ScienceMesh + JupyterLab: Collaborative Data Science services in scientific use cases and in business across different fields of study

Tuesday, 19 October 2021 15:10 (15 minutes)

Data Science became famous in main stream about a decade ago: after Harvard Business Review article which coined Data Scientist as "The Sexiest Job of the 21st Century". In business it is defined as "using data to increase competitive advantage".

ScienceMesh, developed in CS3MESH4EOSC project, creates the Federated Scientific Mesh providing federated sharing of data across different sync-and-share services, federated use of applications (such as collaborative document editing, data archiving, and data publishing), fast transfer of large datasets and remote data analysis (distributed Data Science environments).

As all scientific disciplines nowadays are based on data analysis and Distributed Data Science Environments can support research in all fields of study. It will also support Data Science in the business context in all sectors. In recent report **Critical Capabilities for Data Science and Machine Learning Platforms** (March 2021) Gartner predicts, that in near future **collective intelligence in Data Science** and **cloud-based AI infrastructure** will be among key factors for competitive advantage.

For ScienceMesh Distributed Data Science Environments, **along with** Software Mind (part of Ailleron group, a global IT service provider based in Poland, delivering skilful managed teams for even most demanding projects), **it was developed the JupyterLab extension**, integrating with ScienceMesh –file browsing and additional share and collaboration functionalities for notebooks and resources across federated cloud are now possible in JupyterLab environment. Collaborative Data Science is being now used in products from Finance, IoT, Earth Observation, SmartCities and Pharma, but it is present in virtually every business.

Jupyter Notebook has become No1 platform used by data scientists to build interactive applications and to work with big data and AI. It is a free, open-source, interactive web-based tool which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document. Jupyter has exploded in popularity over the past couple of years, with an enthusiastic community of user–developers.

In this talk, the relevance and benefits of ScienceMesh Distributed Data Science Environments will be presented, starting from **two scientific use cases** (High Energy Physics and Earth Observation), along with various **business-related scenarios**.

In CS3MESH4EOSC project, SoftwareMind is supporting science innovation by providing the expertise on microservices architecture, integration, DevOps, agile software development process and Data Science. It leads tasks on Reference interoperability platform and distributed Data Science environments. In this talk, the speaker will show also how this is as a part of the strategy of growing the usage of application services in the cloud, microservice-based architectures, Data Science, Big Data integration and analytics.

Speaker bio:

Experienced System Architect and R&D Project Manager with 20+ years of enterprise software design and development. Founder and leader of Big Data Lab, focused on R&D, Data Science, data driven innovation and high quality agile software development. He is now involved in CS3MESH4EOSC project (https://cs3mesh4eosc.eu/), leading tasks on Reference cloud interoperability platform and distributed Data Science environments (the project creates Science Mesh integrated with EOSC: https://cs3mesh4eosc.eu/index.php/science-mesh). He developed Big Data solutions before it became mainstream. In the years 2005-2008 he was involved in development of technology for the first web-scale Semantic Web startup: garlik.com - from R&D in alpha phase to commercial launch browsing 4billion web pages. Software Mind's team he was a part of - and finally he

lead - started using Hadoop in February 2006, as one of the first companies in the world. He participated in and lead many commercial projects which included Big Data, high volume and high velocity solutions, in various sectors: telco operators, international telco interoperability hubs, banks and financial institutions or Content Delivery Network providers. He was a Work Package leader and provided Big Data architecture in in a number of EU-funded research projects. He was Chief Software Engineer and integration WP leader in ROBUST (2010-2013). He was a WP leader on technological infrastructure in EU IP WeSenseIt (2012-2016), responsible for designing and implementing a scalable architecture of sensor/IoT platform, management of huge scale geospatial data, scalable backend for mobile apps and cloud-based infrastructure. In Seta (2016-2019) he leads a Work Package on creating Big Data infrastructure for organizing, monitoring and planning multimodal mobility in large metropolitan areas. His current focus in Big Data innovation is state of the art solution for management of geo-located data and low-latency services, including GPU-based acceleration of geospatial indexes. He has experience in all phases of project lifecycle including requirements gathering and analysis, architectural analysis and design, data modelling, implementation, deployment, coordinating and mentoring.

Most suitable track

Innovating services together

By submitting my abstract, I agree that my personal data is being stored in accordance to conference Privacy Policy

Primary author: SIEPRAWSKI, Marcin (Software Mind)

Presenter: SIEPRAWSKI, Marcin (Software Mind)

Session Classification: Innovating Services Together - Presentations