that can simplify the automatic selection of efficient algorithms.

- Application dimensioning workbench

Application analysis is a very important step for enabling any type of optimization and trade-off investigation. Moreover, all applications tend to be composite applications that consist of micro-services. In this context and given the data-intensive nature of all applications, the application dimensioning workbench will allow the identification of dependencies between different atomic services as well as the estimation and prediction of the required infrastructure resources related to the application as well as to the data services.

And finally, UPRC's early stage individual exploitation plan for each component is:

- Data Monitoring

Year 1: Deploy the test version of the Triple monitoring engine and Develop spark exporter for Prometheus using spark measure project.

Year 2: Create a wrapper exporter (Prometheus compatible) for application running in container (docker)

Year 3: Create a unified package for the triple monitoring engine using containers technology and make the triple monitoring engine package available to open source community.

- Process Mining and Mapping
  - Year 1: Develop basic process mining algorithm.
  - Year 2: Apply process mining algorithm to use cases project created processes.
  - Year3: Reach publishable prototype level and release as open-source.
- Data Quality Assessment & Improvement
  - Year 1: Research available methods and related work
  - Year 2: First prototype
  - Year 3: Final working solution available as open source and applied to different datasets.
- Application Dimensioning Workbench
  - Year 1: Patterns and benchmarks applicable to different data and application services
  - Year 2: Prototype of the workbench with load injectors for the identified different data and application services
  - Year 3: Open source, ready-to-use solution evaluated in different real-world use cases.

#### 7.1.14. 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, that 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.

#### Exploitation Strategy

UPM aims at exploiting its components: CEP, the integration of the CEP component with LeanXcale DataBase and the Triple Monitoring of the CEP.

And 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.

The commercialization roadmap for the CEP component will be undertaken by LeanXcale and is not defined yet at this phase of the project.

### 7.1.15. 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.

As an information retrieval group, CLA is primarily interested in how to develop technologies for the efficient and effective extraction of value from big data (streams). For example, GLA is the developer and maintainer of the open source Terrier search engine (terrier.org). Indeed, GLA has expertise in a range of IR technologies, such as search, recommendation, classification and machine learning.

Within BigDataStack, GLA is invited to provide expertise for identifying effective deployment configurations for user applications. When a user wishes to deploy a new application, GLA often will not know what (virtualized) hardware we need for that application. Hence, there is a need for automatic systems that can predict a good configuration given incomplete information. Within the project, we lead the development of automatic machine-learned approaches for this task as part of WP3.

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.

GLA's participation in BigDataStack is part of an exploration of the future research directions for IR. In particular, GLA is examining domains where IR technologies are not yet being fully used but may bring additional value. In this case, GLA is building on item ranking technologies originally developed for search engines and adapting them to rank application deployment options. Indeed, deployment trade-offs between effectiveness (quality of service) and costs of the deployment is new active research area. GLA aims to explore some of the interesting research challenges in this space, in terms of representing the complex application and user needs/preferences and mapping those into general and cost-effective deployment ranking approaches.

#### **Exploitation Strategy**

GLA is responsible of the component Deployment recommendation service, and GLA is focused on publishing novel insights gained using it as a base, as well as demonstrating its research on concrete use-cases and big data.

This component is expected to be released under an open source licence.

GLA's early stage individual exploitation plan is:

- Year 1: Working on the development of component
- Year 2: Public release of application deployment ranking module for Terrier, and deployment of the proposed ranking technology on the ATOS Worldline and DANAOS use-cases
- Year 3: Advancing the core ranking technology to provide explainable recommendations for cluster-state sensitive application deployments and enable knowledge transfer with other end-user partners out-with the current project, e.g. partners from the EPSRC Closed-loop data science project.

## 7.2. Joint Exploitation Strategy

In this section, an initial analysis of a future Joint Exploitation Plan will be outlined. Even though the project is in an early stage to define a Joint Exploitation, here we present all the activities that will be detailed in depth in subsequent deliverables.

BigDataStack consortium is composed of almost 80% Industrial partners, that are industrial partners companies interested in a commercial exploitation of the BigDataStack framework, use case partners interested in exploiting BigDataStack in their own businesses, and finally Academia partners interested in dissemination of scientific relevance results.

Because of this different partners' profiles, the joint exploitation activities could be classified into:

- Commercial exploitation of the framework under one of the Business Models proposed in Section 5. The Business model to exploit the BigDataStack Infrastructure Management System must be selected by the partners.
- Evaluation of the BigDataStack framework, and potential customization to scientific customer needs.
- Continue participation in events to promote BigDataStack framework: workshops, seminars, trainings, conferences.
- Publication of joint articles, case studies in the scientific journals.
- Creation of a spin-off company that will take care of the BigDataStack evolution, maintenance, sales, etc. This would be a separate entity because of project outcomes.

The role of the partners in the joint exploitation strategy will be collected using a template elaborated by the Exploitation Team. This template will collect the partners' interest and involvement in the different possible joint exploitation activities. As said before, in the consortium, different possible roles are presented:

- Industrial Partners (IBM, NEC, RHT, ATOS, SILO, ATC, LXS, UBI) interested in a possible commercial exploitation of BigDataStack framework integrating the know-how or SW developments into their existing offering or establish a new business line starting from BigDataStack project.
- Academic Partners (UPRC, UPM, GLA) interested in dissemination, training, and increasing awareness and interest of Big Data solutions in SMEs.

- Use Case Partners (DANAOS, GFT) interested in the analysis of their data and use BigDataStack platform to implement Big Data features to their businesses and daily operations.

Once the organization model and partners' roles have defined, three activities will started as soon as possible:

1. Elaboration of an Exploitation Agreement which will detail all the partners contribution as well as partners' rights and obligations
2. Explore the possibility of creation of a BigDataStack Foundation to foster the joint exploitation activities.
3. Start working on pre-commercialization activities working along with the Dissemination Team.



## 8 Conclusions

---

During the first period of the BigDataStack project, exploitation activities have been mainly focused on three aspects:

- Define the market context for BigDataStack outcomes
- Identify and describe BigDataStack SW components developed by the partners  
Identify the initial partners' exploitation interest and intentions discovered during the first months of the project.

This deliverable aims at being the starting point for the next planned exploitation activities which will improve the sections included in this first report.

Next activities exploitation activities planned will be:

- Update the Market analysis, focusing on the market context for the specific SW components.
- Market analysis of the use cases sectors: Maritime, Consumers and Banking, and benefits obtained with the implementation of BigDataStack outcomes to their businesses.
- Further investigation on the three business models identified for BigDataStack outcomes and describe more exhaustively the Business Model Canvas components.
- Define the Joint Exploitation Plan agreed with partners and discussed in an Exploitation Workshop.
- Work closely with the ICB to define a realistic Roadmap adoption.

The outcomes of these activities will be reported in the last deliverable due in M34, D7.3 Exploitation Plan and Business Potential V2.