EGU2020-15048
https://doi.org/10.5194/egusphere-egu2020-15048
EGU General Assembly 2020
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Polytope: Serving ECMWFs Big Weather Data

James Hawkes, Nicolau Manubens, Emanuele Danovaro, John Hanley, Stephan Siemen, Baudouin Raoult, and Tiago Quintino
European Centre for Medium-Range Weather Forecasts, Production Services, Forecast Department, Winchester, United Kingdom of Great Britain and Northern Ireland (james.hawkes@ecmwf.int)

Every day, ECMWF produces ~120TiB of raw weather data, represented as a six-dimensional dataset. This data is used to produce approximately 30TiB of user-defined products, which are disseminated worldwide. The raw data is also stored in the world's largest meteorological archive (MARS), currently holding over 300 PiB of primary data -- which is also served around the world on demand. As the resolution of ECMWFs global weather models increase over the next few years, the amount of raw data produced per day will increase into the petabytes, and the distribution of products and archived data becomes impossible. In-situ, on-the-fly data extraction and processing are required to sustain and increase the accessibility of ECMWFs big weather data.

To meet these requirements, ECMWF is developing Polytope -- an open-source service which allows users to request arbitrary n-dimensional stencils ("polytopes") of data from highly-structured n-dimensional datasets. The data extraction is performed server-side (collocated with the data), allowing for large data reduction prior to transmission and less complexity for the user. For example, a user could request a polytope describing a flight path -- simultaneously crossing temporal and spatial axes. Polytope will return just a few bytes of data rather than large structured arrays of geo-spatial data which must be further post-processed by the user.

Polytope is being partly developed under LEXIS, an EU-funded Horizon 2020 project which focuses on large-scale HPC & cloud workflows. The emphasis of LEXIS is on how HPC and cloud systems interact; how they can share data; and methods to compose workflows of tasks running on both cloud and HPC systems. Polytope will be used to provide a cross-centre weather and climate data API which connects to multiple high-performance data sources across Europe, and serves multiple cloud environments with this data.

This poster will present the early developments and future vision of Polytope. It will also illustrate how it is used within the LEXIS project to enable complex weather and climate workflows, involving global forecasts, regional forecasts and cloud-based simulations.