

## BDVA Software Architecture Challenges in Big Data – Workshop report

*“BDVA’s Software Architecture Challenges in Big Data workshop can clearly play a very interesting role in sharing technical details of current Big Data research and development projects.”*

*“This workshop .../... allowed a greater attention to the works presented and it was possible a direct and frank confrontation”*

*“I have felt like home attending this workshop”*

*“Nice diversity of works presented”*

*“Great organisation, nice programme and foremost fruitful discussion”*

The Big Data Value Association, with the support of the BDVe project, has organized the first workshop on “Software Architecture Challenges in Big Data”, as part of the European Conference of Software Architecture, held on 24-28 September 2018 at Universidad Politécnica de Madrid, Spain. The objective of this workshop was to explore the challenges that Big Data and other related technologies such as IoT, cloud computing, HPC and Artificial Intelligence pose to current and future software architectures.

The accepted papers were selected after a rigorous review process, in which each submitted paper was reviewed by at least three members of the workshop committee (see <http://bdva.eu/?q=node/1017>). There were finally accepted 9 papers, that ranged from real industrial applications in sectors such as transport and fishery, to more conceptual works regarding architectures and systems.

The workshop was opened by **Ana García Robles**, Secretary General of the **Big Data Value Association**, who explained how the Big Data Value Public Private Partnership (BDV PPP, from which the BDVA is the private counterpart) is contributing to foster the data driven innovation in Europe. Ana also welcomed this workshop as an initiative where the PPP stakeholders (and particularly the projects that are implementing the PPP strategy) could present relevant technical developments and results. In fact, 4 out of the 9 accepted papers were presented by the following projects in the scope of the Big Data Value PPP:

- **Transforming Transport** (<https://transformingtransport.eu/>), a lighthouse project involving 48 organizations and developing 13 pilots in 7 different domains in the fields of transport and logistics.
- **DataBio** (<https://www.databio.eu/en/>), another lighthouse project working on the domains of forest, agriculture and fishery, and that involves 48 partners from 17 countries.
- **EW-Shopp** (<https://www.ew-shopp.eu/>), a project that aims at supporting companies operating in the ecosystem of the eCommerce, Retail and Marketing industries, to increase their efficiency and competitiveness by leveraging deep customer insights.

All authors had the opportunity to present their works, that were grouped in the following three sessions: “Software architectures in genuine big data scenarios”, “Applying machine learning techniques to big data scenarios” and “Fostering business with big data through cloud analytics”.

### **Software architectures in genuine big data scenarios**

**Michele Ciavotta**, from the University of Milan-Bicocca, presented the work “[Data Wrangling at Scale](#)”, describing the architecture of a subsystem of the **EW-Shopp project** integrated platform aimed at facilitating and enhancing data transformation, integration, and extension processes, with focus on large scale integration of weather and event information.

**Javier Jerónimo**, from Universidad Politécnica de Madrid, presented the work “[Architectural challenges of the analysis of human behaviour in synthetic environments](#)”, where they propose Virtual Reality / Augmented Reality platforms with interaction models centred in the self-representation of the user, and where they also identify performance issues in out-of-lab environments.

**Dr. Ingo Simonis**, from the Open Geospatial Consortium, presented the work “[Container-based Architecture to Optimize the Integration of Microservices into Cloud-Based Data-Intensive Application Scenarios](#)”, where they propose container technologies, such as Docker, as the optimal approach between fully independent micro-services and services provided by a platform with configuration files and executable components.

#### **Applying machine learning to big data scenarios**

**Hamid Bouchachia**, from the Department of Computing and Informatics of the Bournemouth University, presented the work “[Scalable Online Learning for Flink: SOLMA library](#)”, driven by the lack of scalable online learning algorithms to cope with high-velocity streams, and more specifically for the big data platform Flink.

**Rafael Martínez Álvarez**, from the company Gradiant, presented the work “[Large Scale Anomaly Detection in Data Center Logs and Metrics](#)”, where they propose a data processing engine for anomaly detection both based in metrics and in log files, for large amounts of data and a continuous data flowing.

**David Scarlatti**, from Boeing Research and Technology, presented the work “[Real-time prediction of flight arrival times using surveillance information](#)”, explaining how the **Transforming Transport project** applies machine learning for the prediction of current flights and the integration of real-time visualization tools for customer usage.

#### **Fostering business with big data through cloud and analytics**

**Ståle Walderhaug**, from Sintef, presented the work “[Towards precision fishery: standardized modelling of smart services in the fishery domain](#)”, an approach from **DataBio project** on how smart data-driven services for fisheries can be identified and designed using a model-driven approach, using the BDVA big data reference model.

**Michele Ciavotta** presented in “[Using historical and weather data for marketing and category management in eCommerce](#)” the way used by the EW-Shopp project to address the problem to merge the data across consumer journey from various SMEs and to use Big Data analytics along with weather data to provide market insights to help them with marketing activities and category management.

**Mayank Mishra**, from **Tata Consultancy Services** presented the work “[Cracking the Monolith: Challenges in Data Transitioning to Cloud Native Architectures](#)”, where they outline how to overcome the challenges of refactoring legacy applications to micro-services, which will be autonomous and responsible for each single piece of the functionality.

The presentations were followed by a round-table discussion about the biggest architectural challenges of big data applications, and how architectures for data based applications will evolve in the future. Main comments were about the necessity to foster the connection between different architectures and platforms to compete with those from USA and China, and to be able to design platforms that could adapt to changing environments.

The survey conducted among participants after the workshop shows a high level of satisfaction regarding expectations, outcomes, interest and organization, and the majority of attendants expressed their intention of attending next edition.