BigDataStack

Holistic stack for big data applications and operations

A BigDataStack for Industry: a focus on retail, shipping and insurance.

Is Big Data the real future of emerging business?

EUROPEAN BIG DATA VALUE FORUM

bigdatastack.eu

After 3 years of research and innovation, the BigDataStack consortium proudly showcased its infrastructure management system for the holistic analysis of big data. On 5 November, the consortium organised its 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 event brought together 143 attendees from research and academia (50%), industry (33%), Big Data technology providers (15%) , interested in BigDataStack and the benefits of adopting big data solutions for industry.

Data is amongst the most valuable products nowadays, its exploitation for industry will twist the market over the next few years and while a lot of research and innovation is already taking place all over the world, a proper adoption plan to be shared among players is still missing. BigDataStack offers a perfect solution for Industry, handling real-time analytics on big data in flight and at rest, across multiple data stores and from several sources. Here we report back on the final event, the showcased solution and its implementation in Retail, Shipping and Insurance. And we share the insights from experts on Big Data being the real future of emerging business.

EUROPEAN BIG DATA VALUE FORUM

A BigDataStack for Industry - a focus on retail, shipping and insurance

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 architecture of the BigDataStack assets are illustrated in the figure below.

Figure 1: BigDataStack Services

During the event partners illustrated 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.

Smart Insurance

Maurizio Meglioni (GFT) presented an end-to-end overview of the Smart Insurance Use case, highlighting the added value of the solution for potential end-users. Insurance companies increasingly need IT data-based solutions to address their needs about the provision of services according to the customer "tailored" requirements. The BigDataStack allows insurance companies to better develop the customer management, by providing personalised services to both the customer, and the insurance company improving the handling of the customers' profitability. A multi-channel scenario has been developed by **GFT**, facilitating data analytics-powered smart insurance through a 360-degree view of the customer and personalized services. The data-oriented infrastructure of BigDataStack provides value in different areas of the scenario:

- **Customer segmentation:** all the customers are classified into groups by spotting coincidences in their attitude, preferences, behavior, or personal information. This grouping allows developing attitude and solutions especially relevant for the particular customers. As a result, target cross-selling and upselling strategies may be developed and personal services may be tailored for each particular segment (such as lower priced premiums)

- **Lifetime value prediction:** Customers lifetime value (CLV) is typically assessed via customer behavior data in order to predict the customer's profitability for the insurer. Thus, the behavior-based models will be applied to forecast the customer retention. This will allow forecasting which customers are likely to cancel contracts in the near future.

Connected Consumer

Bernat Quesada (Worldline) presented an end-to-end overview of the Connected Consumer Use case, highlighting the added value of the solution for potential end-users. This use case application provides retailers with optimal insights into consumer preferences and improves the effectiveness of marketing strategies for improving consumer shopping experience.
A major Spanish food retailer adopted BigDataStack to offer predictive shopping lists, and tailored recommendations and promotions.The added for the adopter are:

- **Data collection, aggregation, storage and analysis, handling a multitude of heterogeneous sources** which, combined, they generate data at an unprecedented rate, and BigDataStack will manage them and seamlessly analyse them for the 3 predictive services envisioned in the scenario.

- **Efficient and optimized analytics and real-time decision making** enabling the development of data-based value added services such as product logistics, virtual shopping carts and predictive lists, marketing and loyalty management. These services require a real time response, for example actuation of interactive displays in stores or issuing coupons to customers' mobile devices.

- **Process improvement** (with an emphasis on product replacement) exploiting the BigDataStack process modelling and process mining outcomes.

Real-Time Ship Management

Stathis Plitsos (Danaos) presented an end-to-end overview of the third Use case, highlighting the added value of the solution for Real-Time Ship Management. The BigDataStack algorithms optimise and help cut costs on maintenance and spare parts inventory planning and dynamic routing. These predictions are estimated and provided to **DANAOS**, a leading international maritime player with more than 60 container ships and will allow:

- **Performing predictive analytics** on top of both streaming and stored/historical data as key for the optimization of all processes.

- The underlying infrastructure system will allow for **larger datasets to be exploited towards more accurate predictions**, while the CEP approach over cross-streams and federated environments (given that different data are obtained by different sources) will enable the combination and consideration of additional aspects (e.g. inventory locations), which is not feasible today.

- Moreover, the overall maintenance process will be modelled through the Process Modelling framework and process mining techniques will provide insights regarding points of **optimization or potential bottlenecks**.

The three use cases were showcased, each telling their own end-to-end story using BigDataStack.
Watch the recordings of Part 1 of the event.

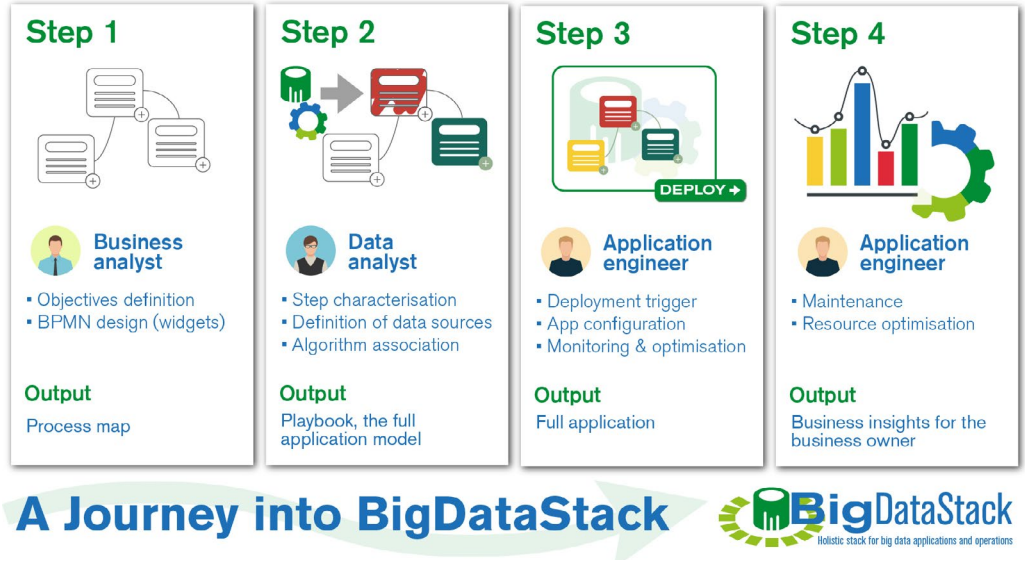


Figure 2: BigDataStack User Journey Smart Insurance

Is Big Data the real future of emerging business?

Celine Xu

Data scientist lead and she's been working in the field for more than 8 years now. She advocates applying analytics to generate action-orientated insights and scalable solutions to steer business decisions. Celine is passionate about leveraging the data and transforming AI capabilities / advanced analytical algorithms to end-to-end products to improve companies' business performance rather than only provide more information. With an excellent track record of analytics and machine learning, Celine has a wide range of industry experiences, e.g. retail, financial service. She has experience in consumer re-segmenting & re-targeting, Omni channel customer behaviour analysis, product portfolio optimization and next best action prediction applied models.

Ray Walshe

Assistant Professor at the Engineering and Computing Faculty at Dublin City University, Chair of BDVA Standards, Chair of Network Development, Gender Responsive Standards Initiative at the United Nations Economic Commission for Europe (UNECE). He is Chair of the External Advisory Board of StandICT.eu, H2020 Partner in StandICT2023.eu and Co-Lead on ELITE-S Fellowship Programme for Emerging Technologies Standardization (Elite-Fellowships.eu). Ray was appointed to the Working Group IEEE European Public Policy Committee on ICT in 2019 and is currently the AI WG Lead for IEEE EPPC.

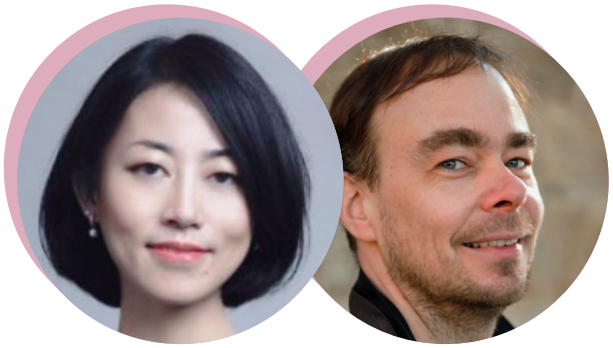
Stathis Plitsos

Senior researcher and Head of Development at DeepSea Technologies, a proud member of Danaos Corporation. He holds a PhD in Operations Research and Decision Support Systems. He has worked in many national and European research projects. Over the past 2 years, he focuses on IoT, big data optimization and AI approaches for the shipping industry.

Tatu Kuivalahti

long experience on strategic utilization of customer data for improving the customer experience. He is a serial entrepreneur and the co-founder of Custobar Ltd. which is a fast growing SaaS company based in Helsinki, Finland. Tatu Custobar has several hundreds of customers in more than 13 different countries in Europe and the US. Through this position, Tatu has a broad view on digital marketing and customer data.

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



Celine Xu

Data Exhaustion can refer to invaluable data or data that goes to waste. This has five possible reasons:

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

Solutions to avoid this can be:

- Efficiency storage of the data and provision of easy access,
- right data format and standardized format as the industry have,
- provide guidelines on how to standardize existing data and workers' education.

Tatu Kuivalahti

Remarked the importance of "visualising the data" and tools for end-users such as retailers, to visualise customers behaviour. Kuivalahti showed the Customer 360° view, an example of visualization of customer behaviour from Custobar. The example illustrated the power of visualisation tools in providing insights.

The BigdataStack software component SQL Data Skipping, provides a solution as it reduces the size of the data scanned for answering the SQL query. Thus it reduces the price to be paid by the customer and accelerates the query.

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

Celine Xu

Remarked that all analytics serve the same purpose: generate insights and actions. Specifically we:

1. Understand which problem to solve
2. The data needed for that
3. How to use that data to get actionable insights

Different data are used to get insights for different target customers. Based on the problem to solve, we would retrieve data from different data sources and accordingly use different models to predict. Other issues crucial for analytics are:

- The different definitions of the data across the databases
- The aggregation
- The limitation of the data

Stathis Plitsos

Agreed that it is crucial to clearly define the problem we want to solve with analytics and the importance of making the right questions and selecting the right data.

The BigdataStack software component Seamless provides a solution as it ensures applications keep up with their requirements in changing environments. The software is flexible, learns on its own and has a fast reaction.

Where does Europe sit in terms of its digital sovereignty towards Big Data? Where are investments still needed in your industry sector?

Ray Walshe

To have data sovereignty, we need to have control over our data, and for that purpose, the EC has developed the General Data Protection Regulation (GDPR) to protect citizens and ensure that personal identifiable data and information is stored and restored securely and safely.

Digitalisation has made that we nove have industry partners all over the world and data passes back and forth all the time. As Europeans, we want control over our data and make sure that data has been treated the same way as if we would like to have it treated within our jurisdiction. We need a standardised playing field, we need everybody to subscribe to best practices when it comes to data, we need data governance and sovereignty, we need to ensure that the international community works together to achieve standardisation, certification, regulation, legislation, concerning citizens data. GDPR is the starting point, but other EU initiatives are being put in place, such as the white paper, EU data strategy. Open Source (OS) standards are very important in this regard.

Regarding the need of investments in the industry, for the Shipping sector, Stathis Plitsos mentioned two pillars that should be addressed regarding investments:

- **Digitalization of the companies**, e.g. IoT solutions onboard, decision support systems, the integration of both tools and digital twins.

- **Decarbonization**: the International Maritime Organisation (IMO), states shipping has to reduce its CO2 emissions by 50%, by 2050. This includes an alternative fuse, components in the vessel to filter the gases, that must be monitored. The Shipping sector needs digitalization to address the challenge of decarbonization.

Tatu Kuivalahti

For the Retail sector, Tatu Kuivalahti stressed that the utilization of data in retail is in an early stage and not yet mature. Although e-commerce is growing at fast speed, the majority of shopping is done in physical stores, and that data is not much utilised, so there is room for improvement of the utilisation of data in this sector. There is also a lack of understanding of the big data possibilities for decision-makers and stakeholders in the retail sector.

Regarding investments, Kuivalahti thinks that it should start with education on how to leverage big data; educating people that are going to work in the retail industry in marketing or sales areas, and of course, executives. Additionally, Kuivalahti states there is also a lack of research projects between industry and the research community, and retail companies are not investing in research.

"BigDataStack is a good example of how retail companies are pushing big data to be more competitive and invest in research."

Tatu Kuivalahti (Custobar)

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

Ray Walshe

Within H2020, the EC is supporting IT standard developments and OS solutions and software and standardization is part of EC strategy. OS can contribute to the big data ecosystem, it provides a platform to develop more solutions and faster. Ray encourages companies to develop standards, because when Standards are published and adopted worldwide, it facilitates market access. This is true for all emerging technologies including Big Data.

Celine Xu

The importance of OS is increasing and can be seen from the users and suppliers sides. From the supplier side, we see big companies such as Microsoft, Apple, Facebook and IBM, actively participating in OS communities and even contributing to their OS solutions. Another big firm, Mackenzie also has an OS solution. From the users’ side and according to a RedHat report, 75% surveyed companies say that OS is quite important for their business and OS adoption acts as a major accelerator.

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

BigDataStack introduced the European open-source initiative towards increased impact of project outcomes. The European Open Source Initiative is led by RedHat and open to the complete research community, welcoming participation from EU-funded projects, industrial and academic institutions, and individual contributors.

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

Ray Walshe

"To break barriers among industry and research and encourage participation in R&D projects, we need to encourage collaboration and that is what EC has done in the H2020 and previous programmes. Billions of euros have been spent to get the best of the best from academia, industry, government, OS communities, etcetera to work together and collaborate."

Tatu Kuivalahti

"From the business perspective I think that companies are running too fast and don't look around for help in research or innovation and are too busy running the daily business and operations. There is not enough understanding of possibilities of utilization of research and it is needed to have a more active dialogue with the research community to explain which industrial problems companies are currently facing."

To enable data operations and data-intensive applications to fully exploit the sustainability of BigDataStack and take full advantage of the developed technologies, the consortium has brought on board three industrial use cases.

Is Big Data the real future of emerging business?

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

- Big Data will help a lot of emerging businesses
- The utilisation of a large amount of data is, definitely, the future for most industries, and it is needed to actively collaborate with different stakeholders, community specifically with the research
- Digitalisation is accelerating due to the Covid-19. Data is everywhere and we need to develop mechanisms to protect, standardize, govern properly, legislate, and certificate while protecting the rights of citizens. This is a challenge but also an opportunity to generate new business and new ways of generates revenues
- The important thing is to transform the data into actionable information. Big data is open our eyes to the options we didn't know we have

Find the slides here:
http://doi.org/10.5281/zenodo.4424956

Watch the recordings:
A BigDataStack for Industry:
a focus on retail, shipping and insurance
Is Big Data the real future of emerging business?

https://youtu.be/U02QZfaSSMk

This project has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement **No 779747**.