



Contribution ID: 24

Type: not specified

From interactive to distributed computing of land parcel signatures using HTCondor

Wednesday, 25 September 2019 09:35 (25 minutes)

In the framework of the Common Agricultural Policy (CAP) of the European Union, a big technological shift is happening. For decades, the correct payment of subsidies to farmers was controlled by means of remotely sensed images, by doing visual interpretation and field visits, to assess that a randomly selected percentage of the land parcels respected all the rules. In recent years, we are witnessing a big evolution in Earth Observation, with the Copernicus Programme and the Sentinel satellites that provide coverage of every piece of land in high resolution every few days, together with the availability of cloud platforms capable of storing the big amount of data captured. Coupling this with the ease of use of cloud computing platforms and a wide number of tools for extracting valuable information from big geospatial datasets using machine learning techniques, puts the CAP controls sector in the edge of a revolution: from 2021, each single agricultural parcel in Europe will be constantly monitored for the full year. This involves the calculation and the constant update of temporal profiles that will monitor the vegetation status of land parcels in all their phases, from ploughing to sowing, from ripening to harvesting.

In this context, the JEODPP (Joint Research Centre Big Data Platform) group was involved in preliminary studies to assess the feasibility of the new CAP Monitoring. On a first stage, by accessing the full catalogue of Sentinel-2 images, we developed an interactive tool to calculate the Normalized Difference Vegetation Index profile for a single parcel at a time inside a JupyterLab notebook (s2explorer application).

When the algorithm was tested and verified, the need for scaling to regional or national level arose. This implied the need to process millions of vector polygons, each of them covered by more than 50 images per year, a perfect workspace for using the HTCondor workload manager services already available inside the JEODPP platform. The C++ routines developed for the interactive prototype were compiled in a standalone executable and the calculation was divided in three phases: 1) compilation of the list of satellite images involved in the selected spatial and temporal range, 2) creation of a job for each image, 3) collection of the result into a single binary file; a typical map-reduce schema. All these phases were performed by using HTCondor jobs, and, in particular, the second phase was heavy parallelized on the hundreds of cores available inside the JEODPP platform.

We executed a first test on a region in Hungary (10K parcels for a full year processed in less than half an hour) and then scaled to the full Catalonia (640K parcels processed in 4 hours).

The need to deeply evaluate the results of the batch processing generated the idea to “close the circle”, that is to provide an interactive tool to visually assess the calculations made by the HTCondor jobs. We developed a Python application running inside JupyterLab that could visualize all the land parcels involved in the calculation and, by clicking on each of them, immediately display the vegetation profile and the imageries extracted from each individual satellite acquisition date. The tool is widely used by the JRC D.5 unit and it is the base for the future characterization of the crops by means of machine learning algorithms, a key component of the CAP Monitoring.

This use case is an example of the involvement of HTCondor services in a complex environment where the need for interactive prototyping goes along with heavy distributed processing needs and contributes to create an integrated solution.

Desired slot length

Speaker release

No

Primary authors: Mr DE MARCHI, Davide (European Commission, Joint Research Centre (JRC) Directorate I. Competences. Unit I.3 Text and Data Mining); Mr WIRNHARDT, Csaba (European Commission, Joint Research Centre (JRC) Directorate D. Sustainable Resources. Unit D.5 Food Security); Mr SOILLE, Pierre (European Commission, Joint Research Centre (JRC) Directorate I. Competences. Unit I.3 Text and Data Mining)

Presenter: Mr WIRNHARDT, Csaba (European Commission, Joint Research Centre (JRC) Directorate D. Sustainable Resources. Unit D.5 Food Security)

Session Classification: Workshop presentations

Track Classification: HTCCondor presentations and tutorials