

ENES climate data infrastructure

Monday, 2 November 2020 11:15 (15 minutes)

In the last decade, the continuous increase in the volume of scientific data has forced a shift in the data analysis approach, leading to the development of a multitude of data analytics platforms and systems capable of handling this data deluge. All these innovations have propelled the community towards the definition of novel virtual environments for efficiently dealing with complex scientific experiments, while abstracting from the underlying infrastructure complexity.

In this context, the ENES Climate Analytics Service (ECAS) aims to enable scientists to perform data analysis experiments over large multi-dimensional data volumes providing a workflow-oriented, PID-supported, server-side and distributed computing approach. Two instances of ECAS are respectively running at CMCC and DKRZ in the scope of the European Open Science Cloud (EOSC) platform, under the EU H2020 EOSC-Hub project. ECAS builds on top of the Ophidia High Performance Data Analytics framework, which has been integrated with AAI solutions (e.g. EGI Check-in, IAM), data access and sharing services (e.g., EGI DataHub, EUDAT B2DROP/B2SHARE), along with the EGI federated cloud infrastructure.

The ECASLab virtual environment, based upon ECAS and the JupyterHub service, aims to provide a user-friendly data analytics environment to support scientists in their daily research activities, particularly in the climate change domain, by integrating analysis tools with scientific datasets (e.g., from the ESGF data archive) and computing resources (i.e., Cloud and HPC-based).

ECAS is one of the platform configurations made available to users from the EGI Applications on Demand (AoD) service. Thanks to the integration into the Elastic Cloud Compute Cluster (EC3) platform, operated by UPV, researchers can very easily deploy and configure a full ECAS environment on the EGI FedCloud. The EC3 service not only takes care of managing the setup and contextualization of the entire ECAS cluster, but also manages the elasticity of the environment by scaling up/down the cluster size on the cloud resources based on the current workload. This integration will effectively support scientists and help advance their research by exploiting a custom ready-to-use environment without the burden of the platform setup. With respect to security and data access, a stronger integration with EGI services will be part of the future work to provide an even smoother experience to ECAS users.

This talk will present the ECAS environment and the integration activities performed in the context of EOSC-Hub, with a special focus on the integration with the EGI federated cloud infrastructure.

Primary authors: Mr ELIA, Donatello (Euro-Mediterranean Center on Climate Change (CMCC) Foundation and University of Salento); Mr ANTONIO, Fabrizio (Euro-Mediterranean Center on Climate Change (CMCC) Foundation); Dr PALAZZO, Cosimo (Euro-Mediterranean Center on Climate Change (CMCC) Foundation); Dr CABALLER, Miguel (Polytechnic University of Valencia (UPV)); Dr WEIGEL, Tobias (Deutsches Klimarechenzentrum (DKRZ)); Dr KINDERMANN, Stephan (Deutsches Klimarechenzentrum (DKRZ)); Prof. BLANQUER, Ignacio (Polytechnic University of Valencia (UPV)); Prof. FIORE, Sandro (University of Trento); Prof. ALOISIO, Giovanni (Euro-Mediterranean Center on Climate Change (CMCC) Foundation and University of Salento)

Presenter: Dr KINDERMANN, Stephan (Deutsches Klimarechenzentrum (DKRZ))

Session Classification: Experience from Around the World with Federated Clouds and Data: Lessons for a European Cloud Federation