Geophysical Research Abstracts
Vol. 20, EGU2018-10934-1, 2018
EGU General Assembly 2018

GSio: A programmatic interface for delivering Big Earth data-as-a-service

Pablo Larraondo, Jian Guo, Joseph Antony, and Ben Evans

Australian National University, National Computational Infrastructure, Research Engagement and Initiatives, ACTON, Australia (pablo.larraondo@anu.edu.au)

Big Earth Data can greatly benefit from the ubiquitous and scalable characteristics of the Cloud. However, the Cloud's storage and computing models differ from those used in classic computing. Transferring traditional workflows and file formats into the Cloud normally results in suboptimal performance. There is an opportunity to define a new Cloud native model for storing geospatial data that enables large scale access.

We present GSio: a generic interface for delivering Big Earth Data, with a Cloud native implementation. Benefiting from the unbounded scalability of Cloud object stores, GSio splits geospatial data into small subsets containing multidimensional arrays, which are stored as individual objects. Fragmenting the data allows compute nodes to access objects in parallel, resulting in higher IO throughputs. GSio also makes use of fast compression algorithms to minimise the size of the data and its access latency. We demonstrate how this simple approach outperforms other traditional formats, such as HDF5 or GRIB, in Cloud environments.

This implementation is demonstrated by exposing a subset of ECMWF's ERA5 reanalysis dataset. Remote access to the data can be performed using a Python client, allowing its interactive use on Jupyter notebooks or deep learning frameworks such as Tensorflow. We will cover the design principles behind GSio, present performance benchmarks and carry out a live demonstration.