

Orchestrated satellite data management

Wednesday, September 25, 2019 11:45 AM (15 minutes)

With the latest missions launched by ESA or NASA, such as Sentinel or Landsat, equipped with the latest technologies in multispectral sensors, we face an unprecedented amount of satellite data never reached before. Exploring the potential of this data with state-of-the-art Artificial Intelligence techniques such as “Deep Learning” could potentially change the way we understand the Earth system and how to protect its resources.

The eXtreme-DataCloud project (XDC), under the umbrella of the H2020 programme, aims at developing a scalable environment for data management and computing, addressing the problems of the growing data volume and focused in providing a complete framework for research communities through the European Open Science Cloud. The target of this project is to integrate different services and tools based on Cloud Computing to manage Big Data sources, and Use Cases from diverse disciplines are represented. One of the goals of the project is to deal with extremely large datasets, including diverse data and metadata types, formats and standards that enable the automatic integration of Big Data.

In order to interoperate those big data sources, the XDC LifeWatch ERIC Use Case proposes a Virtual Research Environment (VRE) deployed on the Cloud that allow the users to preprocess the satellite data to obtain valuable information about the water quality of lakes and reservoirs without the need of using local resources as well as hiding the complexity behind. The architecture of this virtual environment consists of different Docker containers that run automatically with a common distributed storage system (Onedata) capable of storing the data with associated metadata that facilitate the discovery. The workflow of the VRE to preprocess the satellite data is managed by the INDIGO PaaS Orchestrator.

This presentation will describe the architectural design of the VRE and the different components (Jupyter interface, docker deployment for data preprocessing, modelling, etc.) as well as details on how this cloud-based approach can be adopted to many other cases.

Primary authors: Mr GARCIA DIAZ, Daniel (IFCA); Dr AGUILAR GÓMEZ, Fernando (IFCA)

Presenter: Mr GARCIA DIAZ, Daniel (IFCA)

Session Classification: IBERGRID Contributions

Track Classification: Enabling Research Applications in advanced Digital Infrastructures