

# Solutions for Cloud/HPC and big data frameworks interoperability

Thursday, 30 November 2017 16:45 (15 minutes)

Forthcoming EO and scientific space missions create unprecedented opportunities to empower new types of user applications, and to develop a new generation of user services, particularly for the European scientific community. The associated generated data are steadily increasing volume, delivery rate, degree of variety, complexity and interconnection of data. Challenges stemming from such an increase of volume, velocity and variety of data drive the urgent need of new processing concepts which shall ensure the necessary power but also scalability and elasticity to actually exploit those data.

Furthermore, new requirements from the scientific communities are emerging. In particular the requirement to easily integrate their own processing, manipulation and analysis tools into harmonized frameworks (platforms), which on their side should provide basic processing features, like e.g. efficient data access, distributed massive processing, load balancing, etc. Platform users aim at integrating their own processing tools in a seamless and easy way, avoiding software changes and/or development of additional interfaces/components for the sole purpose of their integration and deployment.

These challenges can be faced to with the help of cloud computing, the development of many new computing patterns under the banner of Big Data technologies and a variety of libraries and toolboxes for processing currently available users to support and simplify their processing needs. However, there is no unique Big Data or HPC Framework able to address all computing patterns (Map/Reduce, Streaming, Directed Acyclic Graph...) and all data types (Satellite Imagery, IOT data, Social network stream...). That is why modern scientific computing platforms should be able to combine efficiently Big Data and legacy computing patterns on hybrid on-premise/cloud computing infrastructures.

This presentation will describe the solutions proposed by CS to build such a processing platform.

These solutions are based on a multi-cloud strategy that allows to always have the right offer, to benefit from a maximum of flexibility and to ensure independency against cloud vendors. For this purpose CS developed CS ViP (Critical System Virtual Platform), a multi IaaS system for interoperability with most of popular cloud providers through an unified API. CS ViP uses cutting-edge devops, monitoring and remote desktop technologies.

On top of it, the profusion of Big Data frameworks can be used. Unfortunately, they not interoperable and the choice of one of them is dividing with regards to the ecosystems of the others. In the same way, it is really difficult to access the large valuable code database targeting traditional HPC from the chosen framework.

To ensure interoperability between these frameworks, CS designed SCREW, a PaaS system providing on demand computing platforms that combine major Big Data Frameworks - Spark, Hadoop, Ignite... - with traditional HPC framework - MPI and batch scheduling using DRMAA standard.

A precursor to the future Copernicus DIAS platforms, RUS - Research and User Support Service - <https://rus-copernicus.eu/>, is already running on top of these open source technologies. RUS is a good example of the provisioning of a federated service for research, enabling interoperability between different cloud providers.

**Primary author:** Mr DORGAN, Sébastien (CS Communication & Systèmes)

**Co-author:** Mr D'HOINE, Sylvain (CS Communication & Systèmes)

**Presenter:** Mr D'HOINE, Sylvain (CS Communication & Systèmes)

**Session Classification:** Special focus on Earth Observation