Contribution ID: **156** Type: **Presentation**

JupyterLab for Earth Observation applications with HTCondor scaling and Voilà dashboarding

Tuesday 26 January 2021 10:00 (15 minutes)

The Joint Research Centre (JRC) of the European Commission has set up the JRC Big Data Analytics Platform (BDAP) as a multi-petabyte scale infrastructure to enable EC researchers to process and analyse big geospatial data in support to EU policy needs [1]. One of the service layer of the platform is the JEO-lab environment [2] that is based on Jupyter notebooks and the Python programming language to enable exploratory visualization and interactive analysis of big geospatial datasets. JEO-lab is set-up with deferred processing, using multiple service nodes to execute the Jupyter client processing workflow starting from data stored in the CERN EOS distributed file system deployed on the BDAP. In this context, recent developments were done in these areas:

• Scaling to HTCondor: batch processing submission from the notebook
BDAP uses HTCondor [3] as the scheduler for the batch processing activities. Users can submit complex tasks to the HTCondor master that will then allocate jobs to the HPC nodes and control their progressing. The submit activity is generally done from a remote desktop environment which consists of a linux instance accessible from the browser. For some specific tasks, in particular those involving geospatial datasets, the direct submit of simple processing tasks from the interactive notebook environment has been developed, calling the python bindings of HTCondor. This allows for easy check of the results, thanks to a new GUI created in Jupyter to control the status of the processing jobs and to instantly visualize the produced datasets on an interactive map. This new development complements the usual deferred mode of the JEO-lab environment, which manages big geospatial datasets by processing them at the screen zoom level and at the visible viewport, with a tool that can scale the prototyped processing chains to operate at the full resolution of the input datasets, in order to create and permanently save any type of derived product. The processing is based on python: the user can create a custom python function that is applied to all the input image files to generate the output product by using any Numpy, Scipy, gdal, pyjeo [4] function. Examples will be demonstrated for the numerical evaluation of the effect of the lockdown measures on air quality at European level using the data coming from the Sentinel-5P satellite of the EU Copernicus programme.

• Web dashboards derived from Jupyter notebooks using Voilà
Voilà [5] turns Jupyter notebooks into standalone web-dashboard applications; it supports Jupyter interactive widgets like ipywidgets [6], charting libraries like plotly [7], etc., while not permitting arbitrary code execution, thus posing less security threats. Many applications developed inside the JEO-lab environment have been brought into the Voilà world, where they are accessible without the need for user authentication, and thus greatly expanding the impact of the BDAP platform and providing an easy way to publish complex interactive visualization environments. Among the most important, the CollectionExplorer dashboard that enables users to browse all the geospatial datasets available in the platform with an easy-to-use interface, dedicated dashboards like S2explorer and DEMexplorer, to perform typical GIS operations on Sentinel2 products and Digital Elevation Models, and many dashboards for the monitoring of the Covid-19 spread in various regions and continents.

Both developments were partially financed by the H2020 project CS3MESH4EOSC, led by CERN and to which JRC participates providing support in the Earth Observation use case. In this context, the CERN SWAN Jupyter environment was also deployed in view of future full adoption of the IOP and the connection to the ScienceMesh federated infrastructure.

The JRC Big Data Analytics Platform is a living demonstration of a complex ecosystem of cloud applications and services that allows data scientists'navigation inside a multi-petabyte scale world. In particular, the exploratory visualization and interactive analysis tools in the JEO-lab component can run custom code to pro-

totype the generation of scientific evidence as well as create GUI applications that can be used by end-users ranging from policy makers to citizens.

[1] P. Soille, A. Burger, D. De Marchi, P. Kempeneers, D. Rodriguez, V.Syrris, and V. Vasilev. "A Versatile Data-Intensive Computing Platform for Information Retrieval from Big Geospatial Data". Future Generation Computer Systems 81.4 (Apr. 2018), pp. 30-40.
https://doi.org/10.1016/j.future.2017.11.007.

[2] D. De Marchi, A. Burger, P. Kempeneers, and P. Soille. "Interactive visualisation and analysis of geospatial data with Jupyter". In: Proc. of the BiDS'17. 2017, pp. 71-74.
https://zenodo.org/record/3248741#.XeDvSuhKg2w.

[3] https://research.cs.wisc.edu/htcondor/

[4] P. Kempeneers , O. Pesek, D. De Marchi, P. Soille. "pyjeo: A Python Package for the Analysis of Geospatial Data", ISPRS International Journal of Geo-Information, Volume 8, Issue 10, October 2019. Special Issue "Open Science in the Geospatial Domain".
https://doi.org/10.3390/ijgi8100461

[5] https://blog.jupyter.org/and-voil%C3%A0-f6a2c08a4a93
https://github.com/voila-dashboards/voila

[6] https://ipywidgets.readthedocs.io/en/latest/

[7] https://plotly.com/

Authors: Mr DE MARCHI, Davide (European Commission); Mr BURGER, Armin (European Commission); Mr SOILLE, Pierre (European Commission)

Presenter: Mr DE MARCHI, Davide (European Commission)

Session Classification: Novel Data Science Environments

Track Classification: Main session: User Voice: Novel Applications, Data Science Environments & Open Data