Adaptive Tier Selection for NetCDF/HDF5

Jakob Luettgau*, Eugen Betke*, Olga Perevalova†, Julian Kunkel†, Michael Kuhn†
*German Climate Computing Center (DKRZ), †Universität Hamburg (UHH)

Abstract

Scientific applications on supercomputers tend to be I/O-intensive. To achieve portability and performance, data description libraries such as HDF5 and NetCDF are commonly used. Unfortunately, the libraries often default to suboptimal access patterns for reading/writing data to multi-tier distributed storage. This work explores the feasibility of adaptively selecting tiers depending on an application’s I/O behavior.

Overview

The contributions presented in this work are:

- A proof of concept prototype implementation demonstrating the benefit of adaptive tier selection on a real system.
- An architecture for I/O middleware beyond adaptive tier selection for more intelligent data placement from user space.

Opportunities using HDF5 Virtual Object Layer

Hierarchical Data Format (HDF5): HDF5 is an open source, hierarchical, and self-describing format that combines data and metadata. Advantages of this format make it widely used by scientific applications.

Virtual Object Layer (VOL): The VOL is an abstraction layer in the HDF5 library with the purpose of exposing the HDF5 API to applications while allowing to use different storage mechanisms. The VOL intercepts all API calls and forwards those calls to plugin object drivers. Additionally, external VOL plugins are supported to allow third-party plugin development.

Plugin for Separate Metadata Handling: A VOL plugin was developed to handle data and metadata separately. For adaptive tier selection, this is necessary to keep track of alternating data sources but it also offers additional opportunities. Generated simulation data is routinely published, but at the moment not automatically catalogued. With the VOL plugin, it would be possible to extract a dataset description to make it available for search in a catalogue as the dataset is written.

NetCDF Benchmark

NetCDF Performance Benchmark Tool (NetCDF-Bench) was used to recreate a typical checkpoint restart workload. NetCDF-Bench was developed to measure NetCDF performance on devices ranging from notebooks to large HPC systems. It mimics the typical I/O behavior of scientific climate applications and captures the performance on each node/process. Details: https://github.com/joobog/netcdf-bench

Acknowledgments

The ESIWACE project received funding from the EU Horizon 2020 research and innovation programme under grant agreement no. 675191. Disclaimer: This material reflects only the authors’ view and the EU commission is not responsible for any use that may be made of the information it contains.