Geophysical Research Abstracts
Vol. 20, EGU2018-15967-1, 2018
EGU General Assembly 2018

Semantic storage of climate data on object stores

Neil Massey (1), David Hassell (2), Matt Jones (2), Phil Kershaw (3), and Bryan Lawrence (2)

(1) NCAS / Centre for Environmental Data Analysis, STFC, Rutherford Appleton Laboratory, United Kingdom (neil.massey@stfc.ac.uk), (2) Department of Meteorology, University of Reading, United Kingdom, (3) NCEO / Centre for Environmental Data Analysis, STFC, Rutherford Appleton Laboratory, United Kingdom

High performance computers, data analysis systems and data centres typically use large (often parallel) POSIX filesystems. However, with the advent of exascale climate data, such filesystems cannot deliver price/performance across all aspects of the workflow from simulation to analysis. Alternative methods of storing large amounts of data are necessary, potentially as part of tiered storage environments.

One alternative option is object storage, which potentially addresses many of the challenges of managing data at exascale, including decoupling from traditional POSIX semantics for user management and presenting a universal interface for access from inside the data centre and remotely across wide area networks. In particular, using the Amazon S3 interface allows hosting data more easily on public cloud or adopting hybrid-cloud scenarios. However, this approach raises a number of problems: 1) Determining a means of storage for large files that will facilitate efficient access; 2) Enabling meaningful search of data, when the data is distributed across objects rather than in a hierarchical file system; 3) Supporting existing applications which depend on POSIX semantics; and 4) Managing the distributed objects.

To address these issues large data sets can be broken into "fragments" stored as objects, each fragment containing a subset of the data. This allows for parallel access to the fragments, improving the performance of reading the data across a network - particularly if the application and storage have semantic information about the nature of the objects.

Here we present a new method of splitting netCDF files into subarray fragments and storing each fragment as an object in an object storage system with an S3 interface. The metadata and domain for each climate variable are stored in a master array file, as well as the location in the master array and URI of each subarray fragment. This master array file can be written to a location not within the object storage, for example a POSIX file system on a SSD. It allows for fast search and indexing of the metadata without having to reassemble the subarray fragments. Each subarray fragment is stored as a self-contained netCDF file within the object storage, complete with the metadata and subdomain for that subarray, allowing reconstruction of the data, if the master file is lost, using standard climate data tools.

The method is implemented in Python and Cython as the S3-netCDF4-python package, which is a subclass of the standard netCDF4-python package. This allows any improvements in the standard package, and lower software stack, to be automatically included in S3-netCDF4-python. Also, as a subclass of netCDF4-python, S3-netCDF4-python presents a well-known interface and data model to the scientist, hiding the physical location of the files and allowing scientists to continue using existing workflows based on netCDF4-python.

We conclude by providing some examples and presenting some performance metrics to compare with storing netCDF files on other disk systems or using other interfaces such as OPeNDAP. Some limited comparison with other object store approaches such as the HDF Groups's Highly Scalable Data Server will also be included.