Contribution ID: **34** Type: **not specified**

# Dynamic Batch service with HTCondor and Kubernetes

*Thursday 26 September 2019 09:00 (25 minutes)*

One of the main challenges the JRC Big Data Platform-JEODPP [1] is to offer well consolidated computational services, such as the batch system or interactive data visualization on which users can process large scale geospatial data while ensuring a smooth user experience combined with easy administration of all resources from hardware to applications.
Due to the heterogeneity of the user requirements of the JEODPP platform, many services have dynamic demand over time. As an example users who want to visualise data interactively requires a considerable amount of resources just in short periods, mainly during the core time of the day. Considering that the batch service should use the full capacity whenever possible, the resources should be allocated dynamically. Moreover keeping a physical separation between htcondor, interactive nodes, and all our other services as in our previous setting is not satisfactory from service management, monitoring, and performance perspectives. In that fixed setting, a lot of idle time when some operation was ended was added up to a lack of resources in the peak of need. To address these issues, we are moving to implement all our services under the Kubernetes control.
Running batch workloads in Kubernetes is possible [2] and Kubernetes has a specific resource for that purpose (Job) but it is quite limited. To address this limitation, we present in this work the implementation of HTCondor within Kubernetes. HTCondor daemons are packaged inside pre-configured Docker images and deployed as a service in a container cluster handled by Kubernetes. As a proof of concept, we have successfully run an actual JEODPP use-case (atmospheric correction with Sen2Cor [3]) on Kubernetes and compared its performance with a standard HTCondor pool. Finally, we present some considerations to implement other services based on the namespaces of Kubernetes and some constraints to implement this solution in a production environment such as the JEODPP.

**References**

[1] P. Soille, A. Burger, D. Rodriguez, V. Syrris, and V. Vasilev.; *Towards a JRC earth observation data and processing platform* Proc. of the 2016 Conference on Big Data from Space (BiDS'16), pages 65-68, 2016. Available from doi: [10.2788/854791][4]

[2] D. Aiftimiei, T. Boccali, M. Panella, A. Italiano, G. Donvito, D. Michelotto, M. Caballer et al. *Geographically distributed Batch System as a Service: the INDIGO-DataCloud approach exploiting HTCondor.* In J. Phys. Conf. Ser., vol. 898, p. 052033. 2017.

[3] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm and F. Gascon, *Sen2Cor for Sentinel-2.* In Image and Signal Processing for Remote Sensing XXIII, vol. 10427, p. 1042704. International Society for Optics and Photonics, 2017.

———————————-

## Desired slot length

## Speaker release

Yes

**Authors:** RODRIGUEZ, Dario (European Commission - DG JRC); Mr MARLETTA, Luca (European Commission, Joint Research Centre (JRC) ); Mr SOILLE, Pierre (European Commission, Joint Research Centre (JRC) Directorate I. Competences. Unit I.3 Text and Data Mining)

**Presenter:** Mr MARLETTA, Luca (European Commission, Joint Research Centre (JRC) )

**Session Classification:** Workshop presentations

**Track Classification:** HTCondor presentations and tutorials