Contribution ID: **41** Type: **Poster**

Object storage for climate data storage and analytics

Data analysis in climate has been traditionally done in two different environments, local workstations and HPC infrastructures. Local workstations provide a non scalable environment in which data analysis is restricted to small datasets that are previously downloaded. On the other hand, HPC infrastructures provide high computation capabilities by making use of parallel file systems and libraries that allow to scale data analysis.

Parallel file systems found in HPC show scalability limitations due to constraints of the POSIX standard, which favors consistency of files while penalizing scalability. Object storage consists of a new storage system that tries to favor scalability instead of consistency and is usually provided by commercial cloud storage providers, such as Amazon S3 or Google Cloud Storage.

NetCDF, the standard library for climate data, actually requires files to be stored in a file system, although work is currently being carried out to allow storage on object stores. In this work we show how these new object storage systems can be combined with Python libraries, such as xarray and Dask for distributed analysis and Zarr for data storage in object stores, that allow computations to be easily parallelized without scalability restrictions.

Primary authors: Mr CIMADEVILLA ALVAREZ, Ezequiel (Santander Meteorology Group (unican)); Ms PALACIO, Aida (IFCA); Prof. COFIÑO, Antonio (Santander Meteo Group (UNICAN))

Presenter: Mr CIMADEVILLA ALVAREZ, Ezequiel (Santander Meteorology Group (unican))

Session Classification: IBERGRID Contributions

Track Classification: R&D for computing services, networking, and data-driven science.