

Computing services for climate data analysis

Thursday, September 28, 2023 12:10 PM (20 minutes)

Climate data analysis often entails downloading datasets of several terabytes in size from various sources and employing local workstations or HPC computing infrastructures for analysis. However, this approach becomes inefficient in the era of big data due to the considerable expenses linked with transferring substantial volumes of raw data over the Internet from diverse sources, encompassing observations and model simulations. Climate data analysis tasks involve routine procedures like subsetting, regridding, and bias adjustment. These processes can be effectively executed using existing packages that adhere to best practices, thereby curtailing redundancy. Recent strides in web-based computing frameworks and cloud computing have emerged as feasible alternatives, furnishing collaborative computing infrastructures that improve code reproducibility and reusability. Cloud systems are frequently established on top of object storage, accompanied by the development of pioneering data formats and libraries that harness the potential of this innovative storage paradigm. Consequently, web-based virtual research environments have arisen, grounded in cloud infrastructures. These infrastructures not only refine data analysis workflows but also improve overall productivity. This work provides an overview of these research environments.

Primary author: CIMADEVILLA, Ezequiel

Presenter: CIMADEVILLA, Ezequiel

Session Classification: IBERGRID

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