Science and Technology Facilities Council
Natural Environment Research Council

CLIMATE IMPACT EXPLORER

DIGITAL TWIN EARTH PRECURSOR

ESA Living Planet Symposium, 26 May 2022

Philip Kershaw
Centre for Environmental Data Analysis

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Scientific Computing
Science & Technology Facilities Council

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science and Technology Facilities Council
RAL Space

ESA Digital Twin Earth Precursor: Climate Use Case

- Help decision makers to generate and visualise information relating to regionalised impacts of **climate change** in real-time
- Land surface and drought risk (utilising JULES model) as focus
- Rapid prototyping and explorative
- Aggressive time scales

TELESPAZIO a LEONARDO and THALES company

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UNIVERSITY OF LEICESTER

University of Reading

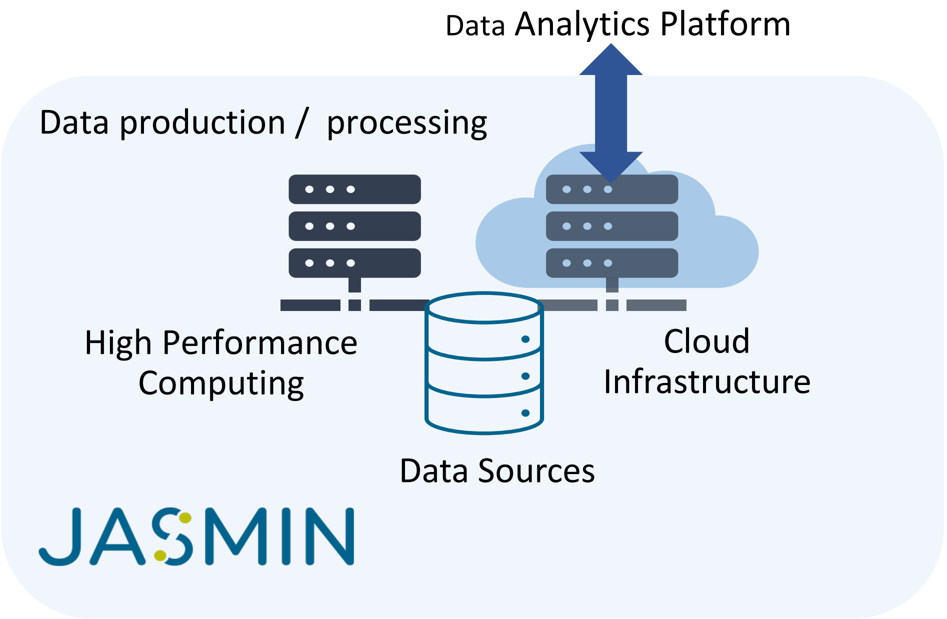
Summary – what did we achieve?

- Challenge: complex modelling required and access to specialised High-Performance Computing

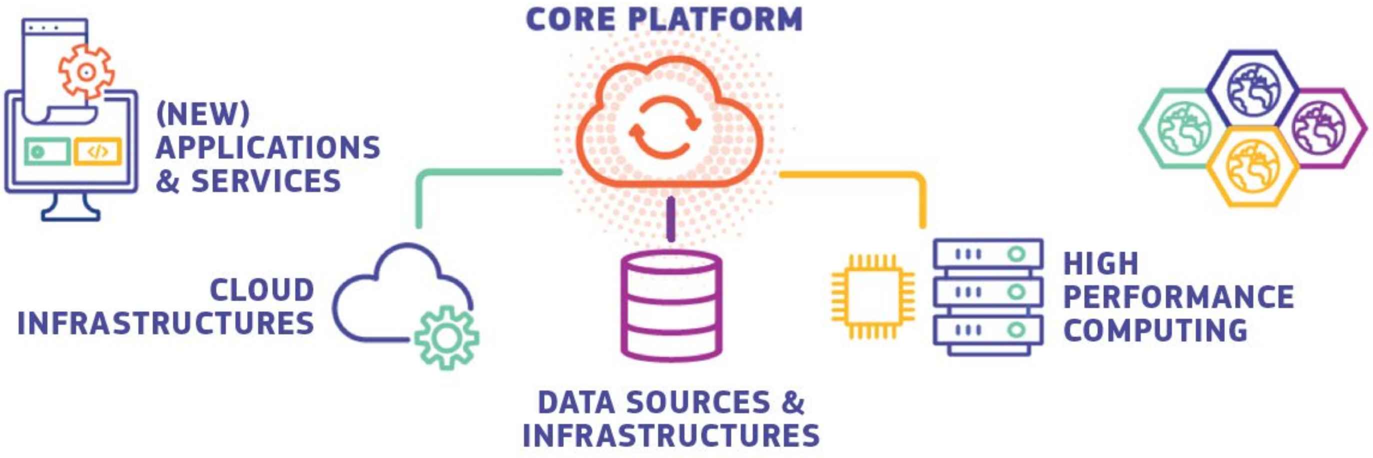
- Applied machine learning to develop an emulator
  - suitable for deployment on a commodity cloud computing environment
  - accessible on-demand when required by end-users

- The above made possible through key elements provided by underlying JASMIN infrastructure

JASMIN – a platform with the essential elements to build a Digital Twin Earth Precursor

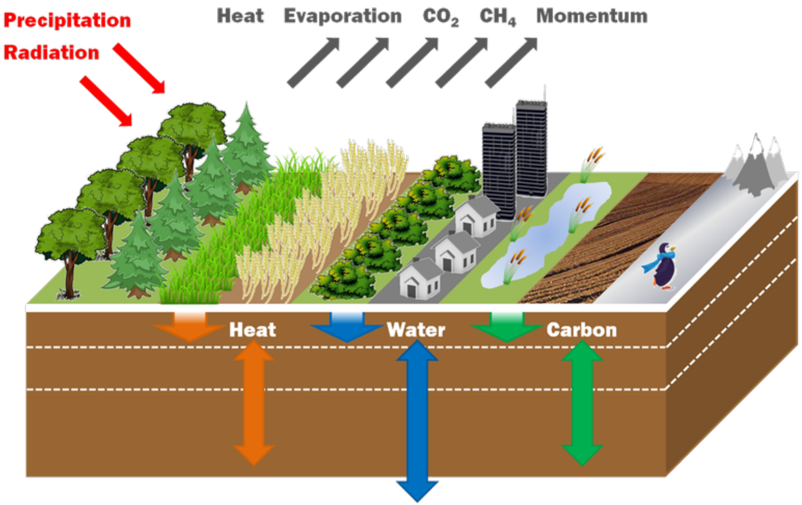


DestinE and Blueprint Architecture



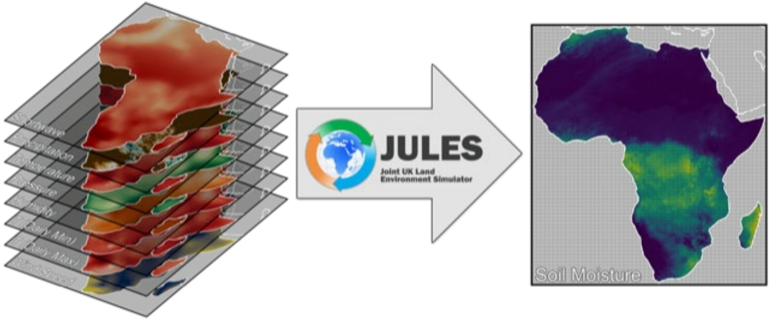
https://digital-strategy.ec.europa.eu/en/library/destination-earth

Climate Explorer focus on land surface modelling use case



- Using JULES (Joint UK Land Environment Simulator)
  - the land surface component in the Met Office Unified Model
- Improvements with LAVENDAR data assimilation system (Reading University)
  - Feed in satellite observations – SIF and SMAP data

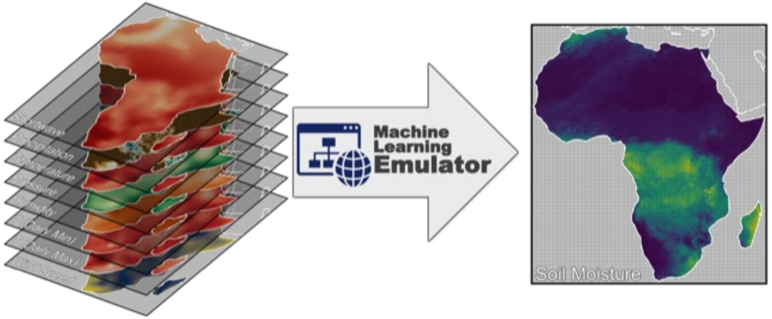
What could be the future impact of climate change on the soil moisture?



JULES driven with climate projections from ISIMIP data

(Inter-Sectoral Impact Model Intercomparison Project)

Make a surrogate AI model to JULES



- Experimented with different Machine Learning (ML) techniques
- Goal: a general-purpose algorithm -

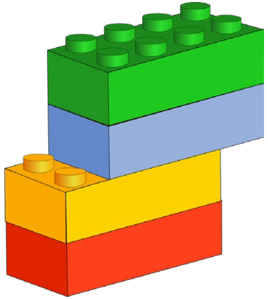
  `time series of daily weather data ➔ time series of soil moisture data`

- Successfully applied XGBoost (eXtreme Gradient Boosting) algorithm.
- Trained on up to 1000 grid cells, representative of the various biomes in continental Africa
- Demonstrated to accurately emulate JULES output at other locations in Africa
- The credibility of the model is enhanced by its transparency and explainability

JASMIN Cluster-as-a-Service to build data analysis environments from pre-configured building blocks

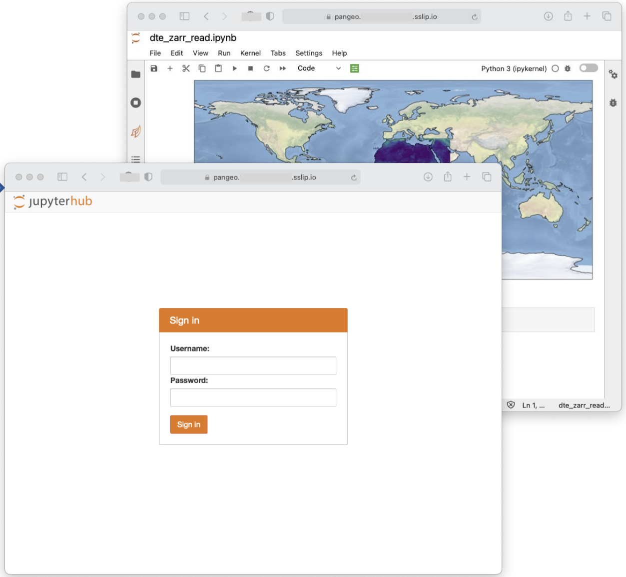
1) Cloud Portal – Cluster-as-a-Service

2) Select required building blocks and deploy

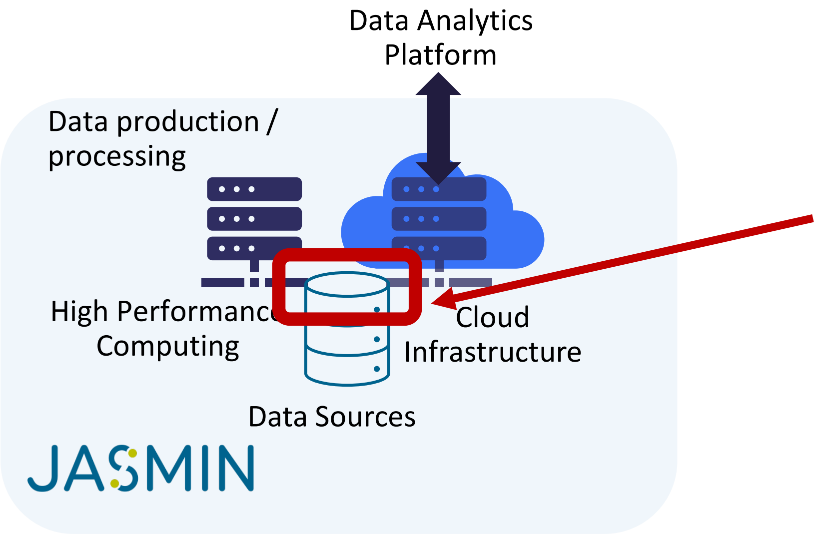


https://www.goodfreephotos.com/

3) Use deployed components



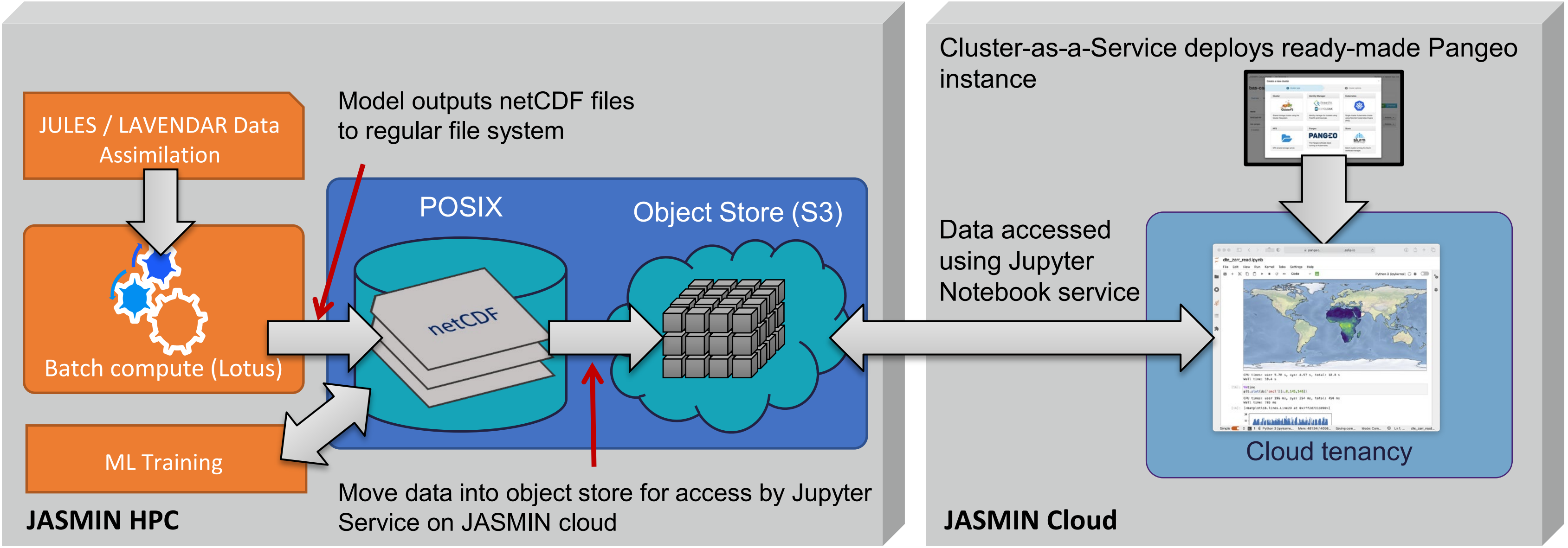
HPC and Cloud data access mismatch



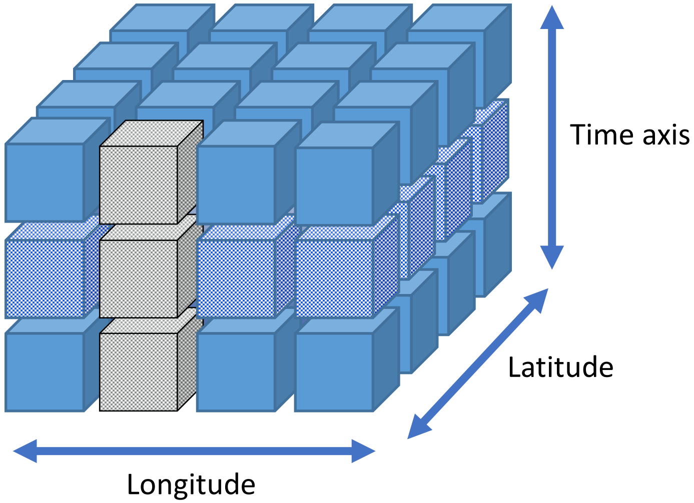
- HPC – POSIX file systems
- Cloud – object store
- Access and format translation needed to map from one to another

Philip.Kershaw@stfc.ac.uk

Bringing together HPC and Cloud-Friendly Storage



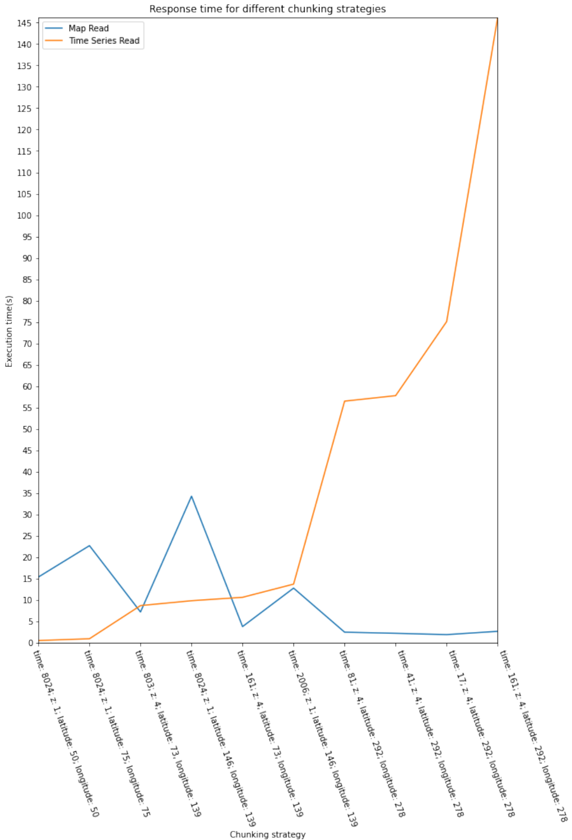
Arrangement of data and efficient access



- Data output from models as netCDF format
- Data in files arranged in spatial dimensions one per time step
- But predominate access pattern for analysis of climate data in the project is time series query (grey blocks)

Philip.Kershaw@stfc.ac.uk

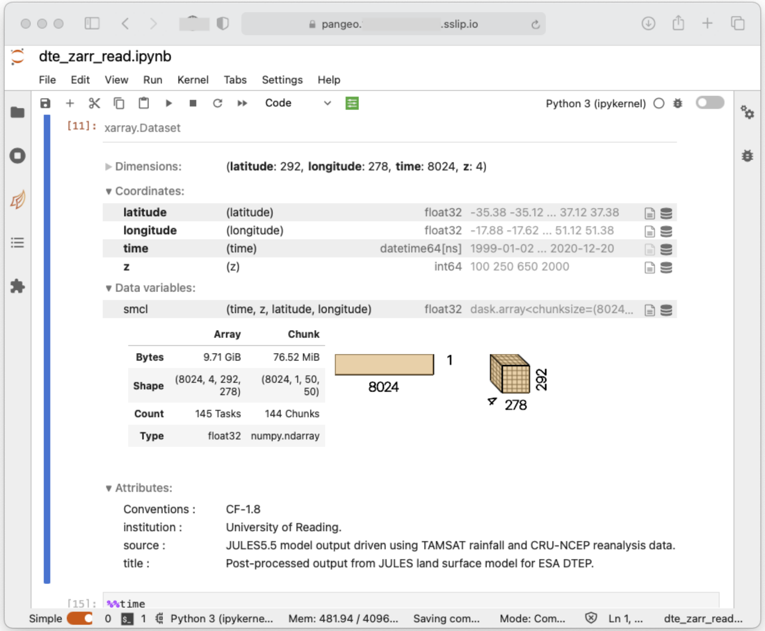
Object Store: Different storage strategies showed radically different performance



- We experimented with different storage chunking arrangements
- 20-year dataset of soil moisture

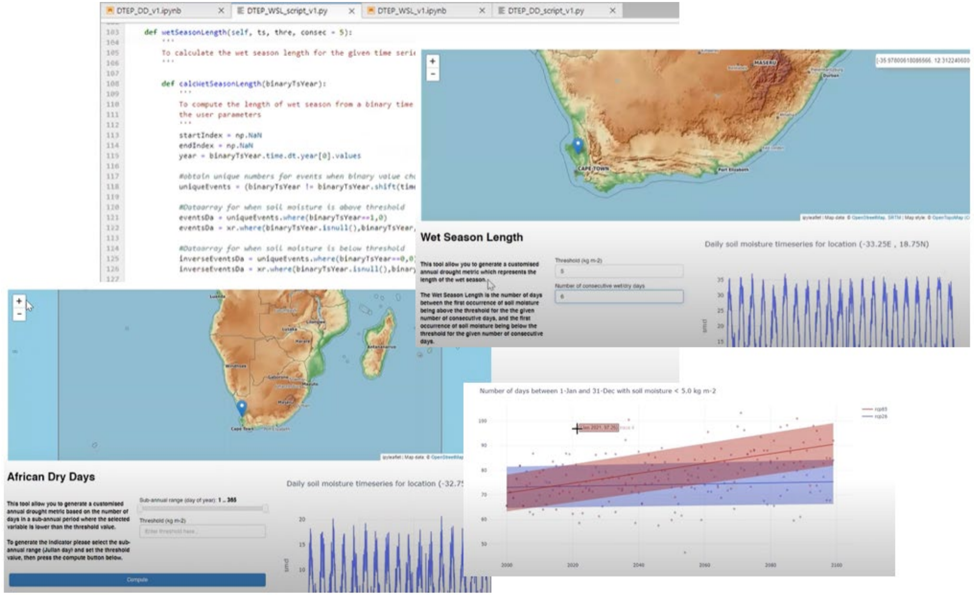
Philip.Kershaw@stfc.ac.uk

Using Object Store for re-arrangement of data to suite our access patterns



- Using zarr and xarray Python libraries to store and access the data
- Chunked data into a series of strips along the time axis

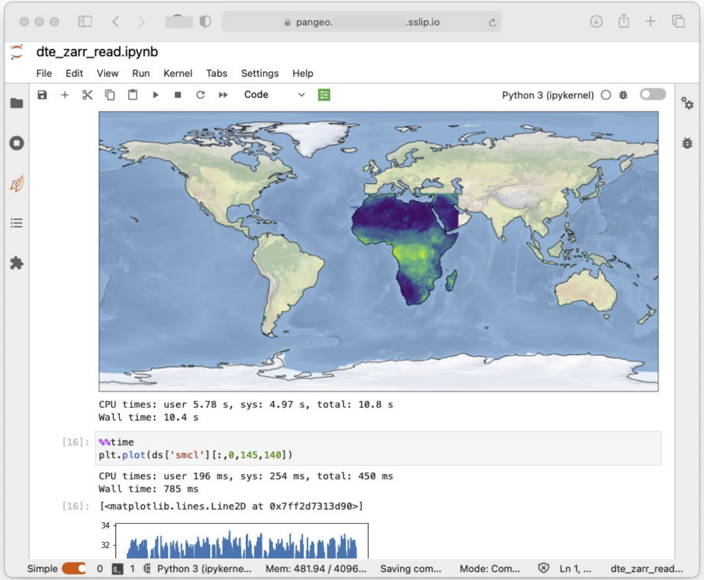
Rechunking of data made possible interactive maps with long time series



Philip.Kershaw@stfc.ac.uk

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Use of object store for analysis-ready cache specific to project needs

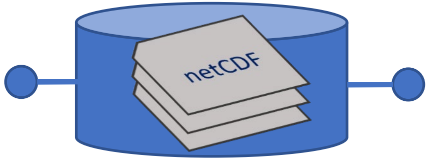


- Object store efficient for access on JASMIN Cloud
- Essential to orient data storage to suit predominant access patterns

Summary

- Emulator: suitable for commodity cloud instead of HPC needed for JULES
- JASMIN: provided HPC, Cloud and storage (object store + POSIX) essential high-level infrastructure constituents needed for the demonstrator
- Cluster-as-a-Service: Rapid prototyping environment possible from day one
- Analysis-ready caches of data on object store
  - Pattern emerging where we make copies of data on the fly arranged to support specific analysis use cases
  - Future: storage for simultaneous POSIX and S3 Interfaces – experiment with *Kerchunk*

POSIX



S3 ⇔ Kerchunk ⇔ zarr ⇔ xarray

Science and Technology Facilities Council
Natural Environment Research Council

Philip.Kershaw@stfc.ac.uk

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Further Information

- Climate Explorer web-site: https://climate-impact-explorer.org
- JULES: https://jules.jchmr.org
- JASMIN: https://jasmin.ac.uk

Philip.Kershaw@stfc.ac.uk