



New DataCloud project to offer innovative solutions to Big Data

DataCloud develops a novel paradigm for Big Data pipelines

Oslo, Norway, 9th of February 2021

Several European organizations, from SMEs to Research Centers, are joining forces to develop new methods and tools for effective and efficient management of Big Data pipelines in the recently started project *DataCloud: Enabling the Big Data Pipeline Lifecycle on the Computing Continuum*. The initiative is funded by the European Commission under the Horizon 2020 Research and Innovation programme and will run for three years between 2021-2023.

With the recent developments in technologies such as the Internet of Things, massive amounts of data are being generated and often become Dark Data, i.e., data that are collected but not used and turned into value. Big Data pipelines are composite pipelines for processing data with non-trivial properties and characteristics, commonly referred to as the Vs of Big Data (e.g., volume, velocity, variety, veracity, value, etc.). They are essential for leveraging Dark Data, but tapping their potential requires going beyond the current approaches and frameworks for Big Data processing. At the same time, the Computing Continuum — federating Cloud services with emerging Edge and Fog computing paradigms — enables new opportunities for supporting Big Data pipelines, although challenges remain in the efficient management of heterogeneous and untrusted resources across the Computing Continuum.

The overall vision of the DataCloud project is the creation of a novel paradigm for Big Data pipeline processing over heterogeneous resources encompassing the Computing Continuum, covering the complete lifecycle of managing Big Data pipelines.

The main goal of the project is to develop a software toolbox, the DataCloud toolbox, comprising of new languages, methods, infrastructures, and software prototypes for discovering, simulating, deploying, and adapting Big Data pipelines on heterogeneous and untrusted resources.

“Situated at the intersection of Big Data processing, software engineering, and heterogeneous infrastructures (Cloud, Fog, Edge), DataCloud addresses the need to lower the technological entry barriers to the incorporation of Big Data pipelines in organizations’ business processes and make them accessible to a wider set of stakeholders regardless of the hardware infrastructure.” Dumitru Roman, SINTEF (DataCloud project coordinator)

DataCloud is coordinated by SINTEF Digital (Norway), one of Europe’s largest independent research organizations. The project partners include companies across Europe, such as Bosch (Germany), dedicated to engineering and electronics; Tellu (Norway), provider of welfare technology and digital healthcare solutions; iExec (France), with its decentralized platform using confidential computing and blockchain technology, pioneers in Cloud decentralization; Ubitech (Greece) that offers leading edge intelligent technical solutions; JOT Internet Media (Spain), a digital marketing company; MOG Technologies (Portugal), a worldwide digital media technology provider; and Ceramica Catalano (Italy), a leading company in the ceramics industry. They will have the full support of the research partners that, in addition to SINTEF, include Sapienza University of Rome (Italy), University of Klagenfurt (Austria), and Royal Institute of Technology (Sweden).

The results of the project will be demonstrated in a wide range of business cases including Smart Mobile Marketing Campaigns, Automatic Live Sports Content Annotation, Digital Health Systems, Predicting Deformations in Ceramics, and Analytics of Manufacturing Assets.

With a budget of 5M euros, DataCloud is funded by the European Commission under the Horizon 2020 Research and Innovation programme and spans from the period January 2021 – December 2023. The Kick-Off meeting was held between the 19th and 21st of January 2021, where the team defined joint strategies and started working together towards making the project results available to the general public as soon as possible.