

CMCC Analytics-Hub

CMCC Analytics-Hub	1
Introduction	1
Registration	1
Data pool content	1
Data pool access	2
Access via Jupyter hub	2
Technical information about the CMCC Analytics-Hub	2
User support	2

Introduction

The [CMCC Analytics-Hub](#) facility provides a data science environment with (i) computational and storage resources, (ii) tools, libraries and services as well as (iii) a set of collections of climate model data in the context of climate model intercomparison experiments. This environment is hosted as part of the CMCC data infrastructure and aims to support user groups with respect to climate data collection access, processing and analysis. The hosted datasets concentrates on model data generated as part of the CMIP climate model intercomparison project. This document summarizes the initial service offerings around this data pool and will be regularly updated to reflect the evolution of these services.

Registration

- To access the CMCC Analytics-Hub, the user needs to be registered at CMCC
 - Registration can be done at <https://ophidialab.cmcc.it/is-enes/registration.php>
- After registration users are enabled to use a set of services (e.g. [Jupyter](#), [Ophidia](#)), data collections as well as a comprehensive set of Python scientific modules for data analysis.
- Registered users can contact analytics-hub-support@cmcc.it for any information request.

Data pool content

CMCC will provide access to a set of specific CMIP variable-centric collections. Data will be downloaded and kept in sync with the ESGF federated data archive using the synda replication tool. About 50TB disk space have been allocated to this purpose. More information about the hosted service and data is available at: <https://ophidialab.cmcc.it/is-enes/>

Data pool access

The data pool is efficiently accessible from cluster resources as well as JupyterLab. The JupyterLab environment is already equipped with a set of Python libraries to support end-users data analysis. Users can request the installation of additional libraries by contacting the user support at: **analytics-hub-support@cmcc.it**

- users can login into the front end node of the CMCC Analytics-Hub and run interactive analytics operators;
- users can login into the JupyterLab user interface and run interactive scripts.

Compute intensive parallel data analysis is supported by the submission of batch compute jobs via Ophidia to the CMCC Analytics-Hub cluster. The pool data is accessible from all compute nodes.

Access via Jupyter hub

CMCC currently provides a pre-production Jupyter hub service for interested users at:

<https://ophidialab.cmcc.it/jupyter/>

The notebooks are running in the CMCC data infrastructure - thus the CMIP data pool is directly accessible as described above.

Technical information about the CMCC Analytics-Hub

Compute resources, storage and installed software:

- Analytics-Hub cluster with 5 compute nodes (20 cores/node, 256 GB RAM/node, 1 TB local disk each), SLURM resource manager, Network 10 Gbit/sec (10 GbE);
- 50 TB shared storage for data analysis (GFS);
- One Virtual Machine (4 cores, 8 GB memory, 12 GB disk) for client-side components
- One Virtual Machine (8 cores, 16 GB memory, 12 GB disk) for front-end components and services;
- Software: the cluster runs the full Ophidia framework stack. Compute nodes are also equipped with additional libraries and tools (e.g. CDO, NCO)
- Data: the shared disk space hosts data collections from CMIP5/CMIP6 which are being populated/updated over time from ESGF using the [synda replication tool](#) (recommended as it supports consistent version updates to keep the copy in sync with the pool content).

User support

- For any request users can send an email to: **analytics-hub-support@cmcc.it**