

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	http://bigdatastack.eu/

D7.5 Dissemination, Standardisation and Adoption Roadmap

Work Package	WP7 – Communication, Exploitation, Standardisation & Roadmapping
Lead Author (Org)	Marieke Willems, Andrea Schillaci, Diego Simoni (TRUST-IT)
Contributing Author(s) (Org)	Silvana Muscella (TRUST-IT), Luis Tomas (REDHAT), Ana Bélen (ATOS), and all partners with their respective activities
Due Date	31.10.2020
Date	08.12.2020
Version	1.0

Versioning and contribution history

Version	Date	Author	Notes
0.1	21.09.2020	Marieke Willems (TRUST-IT)	ToC and structure defined
0.2	8.10.2020	Luis Tomás Bolívar (Red Hat)	Updates for section 3, and added comments
0.3	20.10.2020	Andrea Schilacci, Diego Simoni and Marieke Willems (TRUST-IT)	Sections 1 and 2 and annexes
0.4	25.10.2020	Ana Belen Gonzalez (Atos)	Section 4
0.5	01.12.2020	Marieke Willems (TRUST-IT)	Section 5, overall alignment and executive summary
0.6	02.12.2020	Silvana Muscella (TRUST-IT)	General Review
0.7	03.12.2020	Marieke Willems & Andrea Schilacci (TRUST-IT)	Revisions
0.8	04.12.2020	Yosef Moatti (IBM)	Internal review
0.9	04.12.2020	Ana Belen Gonzalez (Atos), Luis Tomás Bolívar (Red Hat), Marieke Willems & Andrea Schilacci (TRUST-IT)	Version addressing internal review comments
1.0	08.12.2020	Marieke Willems (TRUST-IT)	Final version

Table of Contents

1	Introduction	7
2	BigDataStack Dissemination M1-M36	8
2.1	COVID-19 contingency planning for events.....	8
2.2	Promoting BigDataStack key assets	10
2.3	Promoting BigDataStack Use Cases.....	14
2.4	Onboarding end-users	16
2.5	Promoting BigDatStack Open Source activities.....	20
2.6	Horizontal activities	21
2.6.1	Website	21
2.6.2	Videos	24
2.6.3	Publications & ZENODO.....	25
2.6.4	Social Media Channels	26
2.6.5	Events, webinars and synergies	29
2.6.6	Newsletter.....	35
3	Standards & Open Source Contributions: European Open Source Initiative (M1-M36) 36	
3.1	Realising the European Open Source Initiative to increase the impact of the BigDataStack outcomes	36
3.1.1	Research Days	36
3.1.2	Upstream Contributions	37
3.1.3	Blog Posting	38
3.1.4	OutReachy Internships	38
3.2	BigDataStack Alignment global standards and standardisation of project results	39
3.3	ICT standardisation in support of EU policies.....	41
3.3.1	A common standard for the referencing of open data	41
3.3.2	Analytics insights and scaling policies for microservices	42
4	Adoption Roadmap.....	43
4.1	Big Data adoption challenges and barriers.....	43
4.2	Application of Big Data and challenges by Industry Verticals	47
4.3	Adoption Roadmap for BigDataStack offering.....	48
4.4	Activities to validate the BigDataStack adoption roadmap.....	51
5	Lessons learned and legacy	54
6	Annex 1 - BigDataStack publications.....	55
7	Annex 2 - BigDataStack at events	57
8	Annex 3 - KPI check.....	65

List of Tables

Table 1: COVID-19 caused challenges and how BigDataStack made use of opportunities and impact achieved 10

Table 2: Overview of BigDataStack (co-)organised webinars with date and number of participants ... 35

Table 3: BDA Applications and Challenges by vertical sectors 48

List of Figures

Figure 1: BigDataStack Key-assets 10

Figure 2: Sample of BigDataStack catalogue of software components..... 11

Figure 3: Sample of BigDataStack videos on explainer software components 12

Figure 4: BigDataStack infographic highlighting assets..... 12

Figure 5: BigDataStack infographic on user journey 13

Figure 6: Sample tweet in BigDataStack strategy highlighting project assets..... 14

Figure 7: Infographic on BigDataStack use cases 14

Figure 8: Sample of dedicated and BigDataStack branded use case videos..... 15

Figure 9: BigDataStack branded flyers highlighting tailored to each of the use cases..... 15

Figure 10: Overview of achievements PPC campaign on the BigDataStack branded use case videos 16

Figure 11: Online marketing package set in place for each of the use-cases 16

Figure 12: Sample of the BigDataStack on the Innovation Marketplace 17

Figure 13: Panel questions to highlight the added value of Open Source for the digital future of Europe. 20

Figure 14: Sample of the BigDataStack project website..... 21

Figure 15: Sample of the deliverables dedicated webpage, highlighting the filter function per WorkPackage 22

Figure 16: Sample of the dedicate webpage on BigDataStack liaisons 23

Figure 17: Dashboard monitoring website traffic. 24

Figure 18: Sample of dedicated video webpage, with filtering options according to the category of the video. 25

Figure 19: OpenAIRE explore BigDataStack publication overview 26

Figure 20: Visualisation of the Deliverable section of the website in 2020 26

Figure 21: Sample of 2 tweets illustrating the social media strategy with tailored, branded and content rich messages 27

Figure 22: Overview of social media visibility of September 2020..... 27

Figure 23: Overview business roles BigDataStack LinkedIn community..... 28

Figure 24: Example of a BigDataStack multiplier, in this case on Big Data Standardisation and Open Source 28

Figure 25: Examples of twitter posts connecting to multipliers in the filed 29

Figure 26: Promotional image BigDataStack final event 29

Figure 27: BigDataPilotDemoDays joint branding 31

Figure 28: Promotional image joint event BigDataStack, I-BiDaaS and Track & Know 32

Figure 29: Jointly branded promotional image for the INFINITECH, BigDataStack and BDVA Task force event..... 33

Figure 30: Overview of sectors represented by stakeholder engaged in the BigDataStack webinars . 34

Figure 31: Revenues and cost using BDA 45

Figure 32: Problems envisioned by companies when using big data technologies/analytics 46

Figure 33: Adoption Roadmap for BDA projects..... 49

Figure 34: Adoption Roadmap 51

Executive Summary

This report is scoped under Work Package 7 “Communication, Exploitation, Standardisation & Roadmapping” of BigDataStack project. The main objectives of WP7 are to:

- provide a practical **business model** to which the project’s stakeholders may effectively use with a roll-out plan to different target entities such as decision makers, big data programmers and practitioners, research community, standardisation bodies,
- ensure an effective **communication strategy and plan** has been put in place,
- facilitate and increase the uptake of BigDataStack results **through standardisation and the European Open Source Initiative.**

In order to achieve these objectives, the BigDataStack consortium has carried out several activities, which are reflected in this deliverable.

Dissemination actions are key for positioning the project in the dynamic and rapidly evolving big data ecosystem, to federate big data solution providers communities around the project outputs, to consider local ecosystems and users’ needs in the pilots and to reach to the future innovators of the big data ecosystems such as entrepreneurs, start-ups and SMEs. This deliverable describes the actions carried out and the impact achieved on disseminating the BigDataStack work in three main pillars, set out:

- BigDataStack Assets & SW Components;
- BigDataStack Use Cases and uptake from users;
- BigDataStack contributing to the European Open Source Initiative.

All of the above has contributed to the overall Innovation Management of the project that used the three pillars indicated above to disseminate these results through the pertinent channels of the project.

To support these dissemination and stakeholder engagement activities, horizontal dissemination mechanisms were set in place such as the website, social media, audio-visual and print material production, continuous social media promotion, scientific and non-scientific publications, blog posting and presence at and organisation of events.

One of the standards for software development and products is the “open source way”. Open Source attracts tremendous technical talent and triggers organisations’ support. Moreover, and very recently the European Commission has adopted its new open source software strategy for 2020 – 2023¹ where the internal strategy, under the theme “Think Open”, sets out a vision for encouraging and leveraging the transformative, innovative and collaborative power of open source, its principles and development practices. It promotes the sharing and reuse of software solutions, knowledge and expertise, to deliver better European services that benefit society and lower costs to that society. The Commission commits to increasing its use of open source not only in

¹ https://ec.europa.eu/info/news/european-commission-adopts-new-open-source-software-strategy-2020-2023-2020-oct-20_en

practical areas such as IT, but also in areas where it can be strategic. For the benefit of the Big Data Stack project, the strategy is well aligned with its principles as the collaboration across the Commission, with Member States, companies and the public at large for building new, innovative digital solutions that work across borders and towards technological sovereignty is key – these are paramount in the three pillars mentioned above.

Several innovation activities are taking place in open source communities from individuals contributing on a global scale. Successful examples are Linux, OpenStack, Kubernetes, Git. The BigDataStack project is relying on many successful open source projects and products, especially the ones targeting cloud computing. It is using the 'de-facto' standard tools for the infrastructure to ensure easy adoption of the developed tools, namely OpenStack for the Cloud Infrastructure and OpenShift/Kubernetes for the Container Orchestration Engine. This deliverable describes the efforts of the consortium in not only in engaging in upstream communities but also in active efforts in supporting them and then cross collaborating the services into the results of BigDataStack.

The Open Source Initiative efforts of the BigDataStack project collaborate with the kick off of the "Red Hat Research Day"². The now periodic event gathers both developers and researchers, with Red Hat engineers, customers and partners, to move research ideas into the open source communities, helping to better highlight the benefits of making research together with Open Source tooling.

Finally, an adoption roadmap has been defined for the adoption of BigDataStack solution, identifying the barriers to overcome and other application sectors where the project solutions add value and where it can make savings in terms of resources, time and costs. The adoption roadmap is split into six phases in which you can adopt BigDataStack as a whole solution or the different BigDataStack blocks within some of the roadmap phases. Two internal exploitation workshops and an external expert discussion during the final project workshop validated these roadmap principles. This deliverable includes the main conclusions of the expert discussion, more extensively described in D7.3 Exploitation Plan and Business Potential Y3.

² <https://research.redhat.com/research-day/>

1 Introduction

As one of the key outcomes (in terms of reporting) of Work Package 7 “Communication, Exploitation, Standardisation & Roadmapping”, this report aims at contributing to the objectives of WP7, which are summarized as follows:

- Provide a practical **business model** to which the project’s stakeholders may effectively use with a roll-out plan to different target entities such as decision makers, big data programmers and practitioners, research community, standardisation bodies.
- Ensuring proper communication of BigDataStack outputs, outreach & stakeholder engagement and subsequently raising awareness to the scientific, industrial, and general public communities with the inclusion of three targeted workshops aimed at reinforcing user needs & their results.
- Studying the external scenario for BigDataStack results, providing input and requirements related to market needs and trends and defining the market context for exploitation and positioning against competing solutions.
- Maximising the impact of the **project aligning business opportunities and the roll-out of a credible business model** with the technical and research activity.
- Assisting and complementing the technical development with the business perspective particularly, relating to future uptake and sustainability, feeding from stakeholder needs is a vital exercise.
- Ensure an effective **communication strategy** is put in place.
- Facilitating and increasing the uptake of BigDataStack results, a specific task on **standardisation** was set in place. Specifically:
 - Following, contributing to promoting and ensuring usage of the corresponding relevant standards, while also supporting the liaison and collaboration activities with other EC funded related projects and initiatives. Scoping the European Open Source Initiative that will aim at high impact of the BigDataStack outcomes through relevant contributions to open-source projects.
 - Ensuring that the project work uses and is in line with the relevant global standards.
 - Fostering dialogues with relevant bodies in order to share potential findings and innovations made by BigDataStack that could contribute to standards.
 - Contributing towards open-source standards following the research outcomes of the projects

This deliverable describes the activities performed working towards the above objectives and the impact achieved during the BigDataStack project lifetime.

2 BigDataStack Dissemination M1-M36

Dissemination actions are key for positioning the project in the dynamic and rapidly evolving big data ecosystem, to federate big data solution providers communities around the project outputs, to consider local ecosystems and users' needs in the pilots and to reach to the future innovators of the big data ecosystems such as entrepreneurs, start-ups and SMEs.

At the start of the BigDataStack project, the following key pillars were defined for the dissemination activities:

- Integrating the project into the global big data ecosystem
- Federating big data innovator communities
- Engaging with relevant stakeholders' in the use cases of BigDataStack:
- Disseminating technical results
- Reaching data providers
- Supporting the project's commercialisation and market uptake strategy
- Marketing campaigns for the project offerings

D7.4 Dissemination and Standardisation³ clearly defined the phased approach of the Dissemination and Communication strategy followed in the BigDataStack project:

- ✓ Raising awareness
- ✓ Sharing results
- ✓ Promote results for exploitation and standardisation.

Following this phased approach, BigDataStack communication and dissemination leveraged on the strategies set in place in D7.1 Initial publication package⁴ and D7.2 Exploitation plan and business potential⁵ and D7.4 Dissemination and Standardisation⁶.

The BigDataStack dissemination and stakeholder engagement has followed a Specific, Measurable, Achievable, Realistic, and Time phased (SMART) approach. A set of ambitious KPIs were defined for the project's **Communication Kit and overall visibility, Stakeholder Engagement and Community Development** and Dissemination. The consortium has collaborated extensively to deliver on these KPIs and increase impact. An overview of the KPIs can be consulted in Annex 3.

2.1 COVID-19 contingency planning for events

In March 2020 the World Health Organisation stated COVID-19 officially as a pandemic and measurements were taken all over the globe to minimise the impact of the global health crisis.

³ <https://bigdatastack.eu/deliverables/d74---dissemination-and-standardisation>

⁴ <https://bigdatastack.eu/deliverables/d71---initial-publication-package>

⁵ <https://bigdatastack.eu/deliverables/d72---exploitation-plan-and-business-potential>

⁶ [D7.4 – Dissemination and Standardisation](#)

Today, communities of researchers around the world⁷ are collaborating to address the COVID-19 crisis⁸ - either directly to meet the immediate health challenges, or indirectly to build ecosystems that support and enable communities in a post-COVID world. Nevertheless, the pandemic also shows that despite its efforts the European Union is not yet ready to make full use of the combination of big data and, say for example, AI because of its strong data protection rules restrict the collection, use, and sharing of data which can hamper the bloc's efforts to address the spread of the disease⁹ and monitor citizen health through contact tracing. Already in April 2020, part of this EU-wide data sharing mandate was addressed when EMBL's European Bioinformatics Institute (EMBL-EBI) and partners launched the [COVID-19 Data Portal](https://www.covid19dataportal.org)¹⁰, which enables the sharing and analysis of data related to SARS-CoV-2. The initiative aims to facilitate international collaboration to accelerate scientific discovery, monitor the pandemic and help develop treatments and a vaccine for the new coronavirus.

Reshaping all physical events took immediate effect and the stakeholder engagement activities defined for the 3rd year of the BigDataStack project, namely those activities planned for face-to-face events took on an agile, flexible approach.

BigDataStack WP7 has mitigated the risks and turned challenges into opportunities for synergies with other European Commission initiatives, networks and key players in the field. The table below shows the challenges faced and the solutions found.

Challenge & expected impact	Opportunity & impact achieved
3 F2F workshops with 50-100 participants from end-user communities	<p>1 final online workshop at EBDVF2020, with BigData experts from the field and the BDVA task Force on standardisation: attended by over seventy participants, all resources are made accessible from the project website.</p> <p>1 joint webinar with INFINITECH & BDVA Task Force on FinTech and InsuranceTech: attended by 150 participants, recordings are made accessible from the project and on the BDVA website.</p>
1 properly decked booth	1 online booth at EBDVF2020, 50 direct messages were sent to EBDVF2020 attendees via the whova conference platform ¹¹ .
3 demos & 1 training	1 joint series of 9 demos with 3 H2020 Big Data PPP projects @BDV PPP– 2 demos by BigDataStack, 1 demo by Policy Cloud showcasing adoption of 2 BigDataStack components.

⁷ <https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/>

⁸ <https://www.datainnovation.org/2020/03/how-artificial-intelligence-is-aiding-the-fight-against-coronavirus/>

⁹ <https://www.datainnovation.org/2020/03/COVID-19-crisis-shows-limits-of-eu-data-protection-rules-and-ai-readiness/>

¹⁰ <https://www.covid19dataportal.org>

¹¹ <https://whova.com/>

	<p>Attended by over 420 people from Research & Academia, EC initiatives and Industry. All resources are made available for reuse, from the 4 project and BDVA websites.</p> <p>1 Kubernetes training, attended by IBM and BigDataStack partners, all resources are made available for reuse.</p>
ICT2020 cancelled	<p>1 joint workshop with ICT 14 & 16 @EBDVF2020: 117 people attended from Research and Academia and Industry. A joint report is underway and all resources are made available for reuse.</p>

Table 1: COVID-19 caused challenges and how BigDataStack made use of opportunities and impact achieved

2.2 Promoting BigDataStack key 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. The assets of the BigDataStack project are at the centre of the project’s activities in dissemination and standardisation.



Figure 1: BigDataStack Key-assets

Website & Showcased Assets

Project partners collaborated to showcase the full architecture in the shape of a catalogue of software components¹², clearly mirroring the architecture of services. The catalogue presents the software components in the BigDataStack solution with BigDataStack branded icons, for each of the software components downloadable and branded fact sheets, branded videos and information on their licenses and codes.

A series of news-items were published and promoted via social media to highlight the assets and their added value for end-users as well as major achievements in their developments and availability were highlighted (e.g. the open source contributions of the data skipping component to Spark 3.0, the inclusion of the seamless component in the EC innovation radar and partner companies with BigDataStack related Generally Available (GA-ed) products/services).

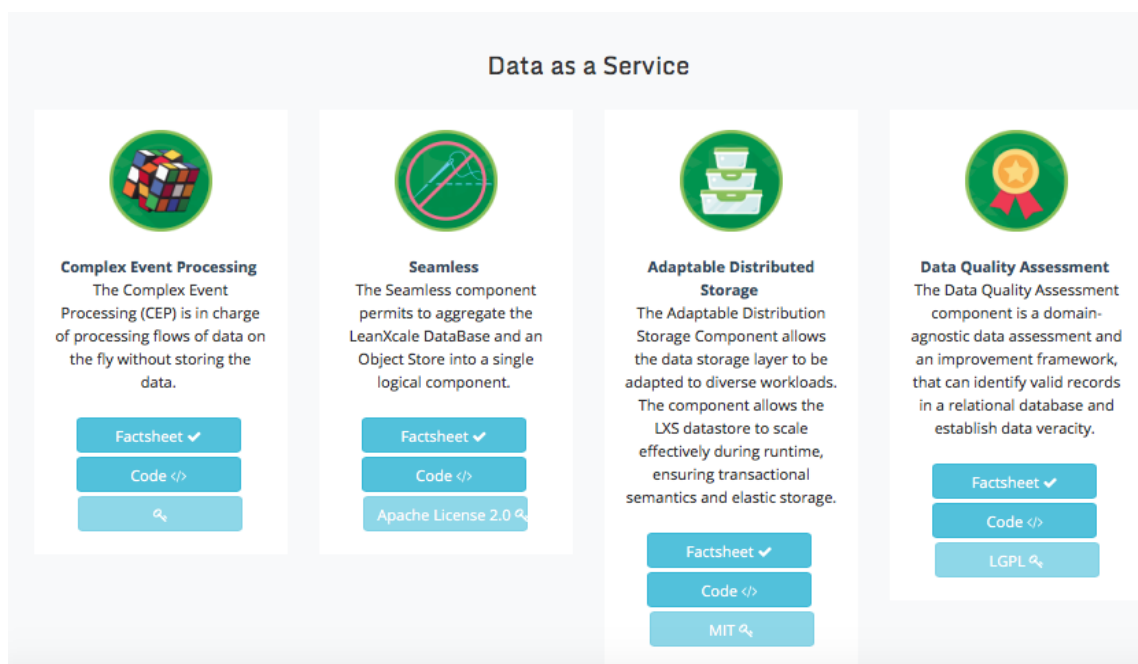


Figure 2: Sample of BigDataStack catalogue of software components

Videos

Partners collaborated on a series of videos explaining and promoting the individual software components by highlighting the specific end-user and the challenge addressed. In total 13 videos were created, branded with the BigDataStack identity, published on a dedicated YouTube playlist, embedded in the relevant **catalogue software component page** and finally regularly promoted on the website in news items and on social media and via the BigDataStack newsletter.

¹² <https://bigdatastack.eu/the-bigdatastack-solution>



Figure 3: Sample of BigDataStack videos on explainer software components

They videos collected a total of 2000 organic views on YouTube and 35k+ views via dedicated promotional campaigns on LinkedIn.

Demonstrations and webinars

During the joint series of BigDataPilotDemoDays with H2020 projects Track & Know, I-BiDaaS and Policy Cloud at the virtual BDV PPP 2020, short demos were given of the Data Visualisation Component and Dimensioning Workbench in the context of the presented use cases. Again, during the flagship event EBDVF2020 a demonstration was given of the User Interface.

During the joint EBDVF2020 session with Track & Know and I-BiDaaS projects, infographics aligned and developed as a joint effort, highlighted yet again BigDataStack software components and showcased the Deployment Recommendation Service. Recordings and slides are published and promoted via the BigDataStack channels.

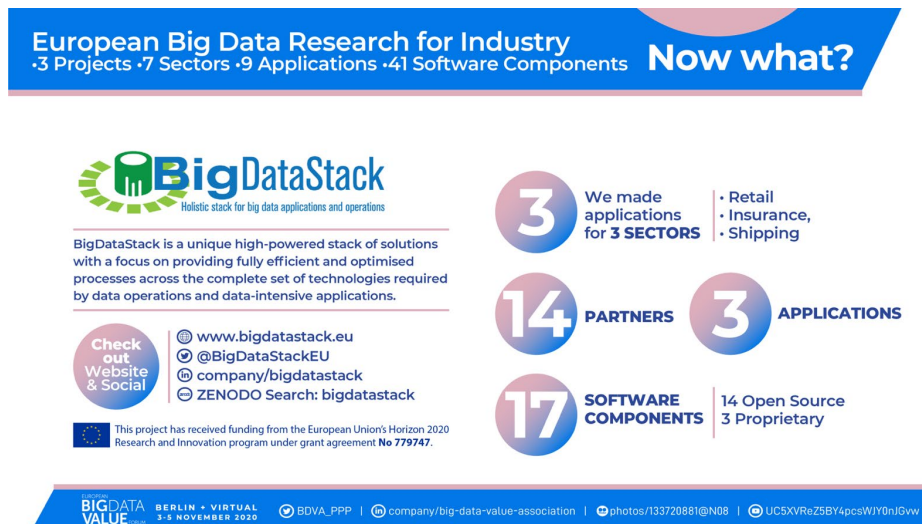


Figure 4: BigDataStack infographic highlighting assets

User Journey

The user journey highlights the different roles needed to fully reap the benefits of the BigDataStack solution. The journey was highlighted during the BDV PPP 2020 and EBDV2020 webinars in presentations for which also a dedicated infographic was designed and adapted to each use case.

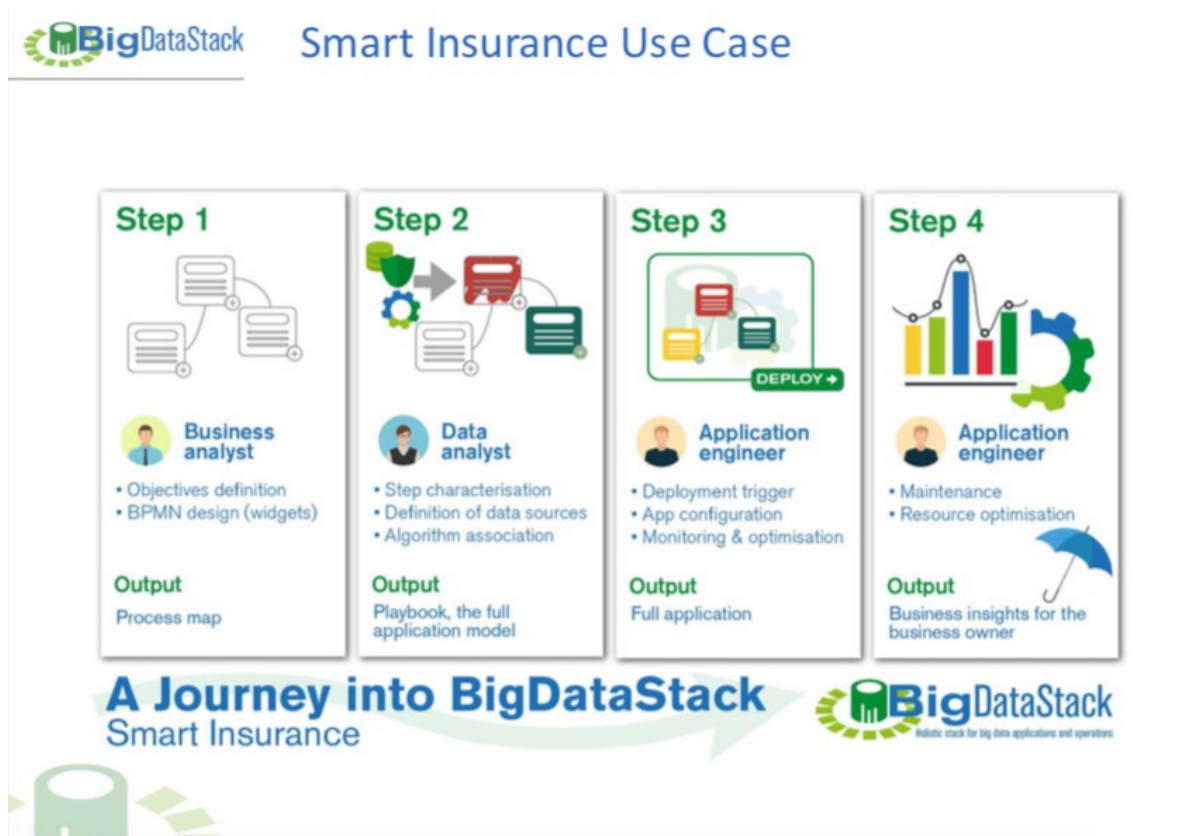


Figure 5: BigDataStack infographic on user journey

Social media campaign

A targeted social media campaign was designed and implemented to increase the visibility of the software components and their added value. All videos were strategically disseminated via Twitter and LinkedIn. The campaign was launched between August and September. Campaigns on both channels gained visibility:

- **Twitter:** with an increased number of tweets (+37.5%) BigDataStack twitter gained impressions (+5,4%), Profile visits (+156,5%), mentions (+30,8) and followers (+2) in the month of September.
- **LinkedIn:** in September a general increase has been registered on the page views (+4%), on the unique visitors (+14%), on reactions (+100%) and on shares (+225%)



Figure 6: Sample tweet in BigDataStack strategy highlighting project assets

2.3 Promoting BigDataStack Use Cases

BigDataStack Use Cases showcase the implementation and added value of the assets developed within the project and implemented in this public private partnership consortium. For communication aspects, use cases are the tangible story to tell of the implementation of BigDataStack and its assets. Leveraging on the possibility to promote industrial application of the software has been of the utmost importance in BigDataStack promotion. Over the course of the 36 months, the three use cases have been promoted through a number of tailored activities highlighted below, but also across the board in the horizontal activities explained in 2.3.

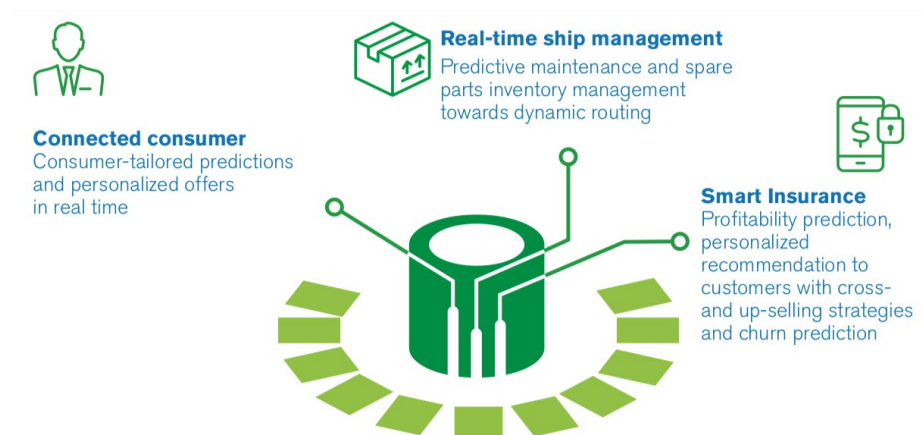


Figure 7: Infographic on BigDataStack use cases

Videos

With video being the most exploited media content all over the web (Hubspot - 2020), BigDataStack decided to make use of them as much as possible, producing a total of 34 videos of which [6 completely dedicated to the use cases](#) and the application of the software solution to real-life scenarios. The more than 8 minutes of footage have then

been synergically disseminated through the website, newsletter, social media and sponsored campaigns throughout the project lifetime.



Figure 8: Sample of dedicated and BigDataStack branded use case videos

Flyers

Towards summarising and contributing to disseminate the main exploitable assets for each Use Case, we've designed and printed a flyer for each one of them. The flyers have been distributed in all the physical events we've participated in and a digital version is available on the website.



Figure 9: BigDataStack branded flyers highlighting tailored to each of the use cases

Paid Campaigns on Social Media

Leveraging on the full potential of social networks, and with the aim to increase the visibility of the BigDataStack solution implemented in the three use cases to a targeted audience, a conservative amount of budget was earmarked through three paid campaigns (one per use case) were launched on LinkedIn. The money for this was taken from the savings made from the eventual travel cost savings not carried out due to the pandemic.

The Target audience was carefully built with people from the bottom of the funnel (active LinkedIn users from the IT industrial domain for retail, insurance, shipping) and showed them the short video interviews with relevant insights from the relevant use case. The campaigns performed well, reaching a total of **more than 33k people**, with 300 clicks of interested members of the audience reached.

Campaign Name ↕	Status ↕	Spent ↕	Key Results ↕	Cost Per Result ↕	Impressions ↕	Clicks ↕
3 selected campaigns		\$330.00	-	-	35,529	324
Engagement - Use cases - Ship ID: 168917303 - Sponsored Content	*** Completed ▼	\$110.00	124 Engagements	\$0.89	15,864	124
Engagement - Use cases - Retail ID: 168926553 - Sponsored Content	*** Completed ▼	\$110.00	116 Engagements	\$0.95	12,275	116
Engagement - Use cases - Insurance ID: 168931163 - Sponsored Content	*** Completed ▼	\$110.00	84 Engagements	\$1.31	7,390	84

Figure 10: Overview of achievements PPC campaign on the BigDataStack branded use case videos

Website and Marketing packages

Each use case has been highlighted with a dedicated webpage on the website. To give better evidence of the work that has been carried out, for each of them we've set up a marketing package with a collection of the informative materials from webinar resources, published news items to deliverables on ZENODO explaining and promoting the use cases.



Figure 11: Online marketing package set in place for each of the use-cases

2.4 Onboarding end-users

The BigDataStack consortium worked on onboarding end-users through its continuous dissemination and stakeholder engagement activities already detailed throughout this deliverable. In this section we highlight a few specific actions undertaken and achievements to onboard end-users such as the BDVA Innovation Marketplace, the European Open Science Cloud Digital Innovation Hub, the EC Innovation Radar, SWForum and showcasing actual adoption in EC projects and initiatives, Industry and Open Source Communities.

BDVA Innovation Marketplace

The Innovation Marketplace offers a view on assets resulting from projects under the BDV PPP umbrella. It provides free access to a catalogue of solutions from the BDV PPP to external organizations, as well a contact point to the organizations behind each asset, and the opportunity to participate in discussion forums, and a contribution channel to upload additional information for new assets.

The main functionality lies in the catalogue of solutions which provides a search and filter actions to help finding the right asset for the people looking for specific solutions for their needs in the scope of BDV PPP. The BigDataStack solution is onboarded by the BDV Innovation Marketplace, and leads directly to the BigDataStack catalogue of software components at the project website.

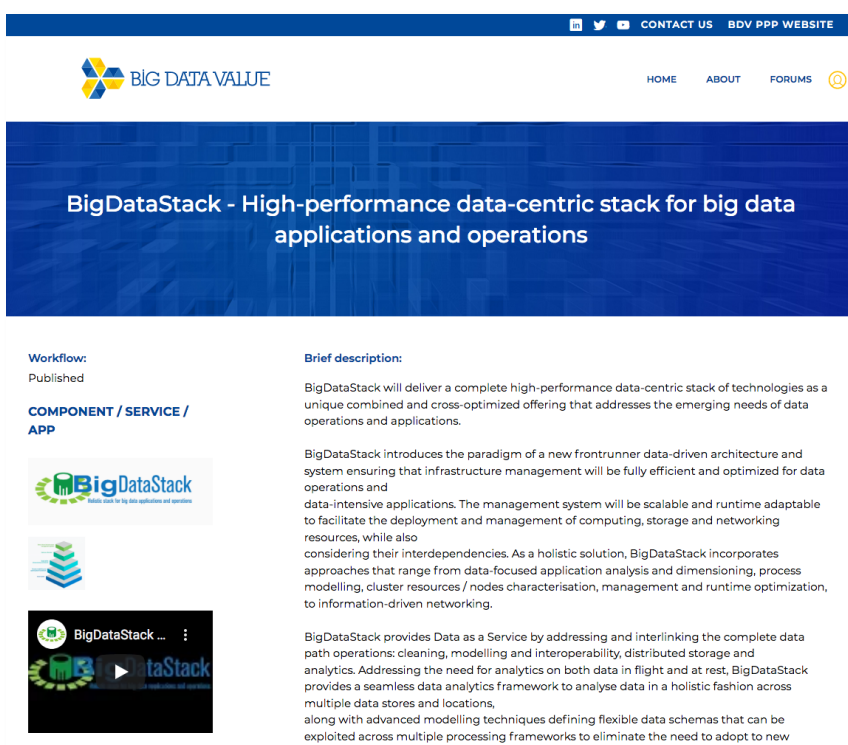


Figure 12: Sample of the BigDataStack on the Innovation Marketplace

EC Innovation Radar

BigDataStack technology Data Skipping has been identified in the EC Innovation Radar and its developer IBM Research as a key innovator and **has been published in its innovation radar.**¹³

¹³ <https://www.innoradar.eu/innovation/35322>

The **EC Innovation Radar platform**¹⁴ builds on the information and data gathered by independent experts involved in reviewing ongoing research and innovation projects funded by the European Commission. These experts also provided an independent view regarding the innovations in the projects and their market potential.

The aim is to make information about EU-funded innovations from high-quality projects visible and accessible to the public via the EU's Innovation Radar platform. This will show citizens and European organisations the many excellent technological and scientific advances being delivered by researchers and innovators around Europe, funded on their behalf by the European Commission. The innovation radar increases the visibility of the technology published, fostering its further adoption.

Showcasing adoption

BigDataStack has highlighted the adoption of BigDataStack technologies in new **EC projects and initiatives**, with the aim to inspire further adoption.

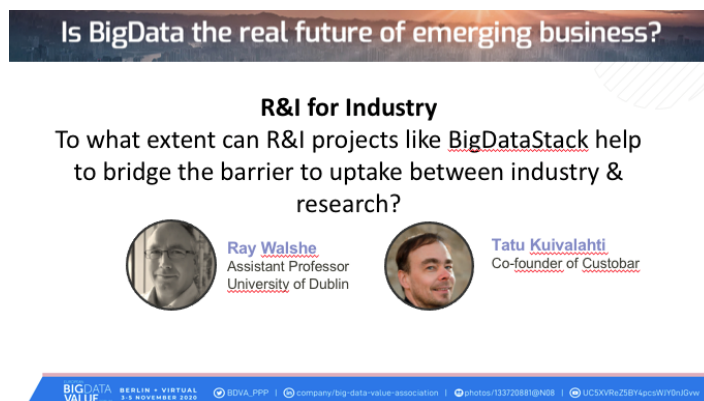
- As explained in section 2.1 BigDataStack joined efforts with the BDVA Task Force Fintech and InsuranceTech and the INFINITECH project in the BDV PPP 2020 hosted webinar titled “FinTech and InsuranceTech Case Studies Digitally Transforming Europe’s Future With BigData And AI-Driven Innovation.” During this webinar, the adoption of two of BigDataStack’s software has been presented and explained, fostering wider uptake. The event was attended
- In the same section we explain the joint series BigDataPilotDemoDays, where adoption of two of BigDataStack’s technologies are adopted in the full architecture and implemented in four pilot cases for public administrations and policy making organisations. This webinar inspired further adoption by telling the story of Policies against radicalisation¹⁵, pilot project at the Maggioli.
- As mentioned in section 2.1, I-BiDaaS Track&Know joined forces at the EBDVF 2020, titled “European Big Data Research for Industry. 3 Projects. 7 Sectors. 9 Applications. 41 Software Components. Now What?” During this session and in the upcoming report, both aimed at fostering further adoption of the technologies developed in all three projects, BigDataStack explained the adoption of one of its technologies in the INFINITECH flagship project.

BigDataStack partners have been engaging with end-users in industry, with SMEs and large enterprises to foster the uptake of the BigDataStack developed technologies. The following examples of adoption by industry have been disseminated to inspire further adoption.

¹⁴ <https://www.innoradar.eu>

¹⁵ <https://policycloud.eu/news-events/news/policies-against-radicalisation-webinar-insights>


- During the DATA+AI 2020 November conference and follow up blog, IBM presented the adoption story of “How the Weather Company Uses Apache Spark to Serve Weather Data Fast at Low Cost”¹⁶.
- During the final project event, BigDataStack invited Custobar, an SME, technology provider for retail businesses. During the discussion Research & Academia agreed on the importance of joining efforts. A next conversation with Custobar will look at possible future collaboration and adoption.




Is BigData the real future of emerging business?

R&I for Industry

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

 **Ray Walshe**
Assistant Professor
University of Dublin

 **Tatu Kuivalahti**
Co-founder of Custobar

BIG DATA VALUE | BERLIN + VIRTUAL 3-5 NOVEMBER 2020 | #DATA_PDP | @company/big-data-value-association | @twitter/13720810408 | @ics2019a2581#ccp@110a01ve

BigDataStack partners have been engaging with end-users to foster the uptake of the BigDataStack developed technologies in the Open Source Community on the one hand, as further explained in chapter 3.

Reaching the end of the project, more adopters have been onboarded. BigDataStack will work on telling those stories not yet told such as the adoption from ongoing H2020 projects Cybele, FogProtect and Accordion. In January 2021, the future H2020 projects Physics will start, adopting BigDataStack developed technology.

The European Open Science Cloud Digital Innovation Hub

The European Open Science Cloud (EOSC) is a supporting environment to accelerate the transition to more effective Open Science and Open Innovation in a Digital Single Market by removing the technical, legislative and human barriers to the re-use of research data and tools, and by supporting access to services, systems and the flow of data across disciplinary, social and geographical borders. EOSC Digital Invitation Hub (EOSC DIH) is an international and multi-partner cooperation that supports companies in easily accessing the digital technologies and services offered by the EOSC. It combines 4 main pillars to help companies become more competitive: Pilot design and co-design, Technical access, Training & support and Visibility. BigDataStack is working on signing a partnership with the EOSC DIH, fostering further adoption.

¹⁶ https://databricks.com/session_eu20/how-the-weather-company-uses-apache-spark-to-serve-weather-data-fast-at-low-cost

European forum of the software research community

Another future alignment will be with the activity being carried out by the CSA H2020 project SWForum.eu which is creating a European Observatory of Software excellence, planned to launch in May 2021, we would propose to channel the relevant BigDataStack components into this marketplace radar to showcase its results.

2.5 Promoting BigDatStack Open Source activities

BigDataStack actively engaged with open source communities for the upstream contribution of software components, but also contributing to the European Open Source Initiative through activities such as the RedHat Research Day, and further detailed in section 4 of this report.

The BigDataStack website contains a dedicated section on the European Open Source Initiative, targeted at raising awareness of its added value for Research & Academia, technology providers and big data providers, those not yet involved in Open Source Communities.

With the aim of raising awareness of the added value of Open Source with the wider community. BigDataStack invited Ray Walshe, Big Data expert, BDVA task force lead on standardisation and project partner to the StandICT.eu project. During the expert discussion at the BigDataStack final event, two of the five questions asked and prepared with the panel, aimed to highlighting the importance of Open Source for the digital future of Europe¹⁷.

Moreover, the consortium members engaged with representatives from the Open Forum Europe to highlight the main priorities of the project and to seek mutual collaborations. This is being brought forward through a dialogue with the Open Forum Europe CEO to understand further measures of collaboration in the coming months.



Figure 13: Panel questions to highlight the added value of Open Source for the digital future of Europe.

¹⁷ <https://youtu.be/xj3iNc41Szo>

2.6 Horizontal activities

2.6.1 Website

BigDataStack’s website [bigdatastack.eu] harvests all the achievements of the project and is used to showcase results and activities. The website is the central hub to interact with our target audiences via all project and partner channels and communication tools. The website also provides the option to sign up for the newsletter and to directly contact the project.



Figure 14: Sample of the BigDataStack project website

Website as knowledge hub

The website reflects the main pillar of BigDataStack work and provides context by linking to the relevant reports, publications and communication resources developed, providing a knowledge hub of all relevant information and placing it in the proper context.

A **dedicated section on the project use-cases** describes the added value of BigDataStack in smart insurance, connected consumer and real-time shipping. The section provides access to a tailored online marketing package for each of the use-cases, as described in section 2.2.

The website provides insights into the BigDataStack solutions in the format of a **software component catalogue** as described in section 2.1, with short descriptions, information on licenses, downloadable factsheets and videos.

A specific section is dedicated to the project results, providing access to BigDataStack publications, deliverables and their digest and communication kit and all 34 videos produced and promoted.

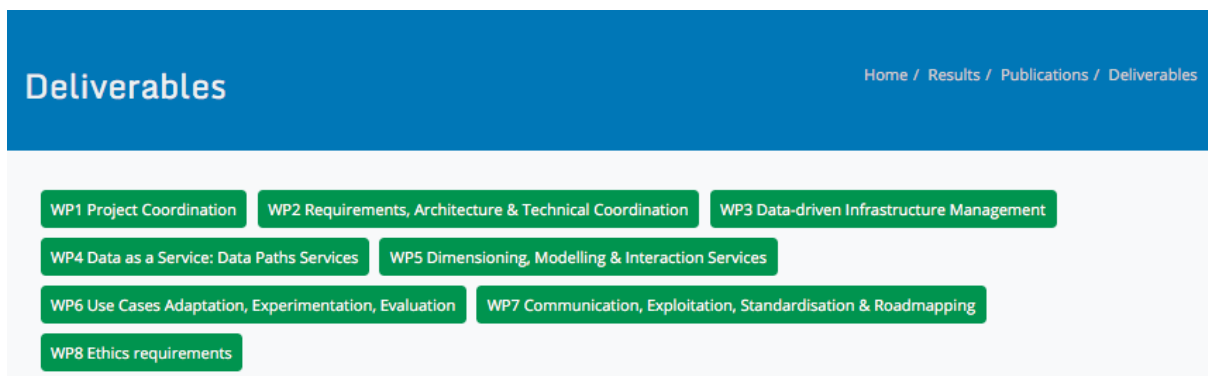


Figure 15: Sample of the deliverables dedicated webpage, highlighting the filter function per WorkPackage

Fostering synergies

A dedicated section on the project's liaisons lists other European projects and initiatives that the consortium has engaged with in terms of communication and stakeholder engagement. A sample is shown in the figure below.

Moreover, the BigDataStack approach to a **European Open Source Initiative** has a dedicated section, linking to relevant information on how to successfully merge open source contributions upstream in the broader open source communities. The information is targeted at industrial or academic institutions, EU-funded projects and individuals who are going through the process of development for new open-source artefacts or projects, open standards and open-source initiatives.

Liaisons

[Home](#) / [Liaisons](#)

Lemo project

Transport researchers and policymakers today face several challenges as they work to build efficient, safe, and sustainable transportation systems. From rising congestion to growing demand for public transit, the travel behaviour and transportation preferences of city dwellers are changing fast.



Big Policy Canvas

Big Policy Canvas will foster collaboration among public sector, enterprises, citizens and researchers and will offer the appropriate knowledge base, research directions and recommendations towards building a more effective, efficient, precise and evidence-based public sector, promoting transparency and restoring trust to public sector structures.



e-sides

e-SIDES is an EU-funded Coordination and Support Action (CSA) that will complement the Research and Innovation Actions (RIAs) of the ICT-18 call on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies and providing a broad basis and wider context to validate privacy-preserving technologies.

Figure 16: Sample of the dedicate webpage on BigDataStack liaisons

Monitoring

The website is continuously updated, with new content on the news and events section as well as in its structure, carefully mirroring the achievements of the project.

The Google Analytics platform¹⁸ is used to track all the traffic on the website and users' behaviours. Google Data Studio is used to visualise and show the user experience statistics with a more user friendly and an easy reading approach.

Over the course of the last year of the project, BigDataStack collected more than 5k session lasting for 3 minutes on average, more than 22k pageviews. 4k users from almost all over the world visited the website, focusing especially on the software component catalogue (1.2k visits).

¹⁸ https://datastudio.google.com/reporting/1rCKeCLSM_BPYS3ytPZuqMz6WPD11Pqlu/page/1M

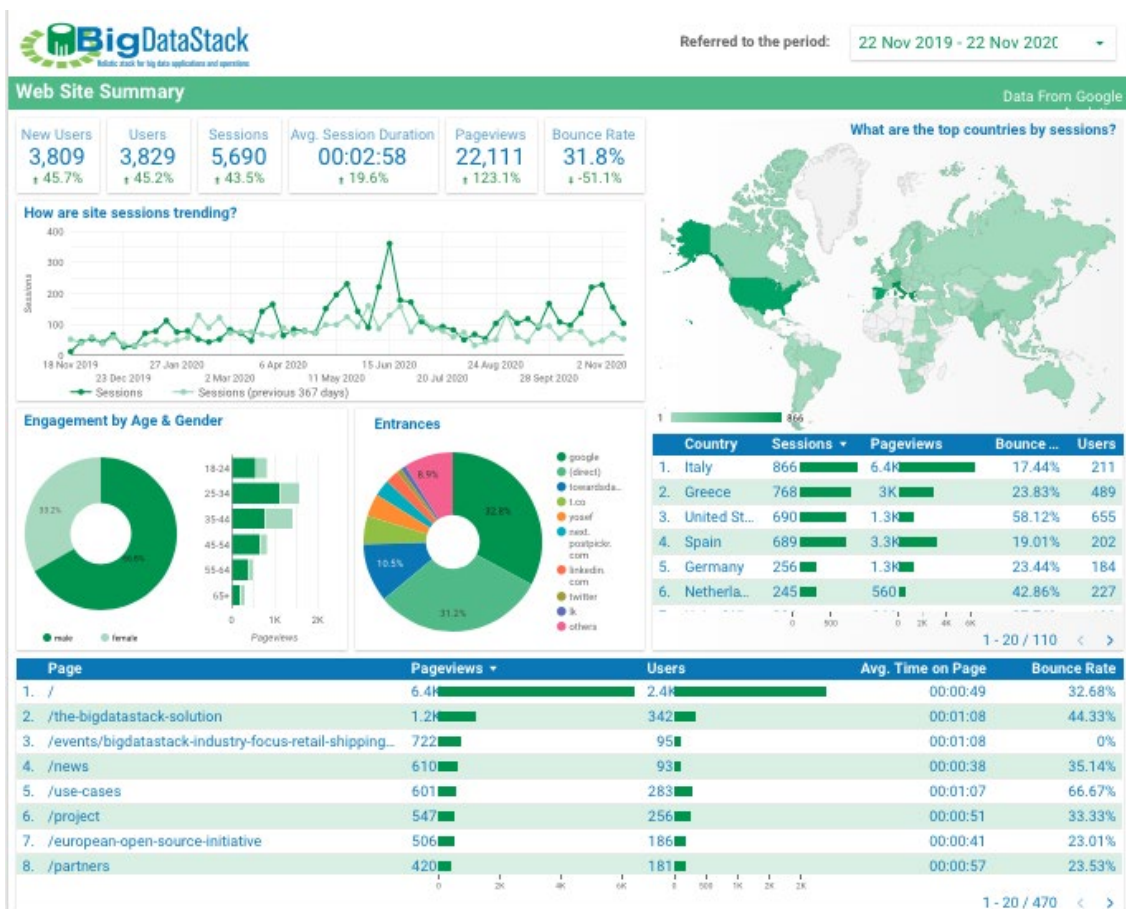


Figure 17: Dashboard monitoring website traffic.

2.6.2 Videos

A significant number of videos have been produced over the course of the project, to better reach our audience and make the content of the project as much appealing, understandable and enjoyable as possible. Over the course of the project, 34 public videos have been published and they collected more than **37.000 views** from all over the internet. The project produced promotional videos, demos, software component explainer videos and training videos. All videos are accessible from dedicated playlists on YouTube and the Video resources page. BigDataStack partners collaborated to produce videos on project insights, training, technology demo and BigDataStack software component explainer videos.

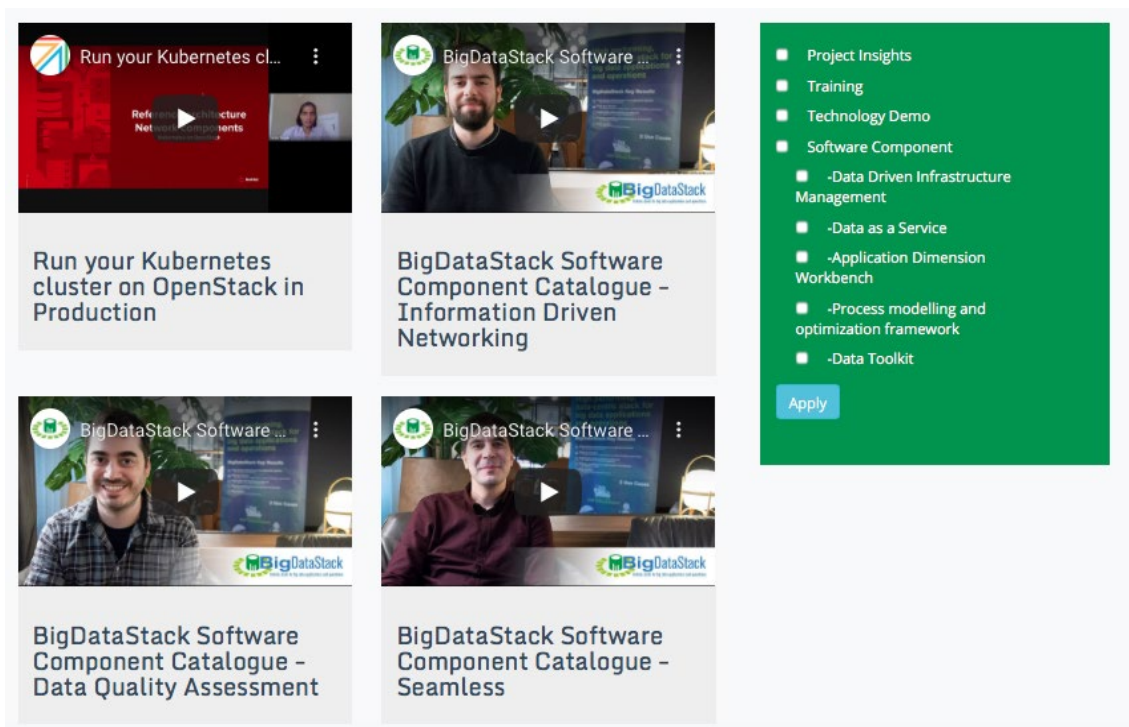


Figure 18: Sample of dedicated video webpage, with filtering options according to the category of the video.

2.6.3 Publications & ZENODO

All the results in terms of Scientific publication, events and conferences attended have been widely published on the main dissemination channels. BigDataStack has made **13 scientific publications, 1 poster presentation and 7 blog posts designed to share knowledge** gained within the BigDataStack project in external communities. The majority of the scientific papers have been uploaded on **open access repositories** (like Arxiv or RecSys). A complete list of these publications can be found in Annex 1.

In addition, in the third year of the project, building on experiences from other H2020 projects, the consortium decided to ensure the long-time preservation and access to its results through the publication of deliverables and reports, year 3 presentations to ZENODO. **BigDataStack now has a ZENODO community with 20 deliverables uploaded, viewed 161 times and downloaded 123 times** at the time of writing of this deliverable.

The OpenAIRE explore tools gives an overview of the BigDataStack community of ZENODO.

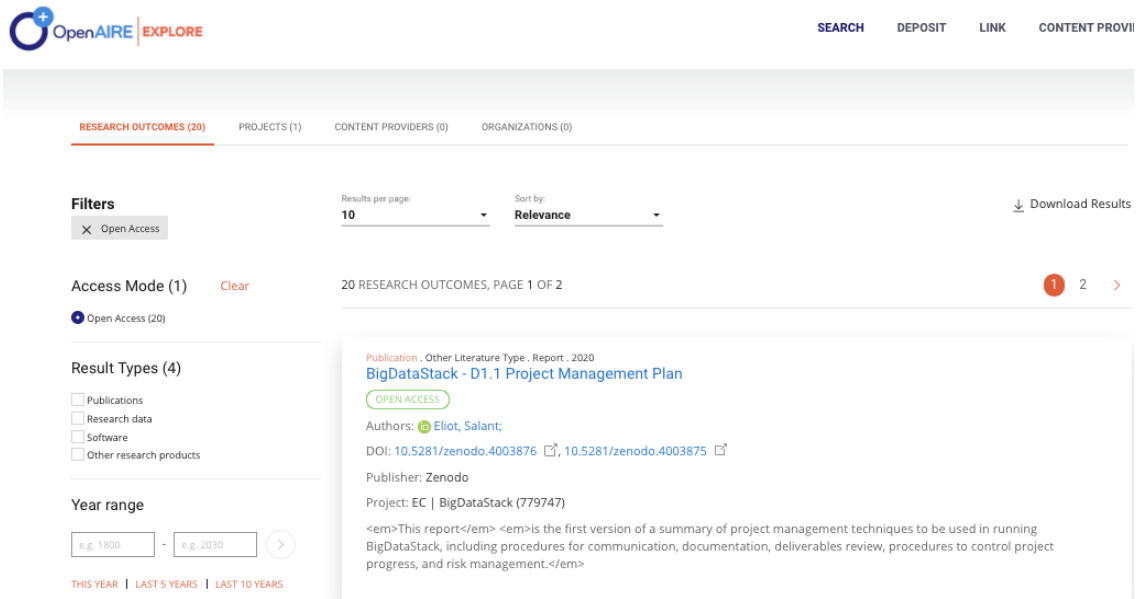


Figure 19: OpenAIRE explore BigDataStack publication overview

At the same time as the set-up of the BigDataStack community on ZENODO, the website dedicated page on deliverables was enhanced with a filter option according to project work packages, followed by a tailored social media campaign on the deliverables and their key results. This resulted in an increased visibility of the deliverables webpage as illustrated by the visualisation graphic below.

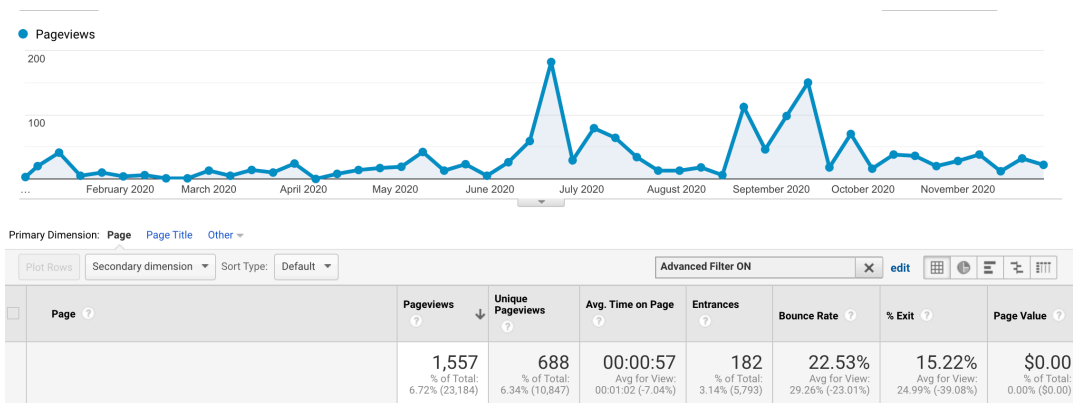


Figure 20: Visualisation of the Deliverable section of the website in 2020

2.6.4 Social Media Channels

BigDataStack is present on LinkedIn, Twitter and YouTube to increase interaction with target audiences. A curated Social Media editorial planning aims at information rich, branded and tailored messages to engage stakeholders and grow the BigDataStack community. The common approach, which has been used for all the social media channels, consist in an informal, easy and warm tone of voice. Even when speaking about complex and technical concepts, with this approach the project social media

community has been engaged in an approachable and interest triggering way. A top performing tweet and LinkedIn post can be seen in the figure below.

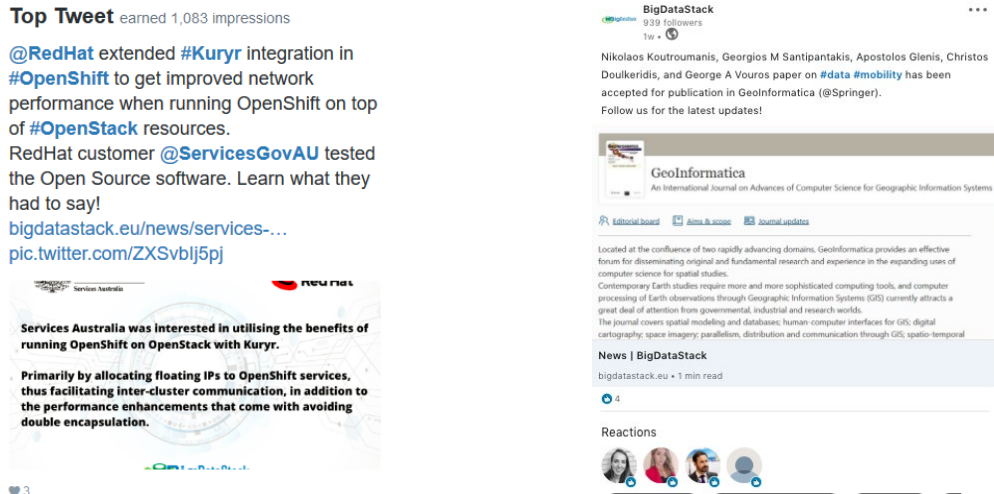


Figure 21: sample of 2 tweets illustrating the social media strategy with tailored, branded and content rich messages

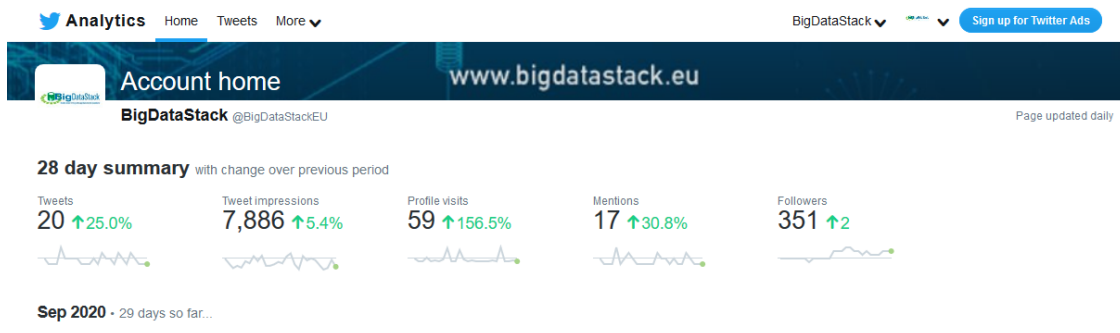


Figure 22: Overview of social media visibility of September 2020

Organic promotion and social media

The massive use of social media helped the project to reach out a wide audience. At the time of writing a total number of 1000+ posts on LinkedIn and Twitter built up a community of more than **+2200 social media connections**, consisting of 1600 connections on LinkedIn, 800+ followers on the LinkedIn company page and 350+ followers on Twitter. The figure below shows the profiles of the BigDataStack LinkedIn community, where technical profiles are well represented, as well as those of business development, sales and marketing ultimately looking to improve customers' experiences.

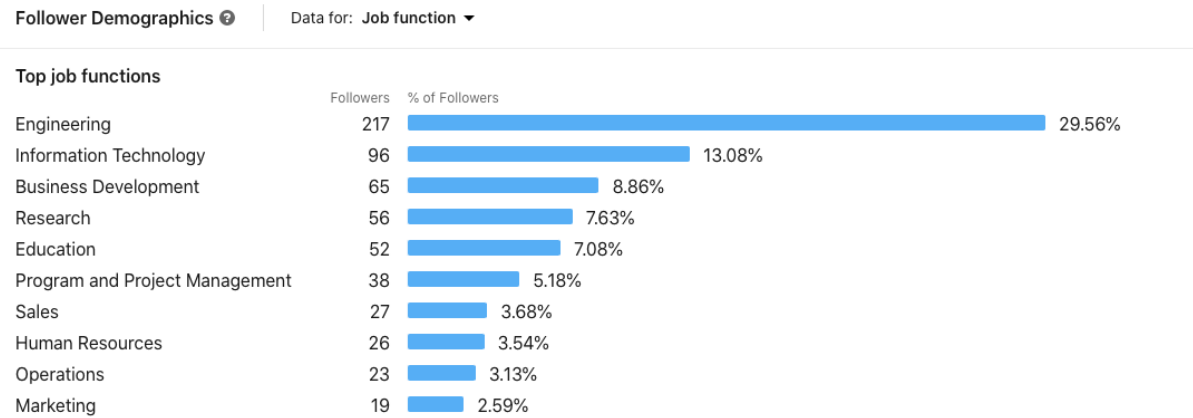


Figure 23: Overview business roles BigDataStack LinkedIn community

Leveraging on multipliers in the field

BigDataStack social media channels connected to target audiences representing the wide variety of BigDataStack stakeholders through following their channels on Twitter and connecting on a one-to-one basis on LinkedIn. BigDataStack leveraged on the visibility of key-players in the field. A sample of such an activity on twitter and LinkedIn is given in the figures below.

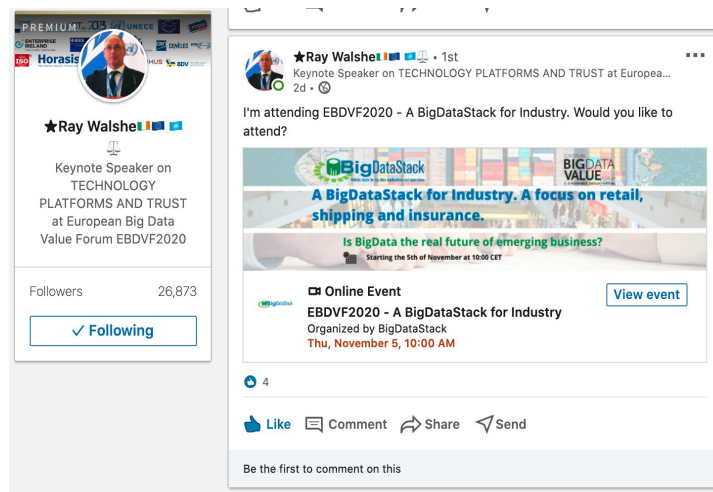


Figure 24: Example of a BigDataStack multiplier, in this case on Big Data Standardisation and Open Source

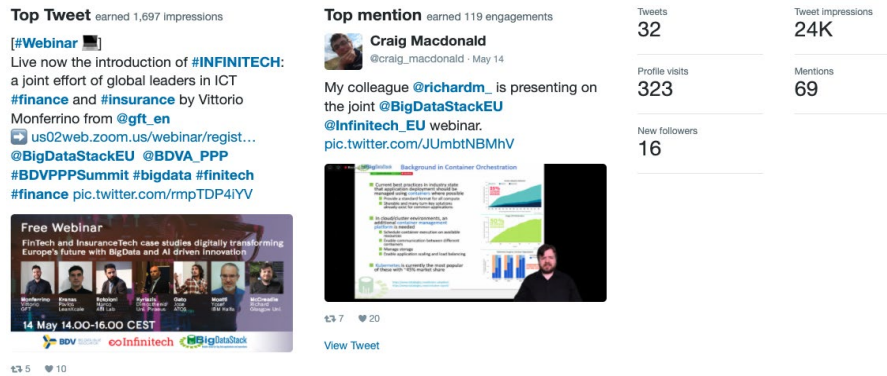


Figure 25: Examples of twitter posts connecting to multipliers in the filed

2.6.5 Events, webinars and synergies

The BigDataStack consortium engaged stakeholders through 54 of third-party events, 3 project webinars, 2 joint webinars, one joint series of 9 online demonstrations and 13 strategic synergies. The purpose behind these activities was to promote the BigDataStack results, share knowledge and foster uptake of the technologies developed with end-users. All events attended can be seen in Annex II.

BigDataStack final event



Figure 26: Promotional image BigDataStack final event

After 3 years of research and innovation the BigDataStack consortium is proudly showcased its infrastructure management system for the holistic analysis of big data. Handling analytics on both data in flight and at rest, across multiple data stores and from several sources. At the final event: “**A BigDataStack for Industry - a focus on retail, shipping and insurance: Is Big Data the real future of emerging business?**” at the

2020 edition of the European Big Data Value Forum (EBDV 2020) titled “Towards a European ecosystem for Big data and Artificial Intelligence. The 3-hour event was structured in 2 parts:

1. **A BigDataStack for Industry - a focus on retail, shipping and insurance:** illustrating the added value of a BigDataStack for industry through a live demo of the BigDataStack User Interface and an end-to-end presentation of the three use cases.
2. **A BigDataStack for Industry. Is Big Data the real future of emerging business?** Big Data Gurus from retail, shipping and insurance and BigDataStack experts Céline Xu, Ray Walshe (Dublin City University) and Tatu Kuvalahti (Custobar) engaged the audience in a panel discussion on the future of Big Data in emerging business. Yosef Moatti (IBM Research) will take the audience and panellists through Big Data for emerging business addressing Data Exhaust, Open Source as a barrier or blessing for adoption and European sovereignty.

The event was attended by 143 participants from a wide scope of stakeholders, ranging from research and Academia to Industry. All resources from the event are accessible via the project website. More on the inputs provided for future adoption in chapter 4.

Synergies

BigDataStack has created synergies to foster knowledge sharing with other European initiatives and projects on the one hand and to join efforts in promoting project results as a group. We list them in the table below.

<p>Synergies established at local, national, or international level for uptake of BigDataStack results</p>	<p>☐KPI achieved 7 synergies established at events:</p> <ol style="list-style-type: none"> 1. ICT 2018: panel discussion - synergy with BDVe 2. RedHat summit Researcher Day 2019 3. EBDVF 2018 - Data4AI everywhere panel discussion and report 4. BDV PPP 2019: panel discussion with BDV PPP projects and co-written policy brief 5. BDV PPP 2020 joint series of 9 demos 6. EBDVF2020 joint session with ICT14 & ICT16 projects, 7. Chapter contributed to the BDVA Open Access Booklet
<p>Synergies at national or international levels for sharing knowledge and standardisation.</p>	<p>☐KPI achieved with 6 synergies:</p> <ol style="list-style-type: none"> 1. KubeCon + Cloud Natives (working group on policies for open source) 2. BDV PPP 2019 coordination and technical meetings, 3. BDV PPP 2020 coordination and technical meetings, 4. Open Source Summit - RedHat Kuryr BigDataStack element presented 5. Joint webinar with INFINITECH & BDVA Task Force on FinTech and InsuranceTech, 6. EBDVF2020: Final event joint discussion with BDV Task Force lead Standardisation and SME

In this section we'd like to highlight the fruitful synergies created with other EC initiatives.

The #BigDataPilotDemoDays



Figure 27: BigDataPilotDemoDays joint branding

During the virtual BDV PPP Summit 2020 BigDataStack, I-BiDaaS, Track & Know and Policy Cloud joined forces in a series of 9 online demonstrations. The #BigDataPilotDemoDays demonstrated innovative Big Data Technologies unlocking the potential of applications in domains spanning from telecommunications, transport, finance, retail, manufacturing 4.0 and health to citizen mobility and policy-making against radicalisation. Policy Cloud highlighted two of its software components to be implemented in the coming two years of the project, adopted from BigDataStack. The projects collaborated on a joint branding, social media campaign, liaising with the BDV PPP Summit organisation and technical support for each of the 9 demo webinars. The series was well attended by over 420 people from all over the world, representing a wide scope of stakeholder from Research and Academia, Big Data technology providers and other Industries. All slides and recordings have been made available for reuse on the four project websites and channels. Comments on this joint achievement can be seen in the final joint communication¹⁹.

¹⁹ <https://bigdatastack.eu/news/reaping-benefits-bigdata-developments-big-data-pilot-demo-days>

European Big Data Research for Industry. 3 projects. 7 sectors. 9 applications. 41 software components. Now what?



Figure 28: Promotional image joint event BigDataStack, I-BiDaaS and Track & Know

After 3 years of research and innovation, I-BiDaaS, Track & Know and BigDataStack joined their forces again in the led expert-led discussion on the impact and uptake of Big Data research. The purpose of the discussion was to (i) Identify shared barriers to adoption of Big Data research in different sectors, and mechanisms to overcome them, (ii) map the current and future impact and sustainability of their Big Data research, (iii) share best practices on the concrete business questions that have been answered in the project pilots. The session was attended by 117 people from 14 different countries a wide scope of stakeholders, from Research and Academia (62%), to Big Data Technology providers (23%). All resources are made available via the EBDVF2020 conference platform and project channels.²⁰ A joint report is underway.

²⁰ <https://bigdatastack.eu/events/ebdvf-2020-european-big-data-research-industry-3-projects-7-sectors-9-applications-41>

FinTech and InsuranceTech case studies digitally transforming Europe's future with BigData and AI-driven innovation



Figure 29: Jointly branded promotional image for the INFINITECH, BigDataStack and BDVA Task force event

INFINITECH, BigDataStack and the BDVA Task Force on BigData and AI for Finance joined forces in a two-hour webinar on data-driven innovation contributing to Europe's digital future in Insurance and Finance. The webinar illustrated the need for digital transformation by defining the requirements of INFINITECH use cases in the financial and insurance service sector. The solutions provided by BigData, Internet of Things and Artificial Intelligence-driven technologies, developed in the BigDataStack and INFINITECH projects were showcased. This two-hour session was attended by 147 participants from Research and Academia, Big Data providers, Big Data Technology providers, Insurance & Finance and policy makers. All resources are made available via the BDVA and both project's websites.²¹

Policy4Data Policy Brief

Big data in Europe for 2020 and beyond: Policy insights and recommendations from current H2020 big data projects. This policy brief reflects current developments within the several Big Data research projects funded under H2020 and, combined with insights from the BDV PPP summit 2019 in Riga, aims to contribute to ongoing challenges in Europe around the regulation of big data. This policy brief is a product of the Common Dissemination Booster, funded under H2020. The policy recommendations are based on projects participating in the CDB services. BigDataStack was involved in and contributed

²¹ <https://bigdatastack.eu/news/fintech-and-insurancetech-case-studies-digitally-transforming-europe's-future-bigdata-and-ai>

to the discussion panel and the policy brief, available via the BDVA and all contributing projects.²²

Webinars

Already in 2019, BigDataStack started engaging its stakeholders through webinars, as a warm-up for the 3rd year in which the consortium had planned to deliver a series of three face-to-face workshops. As mentioned in section 2.1, the COVID-19 global health crisis forced the consortium to come up with creative alternatives for these workshops. BigDataStack had already built strong connections with the BDVA through its synergies created in year 1 and 2 of the project and leveraged on this when delivering a series of joint online demonstrations, an online booth, joint webinars and one final project event in collaboration with BDV PPP projects and BDVA Task Forces at the BDV PPP 2020 and EBDVF2020. The figure and table below show the stakeholder groups represented by the 953 attendees engaged from 43 different countries in the total of 14 webinars delivered.

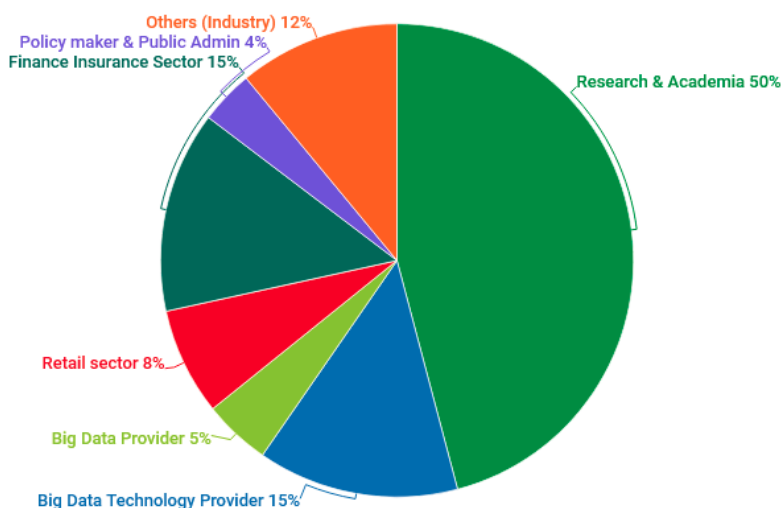


Figure 30: Overview of sectors represented by stakeholder engaged in the BigDataStack webinars

Webinar title	N Attendees	Date
A BigDataStack for Industry - a focus on retail, shipping and insurance: is BigData the real future of emerging business?	143	05/11/2020
Parallel session on European Big Data Research for Industry. 3 projects. 7 sectors. 9 applications. 41 software components. Now what?	196	03/11/2020
Big Data Pilot Demo Days - Track & Know Healthcare: Using patient mobility information to understand the provision of services across large rural and urban communities	85	16/07/2020

²² <https://bigdatastack.eu/news/policy4data-policy-brief>

Big Data Pilot Demo Days - Track & Know Insurance Sector: Using mobility data to understand and mitigate risky driving behaviour	68	14/07/2020
Big Data Pilot Demo Days - I-BiDaaS application to the manufacturing sector - CRF	55	09/07/2020
Big Data Pilot Demo Days - Track & Know Fleet management services: Quality and predictions in location data from GPS devices	102	07/07/2020
Big Data Pilot Demo Days - Policy Cloud Pilot for Policy Cloud For Policies against radicalisation	42	02/07/2020
Big Data Pilot Demo Days: I-BiDaaS Application to the Telecommunication Sector	43	21/05/2020
Big Data Pilot Demo Days - A BigDataStack Seafarer's tale of Real-Time Shipping	31	18/05/2020
Big Data Pilot Demo Days - BigDataStack Connected Consumer	39	28/05/2020
Big Data Pilot Demo Days - I-BiDaaS Application to the Financial Sector	48	21/05/2020
FinTech and InsuranceTech case studies	147	14/05/2020
BigDataStack Connected Consumer Technologies for Retailers	26	29/05/2019
BigDataStack Technologies for Shipping	28	03/06/2019
Total Numbers	953	

Table 2: Overview of BigDataStack (co-)organised webinars with date and number of participants

2.6.6 Newsletter

BigDataStack results and activities were disseminated via a newsletter, sent to a list of people that signed up to receive regular BigDataStack updates. The quality of this audience is among the highest for the project, as these are people who signed in to our newsletter on their own. The contacts have been gathered through the website, during live events and over the course of specific webinars, where an “upselling” strategy has been applied to people that already manifested their interest in the project.

The total number of people receiving this newsletter is 241, the opening rate for the emails is, in fact, 37% on average, and in total 18 newsletters were sent. The mailing list was also used for 3 specific targeted messages on events and press releases.

3 Standards & Open Source Contributions: European Open Source Initiative (M1-M36)

3.1 Realising the European Open Source Initiative to increase the impact of the BigDataStack outcomes

3.1.1 Research Days

Red Hat is one of the leading companies as regards to Open Source activities, being worldwide recognised for its contributions to Open Source Communities as well as open culture. In order to increase the impact that BigDataStack contributions can make, as well as to foster Open Source contributions from research projects, Red Hat started to organise the Red Hat Research Day as part of its research activities²³. The Research Day is a one-day event dedicated to Red Hat's research initiatives that aims to be a showcase for the research Red Hat supports at various universities and research institutions worldwide.

The main goal of the event is to bring together international researchers, Red Hat engineers, industry representatives, customers, partners, and other interested people to share knowledge about the latest research findings, network with research enthusiasts from around the world and learn about moving great research ideas into open source communities.

The first research day was in 2019 as a part of Red Hat Summit in Boston, US. Red Hat organised an extra "Research Day" at its annual summit²⁴ that is usually gathering between 5 and 8 thousand people from many companies around the world. The first Red Hat Research Day addressed "Bridging Between Research and Upstream Engineering²⁵", where the morning talks were focused on different data processing topics, such as machine learning, Multi-Party Computing and Homomorphic Encryption. Including a talk given by Dimosthenis Kyriazis about the Big Data Stack projects "**An Architecture Slack for Data-Driven Infrastructure Management**".

The second half of the day was dedicated to the infrastructure side and exploration of different tools and techniques to improve the way data can be processed, such as adding FPGAs, UniKernels, OS accelerators, etc.

As part of being present at the Red Hat summit, the Big Data Stack project also had a presentation slot at one RedHat booth in the exposition hall. This also led to discussions with related projects that were working on similar tools, such as OpenDataHub (A Data & AI Platform for the Hybrid Cloud²⁶).

Following a successful inaugural event in Boston, Research Day expanded to Europe in 2020. **The second Research Day**²⁷ was collocated with the Brno DevConf event. It brought together an audience of about 130 invited participants from academia, Red Hat customers, Red Hat partners, and also companies interested in research and

²³ <https://research.redhat.com/>

²⁴ Red Hat summit 2019 -- <https://www.redhat.com/en/summit/2019>

²⁵ <https://research.redhat.com/research-day/research-day-us-2019/>

²⁶ <https://opendatahub.io>

²⁷ <https://research.redhat.com/research-day/research-day-europe-2020/>

cooperation with universities to exchange ideas, share their experiences, achievements and common passion for open source research. The main topics were focused on current research projects such as encryption, crypto-algorithms, new forms of attacks, and work on automation of formal code verification.

Finally, a **third edition** has been organised in a slightly different format due to the COVID impact²⁸. This time it spawns among 4 different days in September, fully virtual, and with the focus on a mix of discussions and presentations with focus on how to improve privacy and security, make experimentation and system execution more reproducible, and enhance the performance of cloud systems.

3.1.2 Upstream Contributions

As part of the European Open Source Initiative, we are not only contributing the code developed for most of our components into public repositories, but also contributing some parts to well established Open Source Communities, such as OpenStack, OpenShift, Kubernetes, or Spark. The latter does not mean to only make the code changes public, but to engage with the related communities, make your code go through the community review and testing processes, as well as to engage with the community on discussions about the best way of doing a specific feature and how to make it reusable in more use cases. Examples of this are the contributions made to the OpenStack Kuryr project, or to several OpenShift projects such as Cluster Network Operator or Cluster API Provider OpenStack (related to the Kuryr integration into them) or a contribution to Spark 3.0 to allow us having the Data Skipping to work with it.

Moreover, as part of European Open Source Initiative, several proposals for talks were submitted to be presented at their respective community summits, such as KubeCon or Open Infrastructure Summit (previously known as OpenStack Summit), conferences that gather thousands of people from Open Source communities. The focus of the talks we submitted can be grouped in several topics:

- **Related to the Kuryr project:** We had several talks at the Open Infrastructure Summits around Kuryr²⁹ functionality and its new features.
- **Focusing on how to start contributing to Kubernetes upstream.** This talk was submitted to San Diego KubeCon but unfortunately got not accepted.
- **Focusing on how to adapt your applications to the Kubernetes operators/CRDs model^{30,31}.** A talk about this has been accepted at Open Infrastructure Summit and covers the path we followed in Kuryr to adapt to this model (as part of the Big Data Stack contributions).
- **Focusing on running OpenShift/Kubernetes on top of OpenStack and their benefits and new features³².** Accepted at Open Infrastructure Summit. It covers

²⁸ <https://research.redhat.com/research-days-us-2020/>

²⁹ <https://www.openstack.org/summit/berlin-2018/summit-schedule/global-search?t=kuryr>

³⁰ <https://www.openstack.org/summit/2020/summit-schedule/events/24615/k8s-controller-and-crd-patterns-in-python-kuryr-kubernetes-case-study>

³¹ <https://www.openstack.org/summit/denver-2019>

³² <https://www.openstack.org/summit/2020/summit-schedule/events/24576/run-your-kubernetes-cluster-on-openstack-in-production>

the work done in Kuryr for Big Data Stack for network improvements such as distributed E/W load balancing, Network Policies support, Manila support.

In addition, and to further engage with the community around the Kuryr project, we have organised a **Project Team Gathering session for Kuryr³³ as part of the Open Infra Summit 2020³⁴**. The Project Teams Gathering is an event organised by the OpenStack Foundation to allow the various technical community groups working on OSF projects to meet, exchange and get work done in a productive setting.

3.1.3 Blog Posting

In Open Source communities it is common to publish blog posts around specific settings or configuration, new features or performance evaluation. Those get spread quite fast and may help people to know or test certain features as well as to know how to configure them. We wrote a few blog posts related to the **Kuryr new feature developed for the BigDataStack** needs:

- Performance evaluation of Kuryr networking improvements as well as the advantages of distributed E/W traffic³⁵.
- Kuryr configurations/tuning to adapt to specific customer cases. This is the output of working with real customer needs and making new Kuryr features part of Red Hat portfolio³⁶.
- How to make EU projects contributions into Open Source projects

3.1.4 OutReachy Internships

Outreachy is a paid, remote internship program whose main goal is to support people from groups underrepresented in tech. And the main outcome is to help newcomers to Free and Open Source Software make their first contributions³⁷.

As part of Big Data Stack, and related to Red Hat efforts on Kuryr OpenStack project, we have also requested and received funding for helping people getting familiarised with both the Kuryr project and how to contribute to Open Source communities. The topic of the outreachy position is "Add new IP protocol to Kuryr-Kubernetes"³⁸.

³³ <https://bigdatastack.eu/events/project-teams-gathering-coming>

³⁴ <https://www.openstack.org/ptg/>

³⁵ https://www.openshift.com/blog/accelerate-your-openshift-network-performance-on-openstack-with-kuryr?extIdCarryOver=true&sc_cid=701f2000001Css5AAC

³⁶ <https://developers.redhat.com/blog/2020/10/02/customizing-and-tuning-the-kuryr-sdn-for-red-hat-openshift-3-11-on-red-hat-openstack-13/>

³⁷ <https://www.outreachy.org/apply/>

³⁸ <https://www.outreachy.org/apply/project-selection/>

3.2 *BigDataStack Alignment global standards and standardisation of project results*

One of the standards for software development and products is the “open source way”. Open Source is nowadays attracting tremendous technical talent and companies behind. Much of the innovation is taking place in open source communities, made of people all over the world. Successful examples are Linux, OpenStack, Kubernetes, Git. Red Hat is the largest open source company in the world, building open source products from open source upstream projects. In order to “comply” with the open source standards and being able to have a product out of an open source repository it is not enough to make your code available in GitHub or similar repositories, but to:

- Create a community behind the projects, i.e., engaging more than just one company behind a given project
- Create an open forum for ideas
- Enforce open culture and collaboration, i.e., code reviews, testing, meritocracy (a.k.a., best ideas win)

The BigDataStack project is relying on many successful open source projects and products, especially the ones targeting cloud computing. It is using the ‘de-facto’ standard tools for the infrastructure to ensure easy adoption of the developed tools, namely **OpenStack for the Cloud Infrastructure** and **OpenShift/Kubernetes for the Container Orchestration Engine**. To better align our tools with the software standards in open source communities we were present at the KubeCon & CloudNativeCon event³⁹. Being led by the Red Hat knowledge about the current trends on application development for Cloud Native applications, the project has made a few decisions to pursue increasing portability on different clouds. Thanks to the efforts listed below, it will be easier to run Big Data Stack components anywhere, be it an on-premise cloud (e.g., OpenStack) or a public one (e.g., Azure, GCE, AWS):

- Containerise the different Big Data Stack components
- Components deployment are defined on Kubernetes/OpenShift templates that are the ‘de-facto’ standards for containers orchestration engines
- Use of OpenShift provides extra tools for DevOps that facilitates even further the deployment and management of the applications life cycle (from code to running containers)
- Creation of operators that will handle different components (such as Kuryr Cluster Network Operator) that make it easier to manage both applications and infrastructure in a fully automated and software defined manner.

The Cloud Native Computing Foundation target is to build sustainable ecosystems and foster a community around a big constellation of high-quality projects that orchestrate containers as part of the microservices architecture⁴⁰. The main idea behind this is to ensure, through the open source way, a universal and sustainable cloud native computing. Red Hat is one of the Platinum Members of such an organisation. To

³⁹ <https://events.linuxfoundation.org/events/kubecon-cloudnativecon-europe-2019/>

⁴⁰ <https://www.cncf.io/>

make BigDataStack components more aligned with the CNCF, which in turn will help in possible adoption of their outcomes, we are relying on many of the projects already being part of the organisation. And taking stock of their good practices, such as containerisation of the applications as well as working on operators for full applications lifecycle management. Note the CNCF projects are divided in 3 different maturity levels, depending on their status and adoption:

- **Graduated:** the most mature projects that have widely adopted and high maturity. Among them, we have Kubernetes, Prometheus and CoreDNS, already used in the BigDataStack project and specifically in the infrastructure layer. For instance, changes being made to the Triple Monitoring tool, if adopted by the upstream community, may have a big impact.
- **Incubation:** more recent projects that are already mature and have increasing adoption. Examples of that are CNI, etcd, CRI-O rook. As part of the projects we are working on extending Kuryr, which is a CNI plugin for Kubernetes, helping it to increase its adoption, and already raising interest by big companies that need single SDN control plane for VMs and Containers, as well as requiring extra networking performance.
- **Sandbox projects:** This is the entry point for early stage projects with the main goal of facilitating alignment with the existing projects as well as foster community collaboration/uptake. Example of this are Network Service Mesh or Open Telemetry. This could be one of the targets for specific BigDataStack components in order to ensure full alignment with open standards as well as adoption/maintainability after the project.

We have been watching the different projects in the 3 categories to identify possible projects where we can contribute to (such as the CNI one) or that we can benefit from (such as Prometheus, Thanos or Network Service Mesh ones). We are also investigating what is needed to include some of our components as Sandbox projects at CNCF, which will hugely increase the impact of the project as well as to help standardization around it.

In addition to that, **we have closely followed the patterns followed by those CNCF projects and specially the movement to operators in Kubernetes projects.** Hence, we have followed the same path in some of our components and adapt them to the de-facto standard application development model for cloud native application. We have adapted the next components to fully leverage this model:

- Kuryr-Kubernetes movement to CRD management model for Pods, Services, Namespaces and Network Policies. This not only provides better integration into Kubernetes and improved fault tolerance, but also helps with code-base maintenance as well as with debuggability. In fact, as part of those efforts, we also submitted a talk to the Open Infrastructure Summit to showcase the steps to move an application to the CRD/Operators model through the Kuryr-kubernetes

example⁴¹, and therefore helping to set some standard steps that your application may follow to achieve that level of automation.

- TripleMonitorEngine: Operator was generated to deploy and manage 2 of their components, in this case the Manager and the Ingestor (prometheusbeat).

3.3 ICT standardisation in support of EU policies

The 2019 Rolling plan on ICT Standardisation provides a unique overview of standardisation activities in the field of information and communication technologies (ICT) linked to EU legislation and policies, such as healthcare, cloud computing, intelligent transport systems, security, accessibility, Internet of Things, eGovernment, smart grids and many others. It is crucial to clarify this link with European legislation and policies because ICT standards convey important technological requirements for privacy, security and accessibility.

The rolling plan was developed with the help of the Multi-Stakeholder Platform on ICT Standardisation (MSP). The MSP is an expert group composed of representatives of EU and EFTA countries, ICT standardisation organisations, and industry and societal stakeholders. The MSP advises the Commission on all matters related to ICT standardisation. This inclusive and participatory approach to developing the rolling plan ensures wide support from the European and international standardisation organisations, associations, EU countries, and key global industrial and technology companies.

One of the standards for software development and products is the “open source way”. Open Source attracts tremendous technical talent and organisations’ support. Moreover, and very recently the European Commission has adopted its new open source software strategy for 2020 – 2023⁴² where the internal strategy, under the theme “Think Open”, sets out a vision for encouraging and leveraging the transformative, innovative and collaborative power of open source, its principles and development practices. It promotes the sharing and reuse of software solutions, knowledge and expertise, to deliver better European services that benefit society and lower costs to that society. The Commission commits to increasing its use of open source not only in practical areas such as IT, but also in areas where it can be strategic. For the benefit of the BigDataStack project, the strategy is well aligned with its principles as the collaboration across the Commission, with Member States, companies and the public at large for building new, innovative digital solutions that work across borders and towards technological sovereignty is key – these are paramount in the three pillars mentioned above.

3.3.1 A common standard for the referencing of open data

⁴¹ <https://www.openstack.org/summit/2020/summit-schedule/events/24615/k8s-controller-and-crd-patterns-in-python-kuryr-kubernetes-case-study>

⁴² https://ec.europa.eu/info/news/european-commission-adopts-new-open-source-software-strategy-2020-2023-2020-oct-20_en

With the continuously growing amount of data and the increasing amount of open data, interoperability is increasingly a key issue in exploiting the value of this data. Standardisation at different levels (such as metadata schemata, data representation formats and licensing conditions of open data) is essential to enable broad data integration, data exchange and interoperability with the overall goal of fostering innovation based on data. A common standard for the referencing of open data in the European open data portals would be useful. A candidate for a common standard in this area is the Application Profile for data portals in Europe (DCAT) and the FIWARE open stack-based specification and open standards APIs.

BigDataStack partner NEC is currently applying “dynamic orchestrator” into FogFlow to show the benefit of adaptive orchestration of data-intensive IoT services. FogFlow is a fog computing framework that brings context information management to the edge of the networks. FogFlow is compliant with the NGSi standard utilized by FIWARE for context information management. FogWare is included in the FIWARE catalogue of curated framework of open source platform components (Generic Enablers) which can be assembled together to accelerate the development of Smart Solutions.

3.3.2 Analytics insights and scaling policies for microservices

BigDataStack partner UBI is working on this topic. The work in progress gives guidelines for platform-agnostic elasticity management of elastic-by-design microservices. Elasticity refers to the way a microservice is reacting to the increase or decrease of its load. Microservices that have the ability to scale in/out are considered elastic-by-design. The scope of our work is to set the guidelines for platform agnostic elasticity management. Platform agnostic implies that the mechanism/orchestration entity which will perform the actual scale-in/out process remains agnostic to the cloud infrastructure details. Instead, the definition of the actual elasticity events and the relationship of this definition with the underlying monitoring mechanisms will be formally described.

4 Adoption Roadmap

According to the International Data Corporation (IDC), worldwide revenues for Big Data and Business Analytics (BDA) solutions would achieve &189.1B for 2019 and will maintain this trend of growth throughout the 2019-2022 forecast with a CAGR of 13.2%. By 2022, IDC predicts a worldwide BDA revenue of \$274.3B⁴³.

Several key drivers are boosting the adoption of BDA by the organisations, such as the need of digital transformation; large enterprises dominate the market with a share of more than 60% and SMEs segment is expected to grow at a remarkable pace up to 2022⁴⁴; Gartner predicted that “by 2022, 90% of corporate strategies will explicitly mention information as a critical enterprise asset and analytics as an essential competency, and more than half of major new business systems will incorporate continuous intelligence that uses real-time context data to improve decisions”⁴⁵.

The industries that are currently investing the most in BDA solutions are banking, discrete manufacturing, professional services, and process manufacturing. On the other hand, the industries that will lead the fastest BDA growth are securities and investment services (15.3% CAGR) and Retail (15.2% CAGR)⁴⁶. In terms of deployment, more than 70% of BDA software revenues in 2019 went toward on-premises solutions, while the adoption of BDA public cloud solutions will represent more than 44% of total BDA software in 2022⁴⁷.

Although the use of BDA by companies in all vertical sectors is steadily growing, there are still barriers and challenges to face for the adoption of BDA by enterprises.

The following section provides an overview of the key challenges and barriers in the adoption of Big Data technologies within companies and vertical industries for integrating BDA in their processes decision-making (see sec. 4.2 of this document for potential adopters and 4.3 for methodology).

4.1 Big Data adoption challenges and barriers

Big Data Adoption (BDA) challenges are not only limited to technical ones but also include those that prevent companies from the shift toward a data-driven decision-making culture. Companies’ objectives over the last decade are becoming “data-driven”, and their larger goal is managing enterprise data. Companies have made a huge effort to treat data as an asset and align their culture in a more data-oriented way.

⁴³ <https://www.idc.com/getdoc.jsp?containerId=prUS44998419>

⁴⁴ <https://www.prnewswire.com/news-releases/the-global-big-data-analytics-market-2027-a-105-billion-opportunity-assessment-301014418.html>

⁴⁵ https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

⁴⁶ <https://www.idc.com/getdoc.jsp?containerId=prUS44998419>

⁴⁷ <https://www.idc.com/getdoc.jsp?containerId=prUS44998419>

Technical challenges

The existing technical challenges for the BDA adoption are those related to the ever-growing and disparate nature of enterprises' data. The enterprise data landscape is currently suffering from problems related to the data storage, consumption and quality.

The key challenges identified regarding data management are⁴⁸:

- Incapacity to store and process the amount of data generated by companies
- Understanding and prioritizing the useful data from the noise (90% of all data are noise),
- High latency between when data is generated and consumed
- Lack of mechanisms to verify the authentication and precision of data
- Lack of cost-effectiveness analysis of storing and processing a high amount of data in real-time and determining how long to keep the data
- Complexity in the integration of non-traditional data and traditional data from other existing companies' systems for data analysis
- Deciding on whether to use private or public cloud computing or on-premises to store the data
- Compromising and balancing BDA with data confidentiality requirements

Nevertheless, data management challenges aren't the only technical challenges to face in the adoption of Big Data technologies within companies and across vertical sectors. Here, we present a summary of the main technical challenges identified⁴⁹:

- **Interoperability:** Many companies have already invested in business intelligent systems for their processes. New big data solutions must be integrated with existing BI systems, preferably in a seamless way, which can be quite difficult and expensive. Additionally, some issues related to data ingestion from upstream systems may arise.
- **Manageability:** Infrastructure management becomes a problem when companies must deal with a big cluster of nodes. Some vendors are providing an integrated offering to deal with manageability.
- **Security:** Data must be controlled correctly within the company context. If not, it can lead to compliance issues, exposure of data, data loss, etc.
- **Maturity:** New big data solutions vendors emerge every day. It is crucial to choose a vendor that can offer support for data management and able to provide end to end solutions.
- **Governance:** Identification of data protection requirements and a proper governance strategy, while revisiting existing external and internal data policies regulatory framework.
- **Development maintainability and scalability:** Tools for testing, deployment and administration tools foster the Big Data adoption and help in the maintenance phase. Tools for development and maintenance of Big Data solutions will help companies to faster adopt big data solutions.

⁴⁸ <http://jitm.ubalt.edu/XXV-4/article4.pdf>

⁴⁹ <http://jitm.ubalt.edu/XXV-4/article4.pdf>

As previously mentioned, barriers for BDA adoption include not only technological obstacles but also social, organisational, economic and external barriers. Next, a summary of non-technological barriers is presented.

Economic Barriers

Despite the decreasing cost of information technology, companies have realised the still high cost of BDA adoption, as new IT infrastructures, analysis tools and human skills are needed in most cases.

Additionally, companies experience difficulties in measuring the return on investment and justifying it due to the lag time between the BDA investment and its returns.

According to Bi-Survey50, firms asked about revenues and costs had difficulties to put them into numbers, just a third of participants who had integrated BDA answered this question, and they provided the numbers depicted in the figure below.



Are you able to quantify the benefits from your big data initiative(s)?

Figure 31: Revenues and cost using BDA

Organisational Barriers

According to the Big Data and AI Executive Survey 2019⁵¹ conducted by NewVantage Partners, executives surveyed cite several challenges for BDA adoption, however, 95% of them are attributable to organisational issues. Organisational barriers identified are lack of organisational agility, cultural resistance, value the data as an asset and lack of executive leadership. Almost half of the firms surveyed (40.3%) identify “lack of agility” as the most important challenge to BDA adoption within companies’ business. This “lack of agility” is due to obsolete organisational structure and policies and top management barriers,

Human Barriers

The same report revealed that the main challenge for companies to become data-driven is cultural, due to people’s resistance and the process issues. Lack of awareness, BDA skills/experience and benefits from adopting BDA are some of the barriers identified that prevent the digital transformation that companies need to address for BDA adoption.

⁵⁰ <https://bi-survey.com/big-data-value>

⁵¹ <http://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf>

According to a survey conducted by BI-Survey, the “lack of analytical know-how in our company” represents the 53% of the challenges identified and the 34% of “company processes are not mature enough for the use of BDA technologies”, in contrast with the 26% of “technical problems”⁵². Human barriers are considered one of the top concerns preventing companies from adopting of big data technologies.

External Barriers

External barriers refer to the availability of proper public IT infrastructures, BDA vendors or the governments’ rules and regulations. Governments are called upon to create IT standards, promote regulations to foster BDA adoption and provide public IT infrastructure. BDA combined with Blockchain and IoT are fuelling the Cloud adoption, and it is needed that public organizations support the development of the Public Cloud⁵³. According to IDC, the revenue for BDA software delivered by public Cloud will represent more than 44% of the total opportunity in 2022⁵⁴. Additionally, the existence of a suitable BDA vendor capable of influencing and encouraging companies’ executives in decision-making to adopt BDA is crucial. The BDA vendors who provide customised and easy to use solutions, with data integration capabilities, are the best valued by companies.

A summary of the outcomes from the survey conducted by Bi-Survey is depicted in the figure below. The survey also reveals that the challenges with big data analytics adoption vary by industry.



Figure 32: Problems envisioned by companies when using big data technologies/analytics

⁵² <https://bi-survey.com/challenges-big-data-analytics>

⁵³ <https://www.umbrellainfocare.com/blogs/cloud-is-driving-adoption-of-big-data-analytics>

⁵⁴ <https://www.umbrellainfocare.com/blogs/cloud-is-driving-adoption-of-big-data-analytics>

4.2 Application of Big Data and challenges by Industry Verticals

In recent years, Big Data has become a game-changer in all industries and most organisations have several objectives for adopting Big Data in their business.

The main goal for most companies is to enhance customer experience, moreover, others are cost reduction, make processes more efficient, or develop new business models. According to a recent Deutsche Bank report⁵⁵, the major objective of big data application is customer focus (49%) followed by Operational optimisation (18%) and new business models (14%).

Here, we present a summary of the main Big Data applications and challenges in different industry verticals⁵⁶⁵⁷⁵⁸⁵⁹⁶⁰⁶¹⁶².

INDUSTRY	APPLICATIONS	CHALLENGES
Banking & Securities	<ul style="list-style-type: none"> Prediction and prevention of fraudulent activities Improved market trading analysis Enterprise risk management Customer engagement 	<ul style="list-style-type: none"> Meeting regulatory compliance Inadequate technology to process high-velocity data from the banking market
Communications, Media & Entertainment	<ul style="list-style-type: none"> Create content for the specific target audience Content recommendation on demand 	<ul style="list-style-type: none"> Leveraging social media content The understanding of the real-time pattern of media content usage
Healthcare	<ul style="list-style-type: none"> Electronic Health Records (EHR) Real-time alerting Enhancing patient engagement Predictive analytics in Healthcare Telemedicine Tracking the spread of chronic disease 	<ul style="list-style-type: none"> Electronic data is unavailable, non-accessible or inadequate Use of data from available sensors Ethical and legal issues
Education	<ul style="list-style-type: none"> Enhancing students results Customised programs and courses Reduce the number of dropouts Measure system education effectiveness 	<ul style="list-style-type: none"> Integration of data from different sources on different vendors platforms Privacy and personal data protection
Manufacturing	<ul style="list-style-type: none"> Predictive maintenance of equipment Predictive maintenance after-sale or lease Improve processes efficiency (supply chain) Optimise logistics 	<ul style="list-style-type: none"> Bringing the IIoT capabilities to older or heritage devices Software tools to assess the impact of decisions Infrastructure able to support ever-growing data

⁵⁵ [https://cib.db.com/docs_new/GTB_Big_Data_Whitepaper_\(DB0324\)_v2.pdf](https://cib.db.com/docs_new/GTB_Big_Data_Whitepaper_(DB0324)_v2.pdf)
⁵⁶ <https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications>
⁵⁷ <https://data-flair.training/blogs/big-data-applications/>
⁵⁸ <https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/>
⁵⁹ <https://www.upgrad.com/blog/big-data-applications-in-banking-insurance/>
⁶⁰ <https://www.datapine.com/blog/big-data-examples-in-healthcare/>
⁶¹ <https://www.upgrad.com/blog/big-data-applications-in-education/>
⁶² https://bdva.eu/sites/default/files/BdVA_SMI_Discussion_Paper_Web_Version.pdf

	<ul style="list-style-type: none"> Improve the safety of industrial products and services 	
Government	<ul style="list-style-type: none"> Help support decision-making in real-time Applications to face overcome challenges such as poverty, job creation, terrorisms, natural disasters, etc. Projects to enhance citizens' well-being and engagement in public affairs 	<ul style="list-style-type: none"> Integration and Interoperability among different departments Quality and accessibility of data (legacy data and silos) Data Security Skills gap
Insurance	<ul style="list-style-type: none"> Tailored insurance packages to car drivers Health insurance provides a premium discount for active people (data from wearables and other devices) Pricing strategies Claims processing 	<ul style="list-style-type: none"> Securing digital data Data usage regulation Demand for customer-centric analytical solutions
Transport	<ul style="list-style-type: none"> Traffic control, route planning, prediction of traffic conditions Intelligent transport systems Failure predictions Optimising freight movement Supply chain optimisation 	<ul style="list-style-type: none"> The great amount of data from location-based social networks and telecoms Security issue and personal privacy protection
Energy & Utilities	<ul style="list-style-type: none"> Smart meter for collecting data of consumption and provide tailored recommendations Grid management systems 	<ul style="list-style-type: none"> Sharing data of operational process in such a competitive industry with big players Make data useful and generate actionable insights Deploy new IT systems to collect data
Retail	<ul style="list-style-type: none"> Timey inventory analysis Customer prospecting, customer retention Tailored shopping recommendations Anticipate changes in demand Customer insights and marketing analytics 	<ul style="list-style-type: none"> Ensuring data security and compliance Collecting data from disparate systems Earn customer trust to capture their data

Table 3: BDA Applications and Challenges by vertical sectors

4.3 Adoption Roadmap for BigDataStack offering

While it is certain that Big Data brings multiple benefits for companies, the proper adoption roadmap is still confusing and can vary depending on the industry and the company culture. The adoption of Big Data technologies within the company projects must start with a strong roadmap to overcome the challenges and barriers previously identified and consider the ever-evolving Big Data technologies landscape.

BigDataStack solutions provide a complete high-performance stack of technologies to fully handle data operations and analytic applications of the big data projects.

BigDataStack offers the data management and analytics needs of a BDA project without the need of combining solutions from different providers along the adoption roadmap.

Additionally, BigDataStack components developed within the project provide tools for the different actors in the adoption of big data analytics projects.

Three business use cases from different sectors have been implemented within BigDataStack project, and D6.2 Use case description and implementation Y3 shows how BigDataStack offering has been used from the definition of the business use cases to the implementation and integration with existing organisation IT systems.

Following the technical process for the adoption of BigDataStack offering in the project use cases, and considering the barriers and challenges mentioned in the previous sections, we propose an adoption roadmap that comprises six phases and encompasses the most relevant challenges in the adoption of Big Data projects.

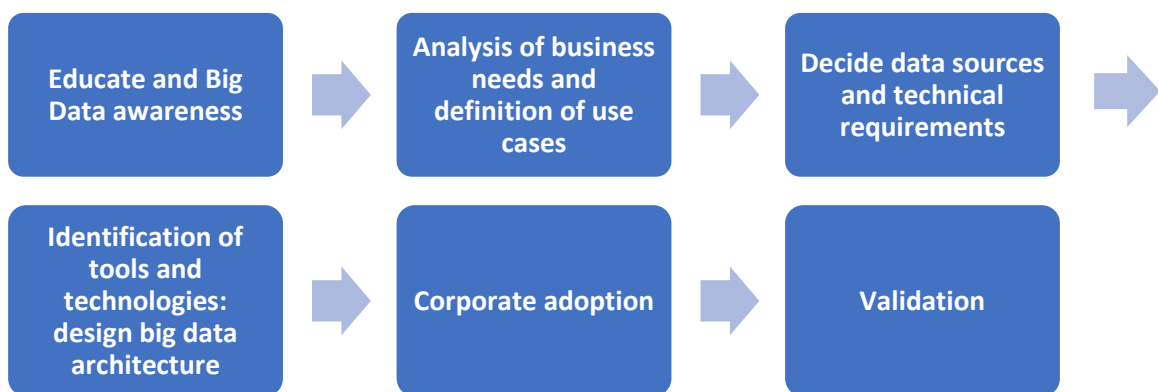


Figure 33: Adoption Roadmap for BDA projects

Phase 1: Educate and Big Data awareness

The starting point for the adoption of big data analytics into organisations is an initial phase of awareness and knowledge about big data benefits and opportunities for the organisation. This phase aims at overcoming the cultural reluctance of the organisations to adopt Big Data technologies.

Within this phase, organisations gather information on their market context, big data landscape (solutions, providers, etc.), realise of the value of analysing information and data, and analyse how Big Data could help to support the creation of new products or services, or develop new business models with new revenue streams or reducing costs.

This phase also addresses the cost of big data adoption considering the digitisation level of the organisation and the stage in its “data-driven” transformation.

Phase 2: Analysis of business needs and definition of use cases

After the initial phase, the organisation must elicit business objectives and priorities and identify the problems and needs to achieve those business goals, after that translate then into the business case based on measurable outcomes. Establish criteria success factors and KPIs, and actors’ role in the use case.

Within this phase, a tight collaboration among different departments (business department, IT department, executives, etc.) is crucial, and different activities namely,

innovation or design thinking workshops and brainstorming are carried out to create use cases and new ideas.

Phase 3: Decide data sources and technical requirements

Once business use cases have been defined, organisations must assess their data landscape and what type of data is needed for the use case: easily available data sources or data that could be obtained externally from stakeholders or social media, data at rest or real-time data, etc. A deep understanding of the organisation's lack of existing data and how they can relate will create new use cases beyond the known new ones.

Additionally, considerations about data governance model, data quality, storage, security, or property are discussed and established in this phase.

Within this phase, technical requirements are evaluated to determine the IT readiness for big data adoption: the right technical team and new technical roles, existing IT systems, infrastructure requirements, data storage capabilities, etc. Furthermore, other technical requirements such as visualisation needs, analytics algorithms for business use cases and deployment requirements are set-up at this stage.

Phase 4: Identification of tools and technologies: design big data architecture

Depending on the business cases and organisation objectives, it is needed to identify what BDA tools and technologies are better to address them. Moreover, an assessment of the existing IT systems and technologies in the organisation that will be integrated with the BDA solution should be carried out to reduce any integration barriers. In many cases, it is better to align and make compatible new systems with the existing platforms.

The new architecture designed must have the scalability, governance, and security to work with the data architecture previously defined.

Phase 5: Corporate adoption

Implementation of the business use case with BDA tools and adoption in production. At this phase, the big data architecture selected is already deployed either on-premises or in the cloud and the BDA application and systems are running and outcomes will trigger process actions.

The organisation must ensure that each team understand how BDA will help in its daily work. Training sessions and workshops will be organised to help departments how to use the BDA solution to obtain actionable insights and turn into organisation competitive advantage.

Phase 6: Validation

Assessment of the KPIs and criteria success factors and the advantages gained using Big Data solutions in the business use cases. Evaluation of the implementation cost of Big Data solutions and measuring of the ROI.

As the Business use case grows, demonstrate that the BDA adoption success is repeatable and reliable, and validate the assumptions made in the previous adoption roadmap phases. Adjust cost while increasing capability and agility.

Here, we present a summary of the Adoption Roadmap for BigDataStack offering from the organisation perspective.

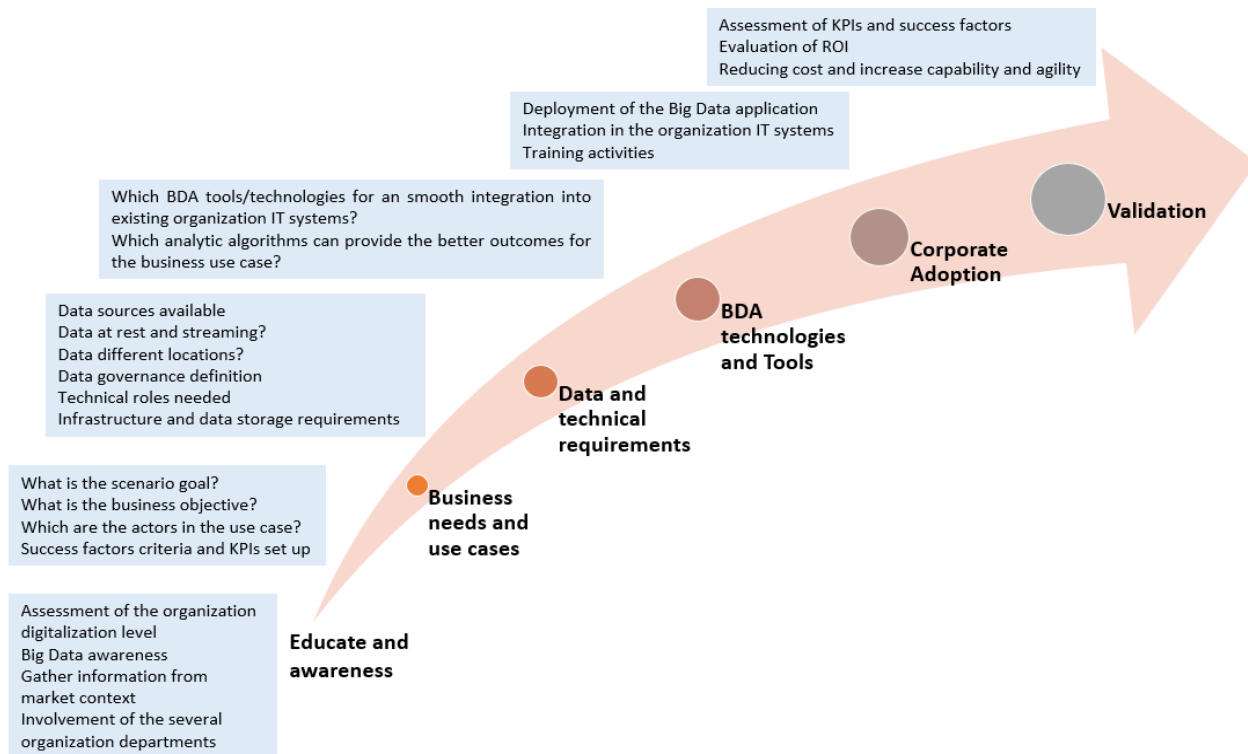


Figure 34: Adoption Roadmap

4.4 Activities to validate the BigDataStack adoption roadmap

During the last year of the project, two activities have been carried out within WP7 to gather feedback from partners and stakeholders regarding BigDataStack adoption and validate the roadmap described in section 4.2.

The BigDataStack consortium organised a remote workshop among partners to discuss barriers to adoption of BigDataStack solutions in other vertical sectors such as Healthcare, Agriculture, Manufacturing, Energy, etc. All partners participated and provided different viewpoints based on their work within the project, as well as providing recommendations for adoption.

<p>Do you consider any relevant technological barrier for the adoption of BigDataStack solutions?</p>	<ul style="list-style-type: none"> • Installation. Maybe a one-click installation would be easier for the adoption • Availability of large data sets • Accurate data collection methods (IoT devices/wearables) • Data warehouse technology is not explicitly included nor validated (through use cases) in the context of the BigDataStack project
--	---

	<ul style="list-style-type: none"> • BigDataStack resource cluster management layer is based on Openshift/Kubernetes and requires a mature underlying infrastructure, ideally, with one or more virtualisation layers. We may not find this maturity in the infrastructure found in the smart agriculture sector • Definition and monitoring of SLOs may require that BigDataStack components (QoS, DO, 3M) access data from the applications that run on BigDataStack. This data might not always be accessible if data is encrypted, which is usually the case in healthcare applications • As a general remark most of the Project deal with infrastructure and therefore is not impacted by the kind of big data that is used
<p>Do you envision any legal barrier to adopt BigDataStack solutions?</p>	<ul style="list-style-type: none"> • Data streaming without encryption • Data privacy, GDPR considerations • The infrastructure needs to be secure, in order to ensure the privacy and security of the data • As far as you are GDPR compliant, not our software stack but the underlying infrastructure, the only problems envisioned are with the security certification of the infrastructure provider (public cloud, data centre) as we already had in this project with testbed ISO27001 certification • Definition and monitoring of SLOs may require that BigDataStack components (QoS, DO, 3M) access data from the applications that run on BigDataStack. This means that the use of this data for scaling applications should be included in the GDPR and any other necessary legal agreements with the users
<p>Do you consider any problem with the type of data that other sectors may use to adopt BigDataStack solutions?</p>	<ul style="list-style-type: none"> • Data privacy and security while data is transferred and stored (Anonymisation in case of Healthcare Sector) • The data have not a universal format. In most cases, every country has a different format, so a standard should be defined (Healthcare sector) • Highly distributed sensors capturing data that will need some hubs in the middle to aggregate the data • It is maybe that image data are not supported by some of the functions of BigDataStack to the extent they could be • Massive volume due to high frequency of produced data • Agriculture sector usually requires the image processing and analysis of unstructured data, where the BigDataStack tools are not suitable for this
<p>What would you recommend considering when adopting BigDataStack?</p>	<ul style="list-style-type: none"> • Data requirements • Amount of data to process and their real-time adaptation needs • Focus on GDR and other legislations for sensitive data processing • Careful evaluation of Quality of Service criteria • Adaptation of techniques from use cases developed during BigDataStack • Involvement of Business department and Data analysts • Privacy considerations regarding scaling out using resources from an external infrastructure provider

- Training effort within sectors less familiar with BDA technologies such as Agriculture
- Follow standards adopted by the sector
- IT organisations must be mature enough to adopt big data technologies and face big data use cases. That is not normally the case, organisations working in the Healthcare sector using small data (conventional RDBS and warehouse systems). You need to mature your data governance practice and have the maturity to face big data analytics. Then, you may think you use BigDataStack to ease your developments in that, but never to substitute your conventional analytical processes and solutions

Additionally, during the final BigDataStack event at the EBDVF2020 held 3-5 November 2020, titled “A BigDataStack for industry – a focus on retail, shipping and insurance: Is Big Data the real future of emerging business?” a discussion panel with Big Data experts from the application domains of BigDataStack as well as standardisation & Open Source was designed to raise awareness on the BigDataStack results and the benefits of adopting big data solutions for industry.

Detailed findings of this event are detailed in **D7.3 Exploitation Plan and Business Potential**. The main conclusions for the benefit of the adoption roadmap are summarized as follows:

- The adoption of Big Data technologies is closely related to the Digitalization of companies
- The big data aim is transforming data in actionable insights for companies
- The crucial importance of an initial phase of education and big data awareness within the company. Raise awareness on how Big Data can help in all the processes of the company and what type the data the company already has.
- The need to correctly define the business use case, making the correct questions, and eliciting the right data, etc. Involvement of the different departments of the company (business department, operation department, sales, etc.) to help to define the use case.
- Data management, data cleaning, data retrieving, data governance, etc. is critical for big data analytics. So, mechanisms to overcome barriers related to data must be put in place.
- Visualization tools are very important to better know the insights of analytics
- CE is focused on the data sovereignty and encourage to the big data ecosystem to develop standards and use of OS solutions and technologies
- OS solutions and technologies are a driver for the adoption of Big Data
- Greater investments are needed in all sectors to adopt big data technologies. Digitalization, decarbonization, education and research are some of the drivers for adopting big data

5 Lessons learned and legacy

Dissemination and stakeholder engagement benefitted highly from the synergies created with other European Initiatives, experts in the Big Data field as well as the activities performed in the Open Source Communities. This helped the consortium to effectively move towards a fully online engagement in the third year in spite of the COVID-19 pandemic.

The BigDataStack legacy is certainly the website as a knowledge hub, with all project results, publications, communication material and training resources. The ZENODO community provides access to the full collection of BigDataStack deliverables, for researchers and big data experts to reuse.

By engaging with **upstream open source communities** and **contributing code to the respective upstream** repositories, not only a bigger impact can be made⁶³, but also a long-term support of the developed features is ensured. The extra effort of contributing to upstream projects is worth its effort, as it ensures not only the feature is more solid and suitable for more use cases, but also better covered by CI systems and hence maintained. In addition, the code will become more visible and potential adoption by other companies is fostered.

The Open Source Initiative efforts collaborate with the kick off of the "**Red Hat Research Day**". This now periodic event gathers both developers and researchers, helping to better highlight the benefits of making research together with Open Source tooling. There are several benefits to this: a) making a bigger impact by making adoption easier; b) ensure innovation on the projects where the contributions land; c) it provides an easier way of reproducing research results, which is a must in research communities.

The extra efforts about supporting others in Open Source communities, such as getting involved with the **Outreachy program** helps raising awareness on certain Open Source projects, and raise the interest as well as the number of contributions and contributors.

The BigDataStack adoption roadmap and recommendations will be made available through ZENODO and the BigDataStack website for companies to consult when interested in adopting the BigDataStack solution, either as single key offerings or as a whole.

⁶³ <https://www.sdxcentral.com/articles/news/att-verizon-5g-deployments-boost-openstack-work/2020/10/>

6 Annex 1 - BigDataStack publications

Scientific Publications	Type	channel
<p>BigDataStack: A holistic data-driven stack for big data applications and operations <i>D. Kyriazis, C. Doulkeridis, P. Gouvas, R. Jimenez-Peris, A. J. Ferrer, L. Kallipolitis, P. Kranas, G. Kousiouris, C. Macdonald, R. McCreadie, Y. Moatti, A. Papageorgiou, M. Patino-Martinez, S. Plitsos, D. Pouloupoulos, A. Paradell, A. Raouzaïou, P. Ta-Shma, V. Vianello</i></p>	Scientific publication	presented at: 2018 IEEE International Congress on Big Data
<p>Integration of Mobility Data with Weather Information <i>N. Koutroumanis, G. Santipantakis, A. Glenis, C. Doulkeridis and G. Vouros</i></p>	Scientific publication	presented at: 2nd Workshop on Big Mobility Data Analytics (BMDA), in conjunction with EDBT'19
<p>Scalable Enrichment of Mobility Data with Weather Information <i>N. Koutroumanis, G. Santipantakis, A. Glenis, C. Doulkeridis and G. Vouros</i></p>	Scientific publication	presented at: Geoinformatica Journal (Springer)
<p>Variational Bayesian Context-aware Representation for Grocery Recommendation <i>Z. Meng, R. McCreadie, C. Macdonald and I. Ounis</i></p>	Scientific publication	presented at: 13th ACM Conference on Recommender Systems (RecSys19) - CARS 2019 Workshop
<p>Parallel Query Processing in a Polystore <i>B. Kolev, O. Levchenko, E. Pacitti, P. Valduriez, R. Jimenez-Peris, P. Kranas, M. Patino-Martinez</i></p>	Scientific publication	presented at: Journal of Distributed and Parallel Databases
<p>Extensible Data Skipping <i>P.Σ Ta-Shma, G. Khazma, G. Lushi, and O. Feder</i></p>	Scientific publication	presented at: 2020 IEEE International Conference on Big Data
<p>Predictive Maintenance Leveraging Machine Learning for Time-Series Forecasting in the Maritime Industry <i>G. Makridis, D. Kyriazis, S. Plitsos</i></p>	Scientific publication	presented at: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)
<p>AutoClust: A Framework for Automated Clustering based on Cluster Validity Indices <i>Yannis Poulakis, C. Doulkeridis, D. Kyriazis</i></p>	Scientific publication	presented at: 20th IEEE International Conference on Data Mining (ICDM'20)
<p>Real-time adaptable resource allocation for distributed data-intensive applications over cloud and edge environments <i>J.D. Totow, D. Kyriazis</i></p>	Scientific publication	presented at: 9th IEEE International Conference on Cloud Computing in Emerging Markets
<p>Exploring Data Splitting Strategies for the Evaluation of Recommendation Models <i>Z. Meng, R. McCreadie, C. Macdonald and I. Ounis</i></p>	Scientific publication	presented at: 14th ACM Conference on Recommender Systems
<p>BETA-Rec: Build, Evaluate and Tune Automated Recommender Systems <i>Z. Meng, R. McCreadie, C. Macdonald and I. Ounis</i></p>	Scientific publication	presented at: 14th ACM Conference on Recommender Systems
<p>Leveraging Data-driven Infrastructure Management to facilitate AIOps for Big Data Applications and Operations</p>	Scientific publication	presented at:

<i>R. Mccreadie, J.D. Totow, M. Fadel Argerich, G. Kousiouris, P. Kranas, O. Avila García, A. Castillo Nieto, D. Kyriazis, J. Soldatos, I. Ounis, C. Macdonald</i>		Book Chapter for Technology and Applications for Big Data Value
Tutor4RL: Guiding Reinforcement Learning with External Knowledge <i>M. Fadel Argerich, J. Fürst, B. Cheng</i>	Scientific publication	presented at: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering 2020
Towards knowledge infusion for robust and transferable machine learning in IoT <i>J. Fürst, M. Fadel Argerich, B. Cheng & E. Kovacs</i>	Scientific publication	presented at: Open Journal of Internet Of Things (OJIOT)
SYSTOR '19: Proceedings of the 12th ACM International Conference on Systems and Storage May 2019 Pages 193 https://doi.org/10.1145/3319647.3325854	Poster presentation	Systor 2019
BlogPost: Accelerate your OpenShift Network Performance on OpenStack with Kuryr <i>R. A. Rodriguez, S. S. Malleni, L. Tomas Bolivar</i>	Publication in external channel	openshift.com
BlogPost: Autoscaling with OpenShift on OpenStack <i>A. Simonelli</i>	Publication in external channel	openshift.com
BlogPost: Customizing and tuning the Kuryr SDN for Red Hat OpenShift 3.11 on Red Hat OpenStack 13 <i>M. Ahmad, L. T. Bolivar</i>	Publication in external channel	developers.redhat.com
Reference Architecture: New OpenShift 4 on Red Hat OpenStack Platform Reference Architecture <i>A. Simonelli</i>	Publication in external channel	openshift.com
Red Hat Research Quaterly <i>Red Hat Research</i>	Publication in external channel	research.redhat.com
Data Skipping for IBM Cloud SQL Query <i>P. Ta-Shma</i>	Publication in external channel	ibm.com
Accelerate Your Big Data Analytics and Reduce Costs by Using IBM Cloud SQL Query <i>G. Khazma, P. Ta-Shma</i>	Publication in external channel	ibm.com
How to Layout Big Data in IBM Cloud Object Storage for Spark SQL <i>P. Ta-Shma</i>	Publication in external channel	ibm.com
Big data skipping in the cloud <i>O. Feder, G. Khazma, G. Lushi, Y. Moatti, P. Ta-Shma</i>	Publication in external channel	scm.org

7 Annex 2 - BigDataStack at events

Event	Stakeholder targeted	Title	Activity	What did we gain?
The Open Infrastructure Summit, May 2018, Vancouver Canada,	Open Source Community	N/A	Network Policy support at Kuryr	
EBDVF18 12-14/11/2018 Vienna, Austria	Industry	AI everywhere: technical challenges	Presentation and discussion BDVe organised panel "AI everywhere: technical challenges"	<ul style="list-style-type: none"> ·Awareness raised with researchers, industry and policy makers on infrastructure-related offerings towards AI ·Liaisons with BDVe & Cross-CPP (ICT-16 project)
ICT 2018 4-6/12/2018 Vienna Austria	Industry	Data Democratization: empowering the citizens in the digital transformation	BigDataStack joined the discussion panel on Data Democratization lead by BDVA	<ul style="list-style-type: none"> ·Co-written report with BDVA, with EW-SHOPP, Network of Living Labs and BDVe. ·Raised awareness BigDataStack & offerings with researchers, industry and policy makers
Posidonia 4-6/06/2018 Athens, Greece	Industry	BigDataStack	Demonstration preliminary results BigDataStack shipping use-case	Awareness raised on preliminary results of BigDataStack at international maritime exhibition attended by 39.000 international maritime community members
Think2019 12-15/02/2019 San Francisco, USA	Industry	Enterprise-Scale Analytics Performance with Cloud Object Storage	Presentation	Awareness raised on BigDatastack Enterprise-scale Analytics Performance with Cloud Object Storage and its practical implementation for real-time shipping among industry stakeholders at THINK2019 (attended by 40.000 IBM customers).
RedHat Summit 2019 27-30/ 04/ 2019 Boston, USA	Industry	BigDataStack	RedHat booth	Awareness raised among Upstream engineers
IEEE2018 02-07/07/2018 Boston, USA	Research & Academia	A holistic data-driven stack for big data applications and operations	Scientific paper presentation	<ul style="list-style-type: none"> ·Awareness raised with technical, researcher and academic audiences ·Open access published scientific paper, viewed 59 times, cited 1 time

BMDA 2019, 26-28/ 03/ 2019 Lisbon, Portugal	Research & Academia	Integration of Mobility Data with Weather - Nikolaos Koutroumanis et.al. (UPRC)	Scientific paper & poster presentation	·Awareness raised with technical, researcher and academic audiences
WF IoT 2019 15-18/ 04/ 2019 Limerick, Ireland	Research & Academia	Reinforcement Learning based Orchestration for Elastic Services	Scientific paper presentation	·Awareness raised with technical, researcher and academic audiences
RedHat Summit 2019 27-30/ 04/ 2019 Boston, USA	Research & Academia and Industry	Bridging between upstream engineering and research	RedHat Research Day around Data processing and booth	·Facilitated the discussion between upstream engineering and research ·join efforts with Open Data Hub: developing our components following the Cloud Native guidelines facilitating adoption BigDataStack outputs Awareness raised among Upstream engineers
SYSTOR 2019 06/06/ 2019 Haifa, Israel	Research & Academia	BigData Skipping in the Cloud	Poster presentation	·Awareness raised with technical, researcher and academic audiences
IEEE 2019 8-13/07/2019 Milan, Italy	Research & Academia	Fog Function: Serverless Fog Computing for Data Intensive IoT Services	Scientific paper presentation	·dynamic orchestrator applied into FogFlow (EU open source initiatives for IoT) to show benefit adaptive orchestration of data-intensive IoT services.
OpenStack summit 2019 29/04-01/05/2019 Denver, USA	Open Source Community	N/A	RHT was present at OpenStack summit with the Kuryr upstream work in relation to BigDataStack developments	•Raised awareness on Kuryr upstream work (network policies implementation), in relation to BigDataStack •Engagement policy makers in relation to open standards
KubeCon + Cloud Natives 2019 20-23/ 05/ 2019 Barcelona, Spain	Open Source Community	N/A	Participation KubeCon/CloudNativeCon	•Learned how to contribute/collaborate with Kubernetes communities •Learned how to build Cloud Native applications •Submitted a talk to KubeCon SanDiego 2019
BDV PPP Summit 2019 26-28/ 06/ 2019 Riga, Latvia	EC initiatives & projects and policy makers	Lowering barriers for the adoption of Big Data Analytics	Presentation and participation discussion panel Policy4Data, organised by BDVe	•Awareness raised on BigDataStack and its offerings with technical and non-technical audience from industry, policy and research. •co-writing policy brief Policy4Data •liaisons with ICT15 and 18 projects: BDVe, Transforming Transport, e-sides

				<p>Technical & Coordination BDVA meetings:</p> <ul style="list-style-type: none"> •Big Data: Techniques, Ecosystems and Value-Driven Applications Booklet •BDVe Data Scientist mobility Program
Thessaloniki International Fair 2019, Greece	Industry & general public	N/A	Booth	Dissemination of BigDataStack results focused on entrepreneurship and innovation
IBM Cloud Research Seminar	Research & Academia	Fog Function: Serverless Fog Computing for Data Intensive IoT Services	Presentation	Knowledge sharing and promotion BigDataStack results
ACM RecSys 15-19 September 2019, Copenhagen Denmark	Research & Academia	Variational Bayesian Context-aware Representation for Grocery Recommendation	Presentation	Knowledge sharing and promotion BigDataStack results
BDV PPP 2020, May, June, July 2020, Online	EC Initiatives & Industry		Joint series of online demos	<ul style="list-style-type: none"> -Knowledge sharing & promotion BigDataStack results -Synergy with ICT14 and 16 projects I-BiDaaS, Track & Know and Policy Cloud
EBDVF2020, 3-5 November 2020, Online	Industry	European Big Data Research for Industry. 3 projects. 7 sectors. 9 applications. 41 software components. Now what?	Joint session with I-BiDaaS, Track & Know	<ul style="list-style-type: none"> -Knowledge sharing & promotion BigDataStack results -Synergy with ICT14 and 16 projects I-BiDaaS, Track & Know and Policy Cloud
EBDVF2020, 3-5 November 2020, Online	Industry	A BigDataStack for Industry - A focus on retail, shipping	Final workshop BigDataStack and booth	<ul style="list-style-type: none"> -Knowledge sharing & promotion BigDataStack results -Synergy with BigData experts from the field and the BDV Task force on standardisation

		and insurance. Is Big data the real future of emerging business?		
Computer Science Researchers @NTNU, Norway	Research & Academia	BigDataStack	Presentation	Knowledge sharing and promotion of results with Computer scientists
The Open Infrastructure Summit	Open Source Community	K8s controller and CRD patterns in Python - kuryr-kubernetes case study"	2 presentations about Kuryr, OpenShift on OpenStack and CRD patterns	Knowledge sharing and promotion of results with Companies working on Cloud technologies Kuryr got mentioned on https://www.sdxcentral.com/articles/news/att-verizon-5g-deployments-boost-openstack-work/2020/10/
Project Teams Gathering, October 2020, Online	Open Source Community	Kuryr	2 sessions for Kuryr	Knowledge sharing and foster the upstream uptake of Open Source technologies used in BigDataStack with developers and operators New contributors for Kuryr
Red Hat research Day, January 2020, Brno	Research & Academia and Open Source Community	N/A	Interaction with research activities around Red Hat and Academia in Europe	Knowledge sharing and foster the upstream uptake of Open Source technologies used in BigDataStack
DevConf, January 2020, Brno	Research & Academia and Open Source Community	N/A	Interactions and foster open source around developers and students	Knowledge sharing and foster the upstream uptake of Open Source technologies used in BigDataStack
SIGMOD Conference 2019	Database researchers, practitioners, developers and users	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Rise Conference 2019	Founders, investors, multinationals in digital environments	N/A	Booth	Visibility for the project, awareness, onboarding end-users

Desafía Hong Kong 2019	N/A	N/A	Attendance	Visibility for the project, awareness, onboarding end-users
South Summit 2019	N/A	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Fintech ICEX 2019	N/A	N/A	One-to-one meetings	Visibility for the project, awareness, onboarding end-users
CPONET 2019	N/A	N/A	Booth	Visibility for the project, awareness, onboarding end-users
IoT Solutions World Congress 2019	Stakeholders in IOT, entrepreneurs,	N/A	Booth	Knowledge sharing and foster the upstream uptake of Open Source technologies used in BigDataStack
Avaloq Headquarters talk 2019	N/A	N/A	Talk	Visibility among few but relevant stakeholders in in community management and IT at large
Web Summit 2019	Digital stakeholder at large	N/A	Booth	Visibility for the project, awareness, onboarding end-users in a top-notch event at global level
Fintech Festival 2019	Industries in fintech and banking, SMEs, entrepreneurs	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Smart Cities Brokerage 2019	Stakeholders in Digital transformation, Urban Environment, Mobility, Governance & Finance, Inclusive & sharing cities	N/A	One-to-one meetings	Visibility for the project, awareness, onboarding end-users and engagement with stakeholders

ET & IoT Technology 2019	N/A	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Seoul National University talk 2019	Research & Academia	A Multi-Resource Dynamic Load Balancing Algorithm for Cache Systems	Talk	Awareness and visibility in a wide and international university environment
IEEE Big Data 2019	Database researchers, practitioners, developers and users	NewSQL : principles, systems and current trends	Tutorial	Visibility for the project, awareness, onboarding end-users
Berkeley talk 2019	Research & Academia	N/A	Talk	Awareness and visibility in a wide and international university environment
SBBD Conference 2019	Database researchers, practitioners, developers and users	N/A	Talk	Visibility for the project, awareness, onboarding end-users
Osaka University talk 2019	Research & Academia	N/A	Talk	Awareness and visibility in a wide and international university environment
UCLA talk 2019	Research & Academia	N/A	Talk	Awareness and visibility in a wide and international university environment
UC Irvine talk 2019	Research & Academia	N/A	Talk	Awareness and visibility in a wide and international university environment
Facebook Headquarters talk 2019	N/A	N/A	Talk	Visibility among few but relevant stakeholders in in community management and IT at large

Asian Financial Forum 2020	N/A	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Brokerage Smart Cities Paris 2020	Stakeholders in Digital transformation, Urban Environment, Mobility, Governance & Finance, Inclusive & sharing cities	N/A	One-to-one meetings	Dissemination among general public and stakeholders in brokering and smart cities
Telecoms World Asia 2020	Entrepreneurs in telcom, start-ups and SMEs	N/A	Booth	Visibility for the project, awareness, onboarding end-users
Petrobras headquarters talk 2020	N/A	N/A	Talk	Visibility among few relevant stakeholders in oil management
Smart Cities Brokerage 2020	SME, Large enterprises, Enterprises, Start-ups, Research institutions, Universities, Public associations/agencies	N/A	One-to-one meetings	Individual sharing of results among stakeholders in smart cities and urban mobility worldwide
Smart Manufacturing Brokerage 2020	Stakeholders in Digital transformation, Urban Environment, Mobility, Governance & Finance, Inclusive & sharing cities	N/A	One-to-one meetings	Knowledge sharing and promotion of results with Companies working on Industry 4.0 and Smart Manufacturing
ISM Brokerage 2020	Stakeholders in Digital transformation, Urban Environment, Mobility,	N/A	One-to-one meetings	Knowledge sharing and promotion of results with Companies working on Industry 4.0 and Smart Manufacturing

	Governance & Finance, Inclusive & sharing cities			
Korea Eureka Day 2020	innovators, companies and leaders from Korea and Eureka member countries	N/A	One-to-one meetings	Visibility for the project, awareness, onboarding end-users among European, and EUREKA countries
South Korea: your partner for smart cities 2020	innovators, companies and leaders from Korea and Eureka member countries	N/A	One-to-one meetings	Visibility for the project, awareness, onboarding end-users among European, and EUREKA countries

8 Annex 3 - KPI check

KPIs for Communication Kit and overall visibility				
Measure	Driver	Action	KPI	M34
Monthly Web content	Regular information updates with SEO-driven approach	Identify and publish new content on a regular basis.	YR1: min 2/month YR2: min 3/month YR3: min 4/month tot 108	<input type="checkbox"/> KPI achieved with 138 items of news, blog, video, articles, press releases
In-house newsletters	Different stakeholders are properly informed in a timely manner	Produce and circulate monthly newsletter based on stakeholder targets	YR1: min 6 YR2: min 8 YR3: min 10 tot 24	18 in-house newsletters and 3 direct messages via mailing list
Promotional material, including video content	Specific audiences receive tailored and timely messages	Design and produce focused material (for stakeholders / events)	YR1: min 3 YR2: min 6 YR3: min 12 tot 21	<input type="checkbox"/> KPI achieved with 1 general project presentation, 3 flyers, 1 roll-up banner, 34 video's with 1.6k+ views
Press releases targeting major stakeholders on supply/demand sides	Raise interest and recruit demand/supply side actors	Produce press releases targeting different media channels and audiences	YR1: min 2 YR2: min 1 YR3: min 2 major stakeholder category tot 5	3 press releases Another 2 press releases are planned for the closing of the project.
Press releases targeting general public	Raise interest amongst non-specialised audiences	Lightweight blog/article targeting non-specialised channels	≥ 2 press clippings	2 press releases are planned for the closing of the project.
Visibility in channels used by different stakeholder categories	Ensure back-links/branding recognition through synergies and social media towards general brand recognition	Liaise and engage with projects/initiatives, journalists and LinkedIn groups	≥ 20 back-links across major stakeholders ≥ 50 entities identifying BigDataStack brand	<input type="checkbox"/> KPI achieved with 40 backlinks from 17 different domains

KPIs for the Stakeholder Engagement and Community Development				
Measure	Driver	Action	KPI	M34
Social media content: Twitter & LinkedIn	Grow community, Regular stakeholder engagement gives insights into interests/concerns	Publish tweets and posts, including SMART-based campaigns & monitor outcomes	YR1: min 9/month YR2: min 28/month YR3:min 56/month tot 708	<input type="checkbox"/> KPI achieved with a total of 1034 posts: 615 tweets, 419 LK updates, 10 articles
Stakeholder database (established before GDPR)	Early identification of prospective stakeholders	Develop a profiled DB of contacts for incremental community development and stakeholder engagement	1.500 profiled & engaged stakeholders by M12; over, 2.500 by M24, and 4,000 by M36 tot 4000	+ 2280 engaged connections , with 1707 personal connection on LK, 939 LK company page followers, 366 TW followers, 214 newsletter connections
Targeted workshops with free admission, focusing on user needs, challenges & results,	Ensure outreach to non-specialised audiences	Show BigDataStack to visitors in lively, lightweight environment	≥ 3 targeted workshop , capitalising on the user needs, showcasing results, and liaising with ICT-14 & ICT-15 projects ≥ 50-100 non-specialised attendees	<input type="checkbox"/> KPI achieved In total BigDataStack engaged +1000 attendees from Research and Academia, policy makers, Industry through: 2 project webinars, 1 joint webinar (with ICT-11 project) and a series of 9 joint demos at BDV PPP 2020 (with ICT-14 and ICT-16 projects) , 1 joint webinar at EBDVF2020 (with ICT-14 and ICT-16 projects), 1 final webinar at EBDVF2020
Online and/or F2F training sessions	Ensure general public is "educated" about need to advanced research to address their needs	Provide a service for non- IT savvy to show what the new service means for them	≥ 1 online sessions ≥ 50 non-specialised attendees	<input type="checkbox"/> KPI achieved , 2 webinars for non-IT savvy audience, 1 webinar on the use-case connected consumers, 31 attendees, 1 webinar on Real- Time Ship Management, 37 attendees
F2F interactions with local clusters	Ensure engagement of "real people" at the local level	Work with local clusters to co-host an open day with media presence	≥ 1 local events ≥ 3 appearance in local media	Involvement in Greek Workshop of EU Research & Innovation Maritime Projects

Free trials for general public	Facilitate and drive uptake through early trial testing	Organise free trials after reaching a maturity level	≥ 5 "testers	BigDataStack software components are tested by IBM and RedHat customers
Marketing events, e.g. trade fairs/exhibitions	Ensure direct engagement with major stakeholders	Host a properly decked exhibition stand with demos	YR2: min 1 YR3: min 2	<input type="checkbox"/> KPI achieved with 1 exhibition stand at Posidonia 2018 and a virtual exhibition at EBDVF2020. Danaos Demonstration at the international Maritime Posidonia Bi-annual Exhibition with over 39.000 participants. EBDVF2020 was attended by 1500+ participants

Dissemination KPIs				
Measure	Why	Action	target KPI	M34
Organisation and/or attendance to conferences and exhibitions	Attract customers	10 Conferences 3 exhibitions	100 visitors 10 speakers	<input type="checkbox"/> KPI achieved with 54 3rd party events attended with 22 speakers, 13 Joint booths.
Synergies established at local, national, or international level for uptake of BigDataStack results	Raise Awareness, Attract users (supply or demand)	Events (any type) for F2F discussions	6 synergies established	<input type="checkbox"/> KPI achieved 7 synergies established at events: ICT 2018: panel discussion - synergy with BDVe, RedHat summit researcher day, EBDVF 2018 - Data4AI everywhere panel discussion , BDV PPP Summit: panel discussion with BDV PPP projects and upcoming co-written policy brief, BDV PPP joint series of 9 demos , EBDVF2020 joint session with ICT14 & ICT16 projects, BDVA Open Access Booklet
Synergies at national or international levels for sharing knowledge and standardisation	Strengthen impact via joint efforts	Meeting attendance and common publications	> 5 projects	<input type="checkbox"/> KPI achieved with 7 synergies: KubeCon + Cloud Natives (working group on policies for open source), BDV PPP Summit 2019 coordination and technical meetings, BDV PPP Summit 2020 coordination and technical meetings, Meeting with Open Forum Europe, Open Source Summit - RedHat Kuryr BigdataStack element presented, Joint webinar with INFINITECH & BDVA Task Force, Final event joint discussion with BDV Task Force lead Standardisation and SME
On-site demonstrations and presentations	Attract customers Raise awareness	6 demonstrations 6 presentations	3 responders 2 on-site demos	<input type="checkbox"/> KPI achieved with 3 online demos in Big Data Pilot Demo Days and final BigDataStack event at EBDVF2020

Open Access publications	Scientific dissemination	Publication to journals and magazines	> 20 publications	<input type="checkbox"/> KPI achieved with 14 scientific articles and underway one chapter of the BDVA “Technologies and Applications for Big Data Value (Open Access Book)” and 7 specialised blogs, and 1 poster
Online publishing (online magazines, blogs, etc.)	Social Awareness	> 15 publications and four blog post per month	> 500 views / publications / year	<input type="checkbox"/> KPI achieved with 6 BDVA publications, 8 CORDIS publications, 3 Open Source blogs, 1 Atos booklet
Customisable marketing packages (videos, how-to demos, press kit etc.), suitable also for trade fairs	Raise awareness Attract users (supply or demand)	Production professional tailored to audiences of material specific	> 10 produced > 50 distributions	3 online marketing kits, 34 videos produced which got a total of +1.6k views,