Orchestrated satellite data management
IBERGRID 2019
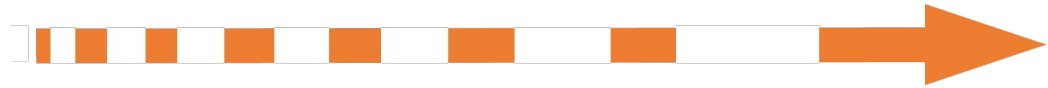
Water Quality Forecast System
Daniel García & Fernando Aguilar
garciad@ifca.unican.es
aguilarf@ifca.unican.es

Data Management for extreme scale computing

Problem

With the **latest missions** launched by ESA and 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.
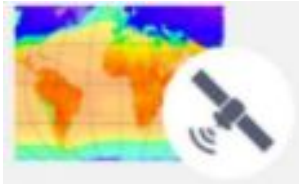
- **Sentinel 2:**
  - 1 file per region (100x100 km2)
  - 1 file every 5 days
  - ~ 2GB per file
  - 13 Bands per file
  - 10, 20 or 60 m of Spatial Resolution

- **Landsat 8:**
  - 1 file per region (170x170 km2)
  - 1 file every 13 days
  - ~ 2GB per file
  - 11 Bands per file
  - 15 and 30 m of Spatial Resolution

Special software needed to analyze the satellite data!!

Our Approach

Our approach: a **Virtual Research Environment (VRE) deployed on the Cloud**. 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 manage by the INDIGO PaaS Orchestrator.
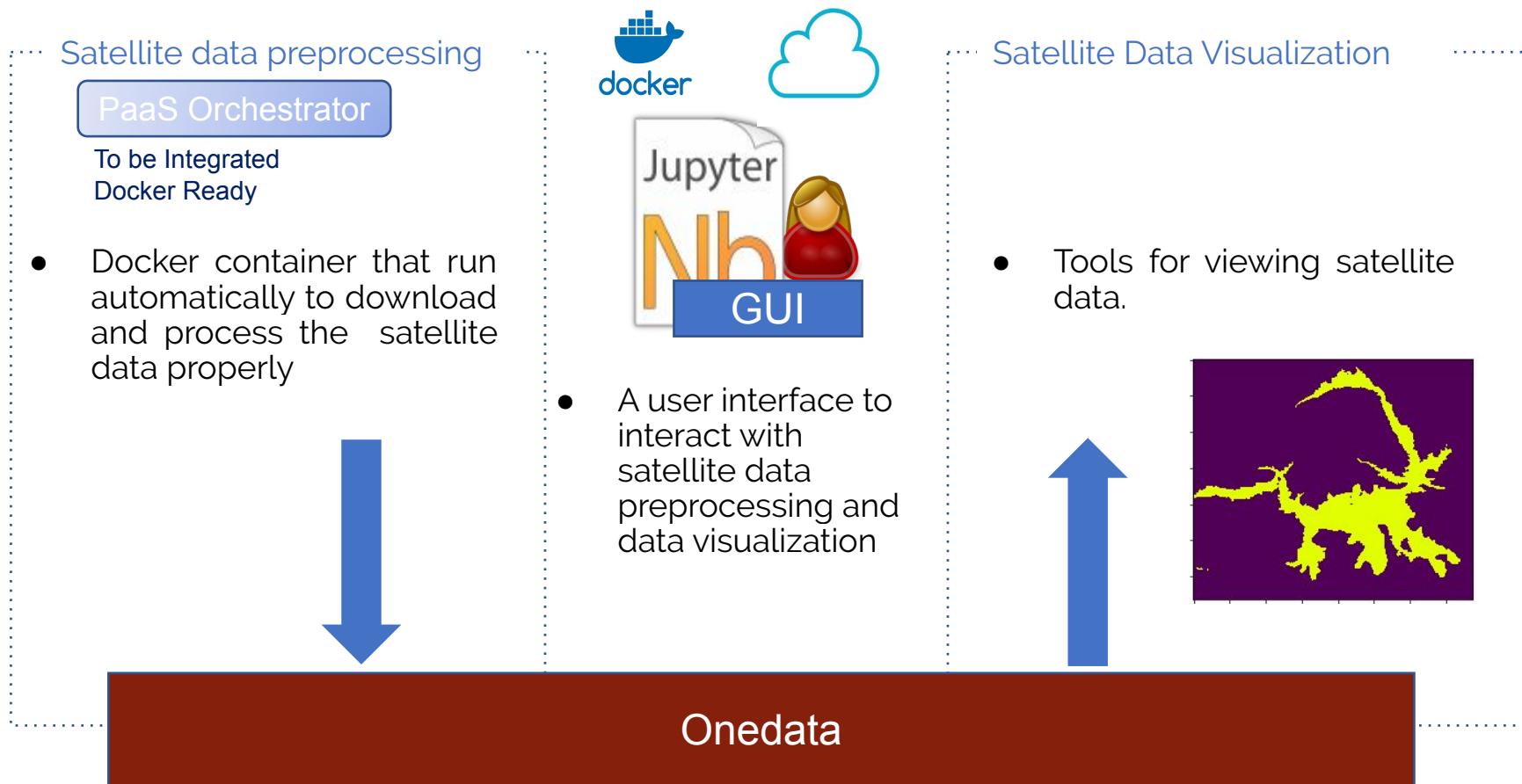
- XDC IAM as a AAI System.
- Jupyter Notebook as a User Interface.
- Different Docker containers that run automatically to process satellite data.
- A common distributed storage system (Onedata).
- The workflow is managed by the INDIGO PaaS Orchestrator.

Architecture

Satellite data preprocessing

PaaS Orchestrator

To be Integrated
Docker Ready

- Docker container that run automatically to download and process the satellite data properly

Jupyter Nb

GUI

- A user interface to interact with satellite data preprocessing and data visualization

Satellite Data Visualization

- Tools for viewing satellite data.

Onedata

Workflow

1- Deploy the Docker image through the log in JupyterHub with the XDC IAM

2- Show the main notebook as a user interface to manage the satellite API

3- Download data by selecting the parameters (range of the date and the place)

4- Saved the data in Onedata with Onedata attachment

5- Plot graphs of the data downloaded.

Demo

DEMO

eXtreme DataCloud
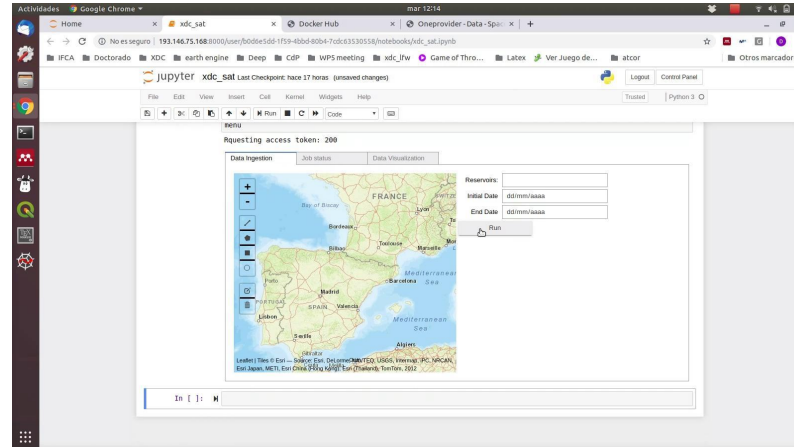
Conclusions

- **Cloud environment** to deal with Big Data sources
  - **Ingestion, analysis and visualization: All in Cloud**
  - AAI with IAM in all the steps: **Integrated environment.**

- **FAIR data management**:
  - Automatic metadata attachment
  - Metadata Discovery to find model input data.

Thanks for listening!!